



HAL
open science

Structural and functional study of the human phosphatase PTPN3 and its interaction with oncogenic viruses

Mariano Genera

► **To cite this version:**

Mariano Genera. Structural and functional study of the human phosphatase PTPN3 and its interaction with oncogenic viruses. Biophysics. Sorbonne Université, 2019. English. NNT : 2019SORUS112 . tel-02947211

HAL Id: tel-02947211

<https://theses.hal.science/tel-02947211>

Submitted on 23 Sep 2020

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

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

Sorbonne Université

École doctorale ED515 Complexité du Vivant

Structural and functional study of the human phosphatase PTPN3 and its interaction with oncogenic viruses

Par Mariano Genera

Thèse de doctorat de Biophysique Moléculaire

Dirigée par Nicolas Wolff

Présentée et soutenue publiquement le 27 novembre 2019

Devant un jury composé de :

Dr Françoise Ochsenbein	Rapporteur
Dr Xavier Morelli	Rapporteur
Dr Aurélie Albertini	Examineur
Pr Catherine Vénien-Bryan	Examineur
Dr Célia Caillet-Saguy	Examineur

*A mis padres
Mercedes y Claudio*

Acknowledgements

I would like to dedicate the first pages of this manuscript to thank the many people that have been involved, directly or indirectly, in this thesis.

First of all, I want to say thanks to Célia. We decided to take this big risk without really knowing each other but with the feeling that it was a good move. And it was. These past three years I had the luck of working with an amazing supervisor. Thank you for the constant support, the encouragement and the patience (lots and lots of it...), and above all, for giving me this life-changing opportunity. I also take the chance to thank Nico, for your support, your sense of humour and your kindness and humanity. It was an immense pleasure to be a part of your group.

I would like to thank Muriel Delepierre and Pierre-Jean Corringer, the leaders of the two Units I had the luck of being a part of during my thesis. I found myself surrounded by great people who made this a fantastic experience both personally and scientifically.

Thanks to Yanlei and Baptiste for your company, your help and your friendship. Also the old members of our team, particularly to Florent, who was always available to teach me and give kind advice during the though first months, and to the students who collaborated with this project: Fariza, Anastasia and Antonin.

To all the members of the old NMR Unit, it was wonderful to share the first half of my thesis with you. To all the members of RCN lab, thanks for your good humour and all the stimulating scientific discussions.

I would like to thank the members of the jury, Dr Aurélie Albertini, Pr Catherine Vénien-Bryan, Dr Françoise Ochsenein, and Dr Xavier Morelli, for having accepted to dedicate their time to evaluate my work and for their insightful suggestions. Thanks as well to all the members of the PPU steering committee, for putting together this wonderful PhD program.

This work would not have been possible without all our collaborators. A big thanks to the members of the Biophysics and Crystallography platforms: Bertrand, Séb, Sylviane, Bruno, Patrick E, Ahmed, Ariel, Patrick W, Cédric and Rafa; for your dedication and for always being willing to teach me something new. And thanks to Barbara and Christine, for helping us to look at this project from a new perspective.

To my dear friends, the fantastic people I met here. An old friend once told me we are just a bunch of orphaned scientists who met in Paris and decided to form a family (deep stuff, eh?). Thanks for making it so easy to be far from the old one.

A mis amigos de Argentina, siempre a un whatsapp de distancia.

To Laura, who stuck with me through thick and thin. It would not be very ecological to try to thank you for everything. So thanks for the unfailing, unconditional support, especially during these last months (yep, I'm meaning the dinners!).

Finalmente, gracias a mis padres, Mercedes y Claudio, y a mis hermanos, Agustín, Santiago y Andrés, por enseñarme con su ejemplo a siempre intentar llegar un poco más lejos.

Index

Acknowledgements	2
Introduction.....	7
1 Protein tyrosine phosphatases	8
1.1 Tyrosine phosphorylation in cell signalling.....	8
1.2 Composition of the PTP family	10
1.2.1 Cystein-based classical PTPs – Structural features and catalysis mechanism	11
1.2.2 Receptor-like and non-transmembrane PTPs	14
1.3 PTPs as tumor suppressors or oncogenic proteins	15
1.4 PTPs in other diseases.....	18
1.5 PTPN3.....	18
1.5.1 Domain organization of PTPN3	18
1.5.2 PTPN3 phosphorylation sites	19
1.5.3 PTPN3 cleavage sites	20
1.5.4 14-3-3 recognition sites	21
1.5.5 Structural features of PTPN3.....	21
1.5.6 PTPN3 functions.....	25
1.5.7 PTPN3 in cancer	31
2 PDZ domains.....	34
2.1 Role of PDZ domains in the cell.....	34
2.1.1 Phototransduction in rhabdomeres	35
2.1.2 Organization of the postsynaptic density by PSD-95	37
2.2 Structure of PDZ domains.....	39
2.2.1 Peptide-binding groove.....	40
2.3 Classification of PDZ domains	40
2.4 Structural determinants of ligand binding.....	41
2.4.1 Specificity at the C-terminal position.....	41
2.4.2 Recognition of position -2	41
2.4.3 Recognition of position -3	42
2.4.4 Recognition of position -1	43
2.4.5 Interactions beyond position -3	44
2.5 The affinity of PDZ-mediated interactions	45
2.6 Non-canonical binding modes and internal binding motifs	46
2.7 Other functions of PDZ domains	48
2.7.1 PDZ multimerization	48
2.7.2 Heterotypic supramodules	50
2.7.3 Phosphatidylinositol and cholesterol binding.....	51
2.8 Regulation of PDZ-mediated interactions.....	52
2.8.1 Phosphorylation	52
2.8.2 Auto-inhibition	53
2.8.3 Allosteric conformational regulation.....	54
2.8.4 Alternative splicing.....	55
3 Targeting of PDZ domains by viruses	56
3.1 Viral interactions with PDZ-containing proteins	56
3.1.1 Cell polarity proteins: a common target	57
3.1.2 Human adenovirus	59

3.1.3	Human T-lymphotropic virus type 1	60
3.1.4	Rabies virus	61
3.2	Viruses targeting PTPN3	63
3.2.1	Human papillomavirus.....	63
3.2.2	Hepatitis B Virus	73
4	Project objectives	81
	Results.....	83
5	Structural and functional characterization of the PDZ domain of PTPN3 and its interaction with the high-risk HPV type 16 E6 oncoprotein	84
5.1	Summary	84
5.2	Article: Structural and functional characterization of the PDZ domain of the human phosphatase PTPN3 and its interaction with the human papillomavirus E6 oncoprotein ...	81
5.3	Comments	94
5.4	Additional results: kinetic studies of PTPN3.....	95
5.4.1	Context.....	95
5.4.2	Results.....	96
5.4.3	Discussion and perspectives	99
5.4.4	Materials and methods.....	103
6	Structural determinants of PDZ ligand selectivity for the NT5 subfamily of phosphatases.....	105
6.1	Introduction.....	105
6.2	Results and discussion	106
6.2.1	Crystal structures of PTPN3-PDZ/peptide ligand complexes	106
6.2.2	Specificity of recognition of PTPN3-PDZ	107
6.2.3	Recognition of atypical PBMs with a C-terminal cysteine	112
6.2.4	Binding specificity of PTPN3-PDZ for cellular proteins	115
6.3	Materials and methods	118
7	Study of the interaction of PTPN3 with the hepatitis B virus core protein	120
7.1	Summary	120
7.2	Article: Molecular basis of the interaction of the hepatitis B virus core protein with the human phosphatase PTPN3	122
7.2.1	Highlights	123
7.2.2	Abstract.....	123
7.2.3	Introduction.....	124
7.2.4	Results.....	126
7.2.5	Discussion.....	134
7.2.6	Materials and Methods	137
7.3	Comments	159
	Concluding remarks.....	160
8	Discussion and perspectives.....	161
	Appendices.....	166
9	Appendices.....	167
9.1	Table: The extended human PTPome.	167
9.2	Table: Summary of Reactome pathway participation of PDZ-containing proteins..	172
9.3	Table: Main functions and molecular targets of E6 and E7 HPV oncoproteins	174
9.4	Table: Data collection and refinement statistics for structures in section 6.2.....	176
9.5	Table: PTPN3-PDZ partners fished by pull-down from HeLaS3 cell lysates	177

10	Index of figures	180
11	Index of tables.....	182
12	List of abbreviations	183
	Bibliography.....	185

Introduction

1 Protein tyrosine phosphatases

1.1 Tyrosine phosphorylation in cell signalling

Cells respond to external physical or chemical stimuli by transmitting these signals to the cell interior through a series of molecular events in a process known as signal transduction. Signal transduction follows pathways that ultimately regulate cellular processes and determine physiological outcomes in response to stimuli. These pathways are interconnected forming networks that function co-ordinately to enable global, diverse and effective cellular responses.

Signal transduction pathways often involve protein phosphorylation. The levels of phosphorylated species are maintained and can be rapidly altered by the action of protein kinases and protein phosphatases. It is estimated that at least one third of the proteome is phosphorylated in eukaryotes¹. Although tyrosine phosphorylation accounts for just a fraction of the phosphoproteome (less than 2%, while 98% corresponds to serine and threonine phosphorylation), accumulating evidence highlights its importance as a crucial regulatory mechanism in eukaryotic cell physiology²⁻¹⁰. An example of this can be found in the mammalian immune system. Briefly, lymphocyte differentiation and activation are mediated by specific antigen receptors, which typically possess conserved cytoplasmic tyrosine-based activation motifs that become phosphorylated upon receptor engagement¹¹. These events lead to the activation of protein tyrosine kinases (PTKs) such as Src, Syk, and Zap70, which trigger a cascade of downstream responses, including effector functions and T and B cell differentiation^{10,11}. Tyrosine phosphorylation is a central mechanism in these processes; accordingly, several protein tyrosine phosphatases (PTPs) have been clearly linked to lymphocyte activation (table 1.1). These PTPs can function as either positive or negative regulators of lymphocyte activation and differentiation. For example, in mice models, CDC45 acts as an activator, and its expression is necessary for adequate levels of T and B cell activation, B cell development and NK cell function. SPH-1, on the contrary, has an inhibitory role on T cell responses, and SPH-1 knockdown mice exhibit augmented B cell receptor-triggered Ca^{2+} fluxes, increased serum antibody concentrations and develop autoimmunity upon aging¹⁰. This highlights that PTPs are signalling molecules that can function as activators or repressors of signalling pathways.

Table 1.1: Protein tyrosine phosphatases involved in lymphocyte activation. Major features of the main PTPs linked to the regulation of lymphocyte activation. DCs, dendritic cells. Adapted from Rhee and Veillette 2012¹⁰.

Type	Name	Lymphocyte phenotypes in deficient mouse strains	Involvement in human autoimmunity
Receptor	CD45 (PTPRC)	Considerable block in T cell development; partial block in B cell development and B cell activation (severe when combined with CD148 deficiency); lower NK cell function	Intronic polymorphism (C77G) associated with multiple sclerosis and autoimmune hepatitis (observed in one cohort only)
	CD148 (PTPRJ)	Partial block in B cell development and B cell activation (severe when combined with CD45 deficiency)	None
Nonreceptor	PTPN22 (LYP, PEP)	Accumulation of effector-memory T cells; expanded germinal centres; higher antibody concentrations; enhanced TCR-mediated signalling and responses	Coding polymorphism (C1858T; R620W) linked to type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, Grave's thyroiditis, myasthenia gravis and generalized vitiligo (observed in several cohorts)
	SHP-1	Greater CD8 ⁺ T cell responses; altered B cell development; augmented BCR-triggered Ca ²⁺ fluxes; higher serum antibody concentrations; autoimmunity (older mice)	None
	PTPN2 (TC-PTP)	Greater thymocyte positive selection; accumulation of effector-memory T cells; autoantibodies; inflammatory illness; greater TCR- and IL-2 receptor-mediated signals and proliferation	Intronic polymorphism (rs1893217(C)) associated with type 1 diabetes, rheumatoid arthritis, Crohn's disease and celiac disease (observed in several cohorts)
	PTP-PEST	Lower secondary T cell responses; greater sensitivity to T cell anergy; less susceptibility to experimental autoimmune encephalomyelitis	None

1.2 Composition of the PTP family

The mammalian PTPome is the ensemble of protein with Tyr-dephosphorylation or Tyr-dephosphorylation-like (nonproteinaceous substrates, including phosphoinositides and phosphorylated carbohydrates) activities (detailed in section 9.1). The PTPome comprises structurally diverse protein families that share a common catalytic mechanism, in which a nucleophile attacks the substrate's phosphate group, forming a transient enzyme-substrate intermediate¹² (described below for Cys-based PTPs). Three PTP families can be defined based on the catalytic residue involved in this mechanism: Cys-based PTPs, His-based PTPs, and Asp-based PTPs. Cys-based PTPs, defined by the conserved motif C(X)₅R (where X is any amino acid) in their active site, are the largest of these families, and can be divided into three classes based on their evolutionary origin (figure 1.1)^{4,13}.

					Substrate
Cys-based CxxxxxR	Class I	Subclass I	Classic	37	pTyr, PIPs
		Subclass II	VH1-like	64	pTyr, pSer, pThr, PIPs, other
		Subclass III	SACs	5	PIPs
		Subclass IV	PALD1	1	? = unknown substrate
		Subclass V	INPP4s	2	PIPs
		Subclass VI	TMEM55s	2	PIPs
	Class II		LMW-PTP	1	pTyr
			SSU72	1	pSer
	Class III		CDC25s	3	pTyr, pThr
	Asp-based	HAD	EYAs	4	pTyr, pSer
His-based	PGM	UBASH3s	2	pTyr	
	Acid-phosphatases	ACPs	3	pTyr, pSer, pThr, Other	

Figure 1.1: Classification of the Tyr phosphatase superfamily. Tyr phosphatases are grouped into three families on the basis of their nucleophilic catalytic residue (Cys, Asp, or His) and topology. Each colour designates a family of Tyr phosphatases that share a common nucleophilic catalytic residue. Different families inside a group are indicated by different tones of the same color. The Cys-based PTPs Tyr phosphatases all share a CxxxxxR catalytic signature motif and are grouped into three families on the basis of their nucleophilic catalytic residue (Cys, Asp, or His) and topology. The numbers indicate how many members are included in each group. The generic substrates of each family are indicated in the right column. Figure from Alonso and Pulido, 2016¹⁴.

Class I is the largest group of phosphatases identified so far, with over 100 members, most of which share conservation in their primary sequences and in the fold of their catalytic

domain. These are the ‘classic’ phosphor-Tyr specific PTPs. More divergent members have been recently added on the basis of the presence of the conserved C(X)₅R motif¹⁴. Among these, an important subgroup is the subclass II, which contains dual specificity phosphatases (DSPs) that dephosphorylate phosphoserine (p-Ser) and phosphothreonine (p-Thr) residues, as well as p-Tyr, thanks to differences in the configuration of their active sites (see below). Class II contains the low molecular weight PTP (LMW-PTP), a small, widely expressed phosphatase group implicated in dephosphorylation of growth factor receptors, cell signalling proteins and cell migration proteins; and SSU72, which displays similar topology to LMW-PTP and is essential for cell viability. Class III is composed of the CDC25A, CDC25B, and CDC25C phosphatases. These proteins are cell cycle regulators and participate in control of the DNA damage response. They are structurally unrelated to the rest, as they lack a PTP domain, although they contain a conserved C(X)₅R motif. The “classical” PTPs are described more in detail below.

1.2.1 Cystein-based classical PTPs – Structural features and catalysis mechanism

The classical PTPs are encoded by 37 genes¹⁴, and possess catalytic domain that span over 250 to 280 aminoacids. The catalytic domains of classical PTP display a well-conserved fold, and several loops that define the active site architecture and contain conserved residues that facilitate the progression of the two-step hydrolysis reaction. The structural features of the conserved classical PTP domain, exemplified by PTP1B, are depicted in figure 1.2. In the following description, the numbering of the residues refers to the PTP domain of PTP1B as shown in figure 1.2 (PDB ID 2HNQ).

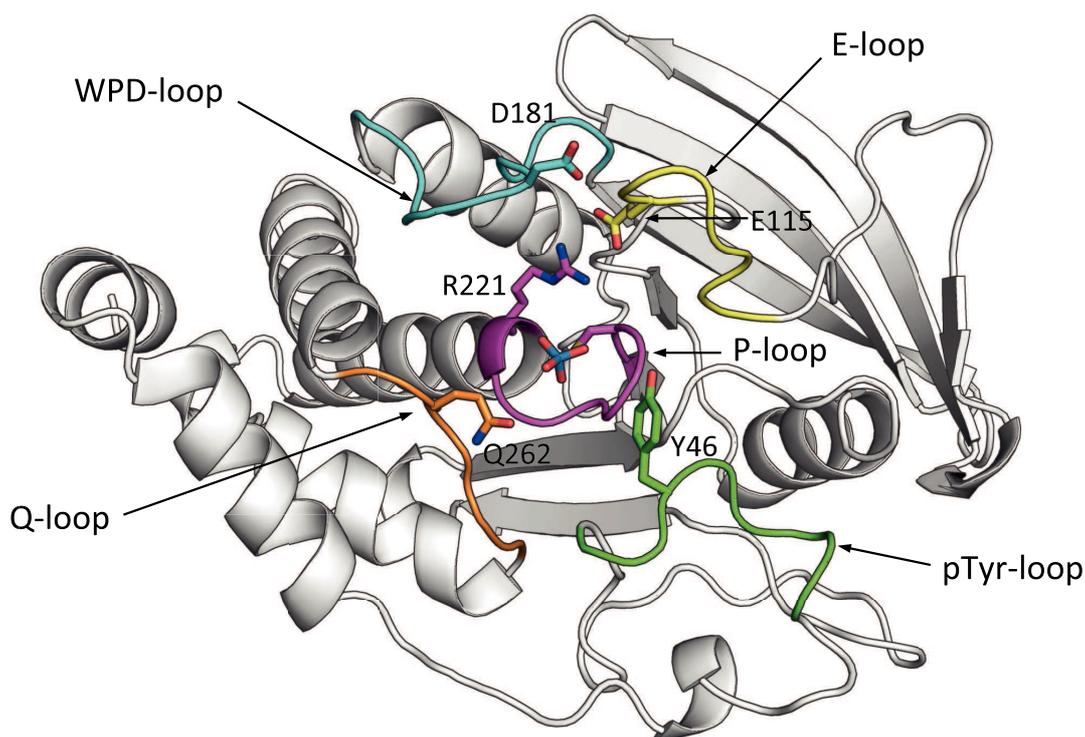


Figure 1.2: Structure of a classical class I PTP catalytic domain bound to tungstate (shown as sticks in CPK colours)(PTP1B; PDB ID 2HNQ). Conserved residues important for catalysis are shown as sticks: the catalytic cysteine (C215) and the invariant arginine (R221) in the P-loop; the catalytic acid/base aspartate (D181) in the WPD-loop; the conserved glutamine (Q262) of the Q-loop; tyrosine (Y46) of the pTyr-recognition loop (pTyr-loop); the conserved glutamate (E115) of the E-loop.

The P-loop, where the C(X)₅R signature motif is located, is present in all PTPs and constitutes the base of the active site. The conserved arginine residue (R221) in the P-loop enhances the positive microenvironment of the catalytic pocket, which increases affinity for the negatively charged phosphate group, and also promotes the nucleophilic role of the catalytic cysteine (C215) by lowering its pK_a from its usual value of around 8.5 to 4.5-5.5¹⁵, enabling it to function as a nucleophile at physiological pH (figure 1.3). Other loops are: the WPD-loop, which contains an invariant catalytic acid/base aspartate (D181); the Q-loop, in which a conserved glutamine coordinates the water molecule necessary for hydrolysis; the pTyr-loop, where a conserved motif specifically recognizes p-Tyr; and the E-loop, that is thought to modulate the dynamics of the WPD-loop¹⁶. In the first step of the reaction, the catalytic C215 of the P-loop initiates the nucleophilic attack on the phosphorus atom to break the phosphorus-oxygen bond (figure 1.3). The catalytic D181 in the WPD loop acts as a general acid, giving a proton to the OH-leaving group. This results in the formation of a phosphocysteine intermediate and the release of the dephosphorylated substrate. In the second step, which is rate-limiting, D181 functions as a general base, taking a proton from a water

molecule and thus promoting the hydrolysis of the phosphorous-sulphur bond of the intermediate (figure 1.3). This way, the free phosphate is released. The catalytic mechanism of His-based and Asp-based PTPs operates through a similar two-step reaction.

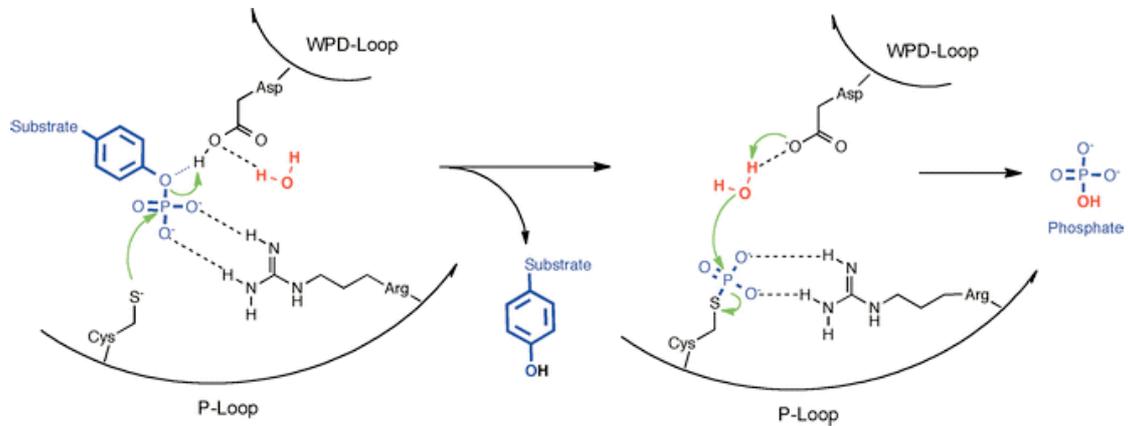


Figure 1.3: Scheme of the catalytic mechanism of classic phosphatases. Dephosphorylation of Tyr-phosphorylated proteins follows a two-step mechanism. In the initial step (left), the catalytic Cys in the P-loop initiates a nucleophilic attack on the phosphate group of the substrate, facilitated by the conserved R, forming a phosphoenzyme intermediate and releasing the dephosphorylated substrate. An Asp in the WPD loop works as general acid in this step of the reaction, donating a proton to the tyrosyl group. The second step of the reaction (right) restores the enzyme after hydrolyzing the phosphoenzyme intermediate. The catalytic Asp working as a general base deprotonates a water molecule, which now acts as a nucleophile to free the phosphate group from the enzyme. Figure from Tautz *et al.* 2013¹⁶.

It is worth noting that differences between different families of PTPs have resulted in different structural requirements for catalysis. Indeed, only the P-loop is conserved in all PTPs. The WPD-loop is largely conserved in most PTPs, with the exception of myotubularin-related phosphatases and CDC25, in which it is not needed for catalytic activity. The p-Tyr loop constitutes the structural basis of the different specificities of classical PTPs and DSPs. The p-Tyr loop provides the structural basis by which classical PTPs can discriminate p-Tyr from p-Ser/p-Thr¹⁶. As it was mentioned before, the catalytic Cys of the C(X)₅R motif is located at the base of the active site. The p-Tyr loop contains a KNRV motif, with a highly conserved V (V46) that is positioned in such a way that generates a deeper catalytic pocket than a p-Tyr, but neither a p-Ser nor a p-Thr, can reach. Moreover, it facilitates p-Tyr binding by establishing aromatic π - π interactions (figure 1.4). DSPs lack this loop, which results in a more accessible active site that can accommodate p-Ser /p-Thr¹⁶.

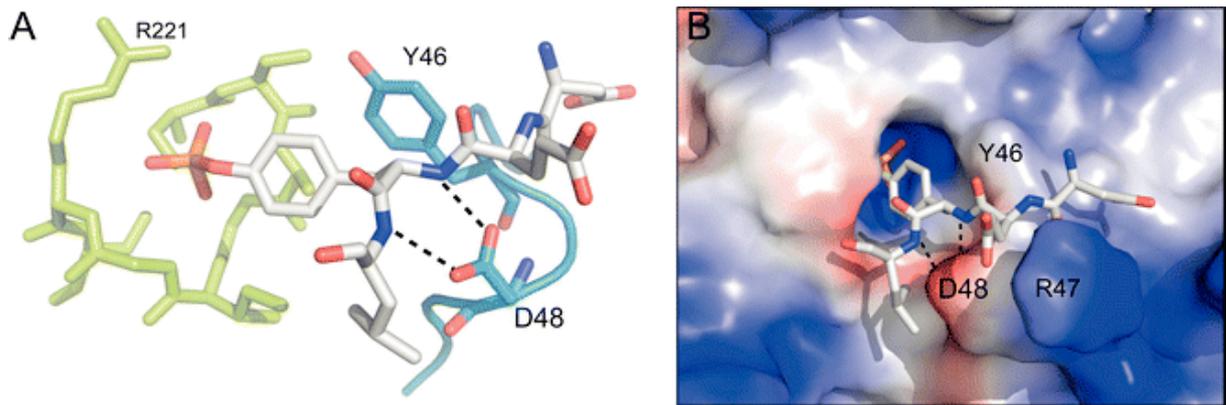


Figure 1.4: The pTyr-loop. (A) pTyr-loop (blue) relative to the P-loop (lime), with bound pTyr-peptide (white) (PTP1B; PDB ID: 1PTT). The conserved tyrosine (Y46 in PTP1B) determines the depth of the catalytic pocket, and favours pTyr binding by establishing aromatic π - π interactions. The conserved aspartate or asparagine (D48 in PTP1B) stabilizes the bound substrate establishing hydrogen bonds with backbone nitrogen atoms of the substrate peptide. (B) Same complex as in (A) with PTP1B rendered in electrostatic surface representation (blue, most positive; red, most negative). Figure from Tautz *et al.* 2013¹⁶.

1.2.2 Receptor-like and non-transmembrane PTPs

Classical PTPs can be in turn classified based on their subcellular localization, falling into two categories: receptor-like PTPs (RPTPs) and non-transmembrane PTPs (NTPTPs)(figure 1.5). RPTPs are transmembrane proteins with a single transmembrane helix of around 30 residues¹⁷. In the cytoplasmic region, they display one or two conserved PTP domains, of which the membrane-proximal one is catalytically active. The function of the membrane-distal domain is not fully understood. It has been proposed to determine substrate specificity of the membrane-proximal domain *in vitro*¹⁸ or enzymatic stability and activity¹⁹, as well as mediating protein-protein interactions that regulate dimerization of RPTPs²⁰. On the extracellular end, they are highly variable and display a combination of ligand binding or cell-cell or cell-matrix interaction domains, such as fibronectin type III-like and immunoglobulin-like domains. RPTPs are thus sub-classified into different types according to the domains they display in their extracellular region²¹ (figure 1.5). Non-transmembrane PTPs are cytosolic proteins with a single catalytic domain. Many of these proteins are modular and display a variety of domains that can regulate activity, mediate interactions, or determine their subcellular localization (figure 1.5). This categorization based on the presence or absence of a transmembrane and extracellular domain is broadly used.

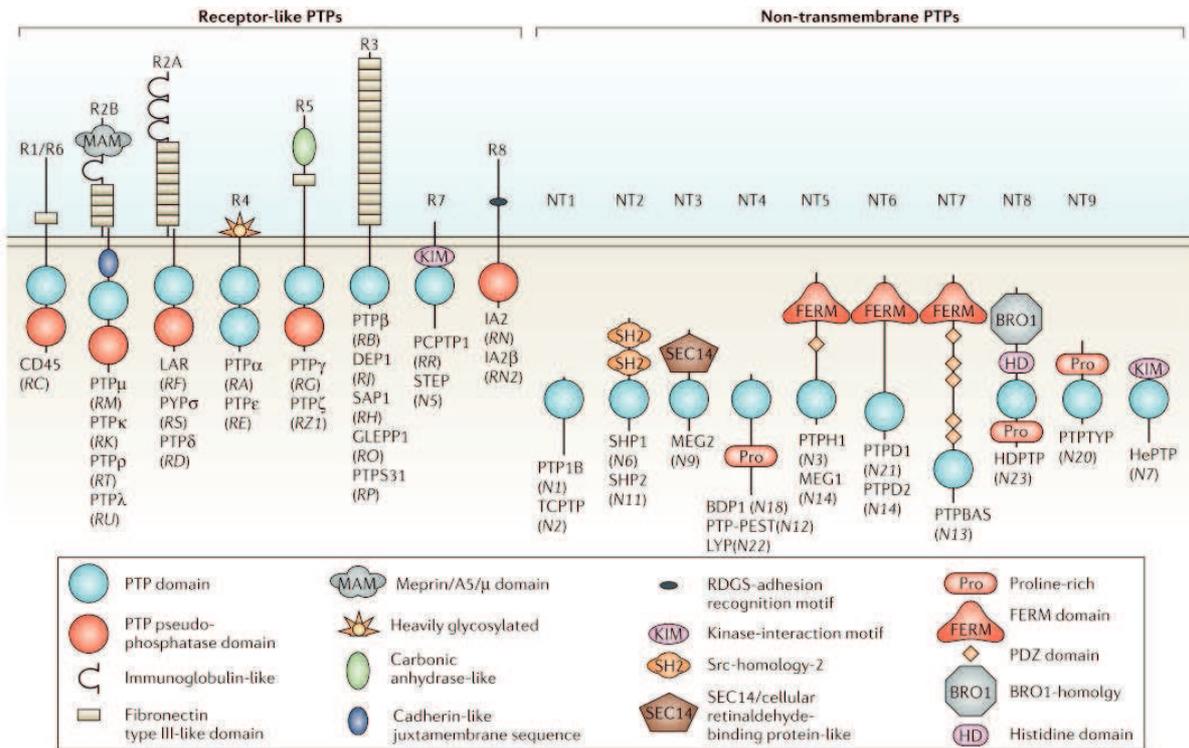


Figure 1.5: Classification of classical PTPs based on their subcellular localization. The classical protein tyrosine phosphatases (PTPs) can be categorized as receptor-like (R) or non-transmembrane (NT) proteins. In each case, the PTPs have been designated by a name that is commonly used in the literature. Where this differs from the gene symbol, the latter is included in parentheses for clarification. In each case, the various subdivisions are based upon sequence similarity. The numbers above the schematics indicate categories based on catalytic domain sequence conservation defined by Andersen et al²². Figure from Tonks 2006⁶.

Splicing variants or alternative use of promoters can generate transmembrane and cytosolic PTP isoforms from the same original gene, as it has been described for PTP ϵ ²³. Additionally, PTPs can be post-translationally modified. For example, PRL-2 is farnesylated, and this is a requirement for the interaction of the phosphatase with the β -subunit of geranylgeranyltransferase II²⁴. These variations add to the complexity of the classical PTPs family, which greatly surpasses the number of genes.

1.3 PTPs as tumor suppressors or oncogenic proteins

During the development of cancer, cells acquire certain characteristic alterations: independency of growth signals, insensitivity to growth inhibition, evasion from apoptosis, loss of differentiation, sustained angiogenesis and capacity to invade tissues and metastasize. These properties arise from aberrant functioning of the signalling pathways that regulate

proliferation, motility and survival of normal cells. Tyrosine phosphatases play a major role in positive and negative regulation of such signal transduction pathways, and disturbance of tyrosine phosphorylation is known to be essential for the development of many types of cancer²⁵ (table 1.2). PTPs were initially considered to counteract the oncogenic activation of PTKs, and were therefore expected to act as tumor suppressors. For example, PTEN, known to act as a tumor suppressor in various common malignancies (breast and prostate cancer, glioblastoma, and melanoma^{26–30}), dephosphorylates inositol phospholipids, counteracting the proliferative signals of phosphatidylinositol-3-kinase (PI3K). Furthermore, a study of mutations in tyrosine phosphatase genes in colorectal cancer by Wang et al.³¹ revealed 83 somatic nonsynonymous mutations affecting 6 PTPs: the RPTPs PTPRT, PTPRF and PTPRG, and the FERM-containing NTPTPs PTPN3, PTPN13 and PTPN14. In line with this, it has been shown that the loss of PTPRJ increases susceptibility to colon cancer^{32,33}. Downregulation of another NTPTP, SPH1 (*PTPN6*), has been observed in myelodysplastic syndrome³⁴ and T-cell lymphomas³⁵. PTPN12 functions as a tumor suppressor in triple-negative breast cancer cells by inhibiting multiple oncogenic tyrosine kinases, including HER2 and EGFR, and is frequently inactivated in breast cancer³⁶.

Accumulating evidence has demonstrated, however, that PTPs can also act as activators of signalling pathways, and a number of PTPs are overexpressed or present activating mutations in cancer cells. PTPN11 was the first PTP oncogene to be identified^{37,38} and normally facilitates activation of Ras, a class of small GTPases involved in transmitting signals within cells. Activating mutations in PTPN11 can favour the development of certain malignancies, such as sporadic juvenile myelomonocytic leukaemia, B-cell acute lymphoblastic leukemia and acute myeloid leukaemia^{39–42}. Other NTPTPs present similar activities: PTPN1 and PTPN7 are overexpressed in various cancers. Then, the dual-specificity phosphatase DUSP1, and the cell-cycle regulators PTP4A3 and the CDC25 A, B, and C isomers, are overexpressed in a number of malignancies^{25,43–46}. Overexpression or activation of RPTPs such as PTP ϵ has also been reported to promote tumor growth⁴⁷.

PTPN3 has been linked to some types of cancer, and this will be discussed in section 1.5.7.

Table 1.2: Protein tyrosine phosphatases involved in cancer. Adapted from Östman, Hellberg and Böhmer 2006, and Tonks 2006^{6,48}.

PTPs with a tumor suppressing function	
PTP (gene)	Observations
PTPBAS (<i>PTPN13</i>), PTPD2 (<i>PTPN14</i>), PTPH1 (<i>PTPN3</i>), PTP ρ (<i>PTPRT</i>), LAR (<i>PTPRF</i>), PTP γ (<i>PTPRG</i>)	Mutations identified in colon cancer; inactivating mutations confirmed for some variants of <i>PTPRT</i>
DEP1 (<i>PTPRJ</i>)	Overexpression reverts the transformed phenotype of different tumour cells; loss of heterozygosis is seen in colon cancer; <i>Ptprj</i> is the gene present at the colon-cancer-susceptibility locus in STS/A mice
SHP1 (<i>PTPN6</i>)	<i>PTPN6</i> is inactivated by promoter methylation in leukaemia, lymphoma and multiple myeloma
GLEPP1 (<i>PTPRO</i>)	<i>PTPRO</i> is inactivated by promoter methylation in lung and colorectal cancer; re-expression is associated with tumour suppression
PTP1B (<i>PTPN1</i>)	Generation of <i>Trp53</i> and <i>Ptp1b</i> double-knockout mice revealed decreased survival and increased development of B-cell lymphomas compared with <i>Trp53</i> knockout mice
PTEN (<i>MMAC1</i>)	Tumour suppressor mutated in various human cancers.
PTP κ (<i>PTPRK</i>)	Potential tumour suppressor in primary central nervous system lymphomas
PTPN12	Functions as a tumor suppressor in triple-negative breast cancer, its function frequently compromised due to deletions, mutations or loss of expression
MKP3 (<i>DUSP6</i>)	Candidate pancreatic tumour suppressor. Downregulation by promoter methylation
PTPs with a confirmed or potential oncogenic role	
SHP2 (<i>PTPN11</i>)	Mediates signalling of oncogenic protein-tyrosine kinases; activating mutations have been found in leukaemia and other tumours; SHP2 mutants are oncogenic in mouse models; might also have a role in <i>Helicobacter pylori</i> -induced gastric carcinoma
PTP α (<i>PTPRA</i>)	Has transforming activity in rodent fibroblasts; is overexpressed in some solid tumours; has an established function in the activation of the Src family kinases.
PTP ϵ (<i>PTPRE</i>)	Cooperates with the Neu (activated rat ERBB2) oncoprotein in mammary carcinogenesis in mice; can activate SRC
SAP1 (<i>PTPRH</i>)	Overexpressed in gastrointestinal and colorectal tumours. Inhibits cell growth and cell spreading; promotes apoptosis in transfected cell lines
PTP-LAR (<i>PTPRF</i>)	Overexpressed in mammary and thyroid cancer
PTP1B (<i>PTPN1</i>)	Has the capacity to activate SRC; overexpressed in mammary and ovarian carcinoma; reduced Ras activation by growth factors in <i>Ptpn1</i> ^{-/-} mouse fibroblasts indicates a positive role of PTP1B in this pathway
SHP1 (<i>PTPN6</i>)	Can activate SRC; is implicated in the positive regulation of Ras in some cell types; is overexpressed in a large fraction of mammary carcinomas, and in ovarian epithelial tumours
HePTP (<i>PTPN7</i>)	Overexpressed in myeloid malignancies and TNBC. Can transform rodent fibroblasts
CDC25	Cell-cycle control. Target of Myc and overexpressed in primary breast cancer
PRL3 (<i>PTP4A3</i>)	Upregulated in metastases of colon cancer

1.4 PTPs in other diseases

Besides their role in carcinogenesis, altered function of tyrosine phosphatases is also implicated in a range of human diseases including cardiovascular, immunological, infectious, neurological and metabolic diseases²⁵. The implication of PTPs in hereditary diseases has been extensively reviewed by Hendriks and colleagues⁴⁹; mutations and allelic variants in a number of phosphatases are linked to a wide range of hereditary pathologies including diabetes, arthritis, Noonan and Leopard syndromes, multiple sclerosis, restless legs syndrome and inflammatory bowel disease. A good example is a single-nucleotide polymorphism in the PEST-domain containing phosphatase PTPN22, which has been reported to be the causative agent of autoimmune diseases such as type 1 diabetes⁵⁰, rheumatoid arthritis⁵¹, systemic lupus erythematosus⁵², among others. Another example is the striatal-enriched protein tyrosine phosphatase (STEP), or PTPN11, thought to be involved in modulation of synaptic plasticity. Dysregulation in its catalytic activity or expression levels is the basis of neurological disorders such as Alzheimer's disease, schizophrenia, fragile X syndrome, Huntington's disease, stroke/ischemia, drug abuse and inflammatory pain⁵³.

The cases mentioned above are a good illustration of the pleiotropy and significance of tyrosine phosphatase function in cell physiology and do not compose an exhaustive list.

1.5 PTPN3

The human protein tyrosine phosphatase non-receptor type 3 (PTPN3, also known as PTPH1) was first discovered by Nicholas Tonks in 1991, by amplification of PTP-related cDNAs from a HeLa cell library using primers corresponding to conserved sequences in PTP catalytic domains⁵⁴. The PTPN3 gene is located in chromosome 9q31⁵⁵. In 1995, Zhang and collaborators expressed and purified it for the first time from insect cells, and demonstrated its tyrosine-dephosphorylating activity⁵⁶.

1.5.1 Domain organization of PTPN3

PTPN3 is a classical PTP and forms the NT5 subtype, together with its homologue protein tyrosine phosphatase non-receptor type 4 (PTPN4, or PTP-MEG), which was previously studied by our team. These two proteins have the same modular organization comprising a N-terminal FERM domain, a central PDZ (PSD-95, Dlg1, ZO-1) domain and a

C-terminal PTP domain (figure 1.6). PTPN3 and PTPN4 share 54% of global sequence identity, with 64% of identity for their PTP domains, 71% for their PDZ domains, and 55% for their FERM domains.

The full-length PTPN3 (UniProt P26045-1) has 913 residues and a molecular mass of 104 kDa. Two other isoforms (UniProt P26045-2 and P26045-3) result from alternative splicing and have incomplete FERM domains (figure 1.6). PTPN3 is a cytosolic protein, enriched at the plasma membrane⁵⁷.

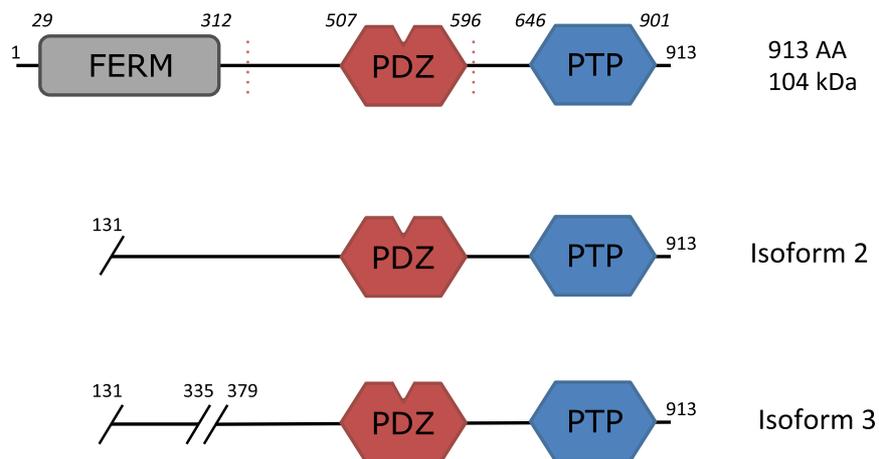


Figure 1.6: Modular architecture of PTPN3. Schematic representation of the PTPN3 isoforms. Numbers on both extremities indicate the boundary residues for each isoform. Numbers in italic above the schematic construct of full-length PTPN3 correspond to the boundary residues of each protein domain. The dotted red lines represent the cleavage sites of trypsin, and the diagonal lines represent the splice sites of the alternative isoforms.

1.5.2 PTPN3 phosphorylation sites

Phosphorylation is the only posttranslational modification that has been reported for PTPN3. In Sf9 insect cells and A431 and 293 mammalian cells, PTPN3 is phosphorylated exclusively on serine by unknown kinases, in multiple sites⁵⁸. *In vitro*, PTPN3 is phosphorylated mostly in serine residues, but also in threonine, by protein kinase A, protein kinase C (PKC), and the MAP kinase 1⁵⁶. Phosphorylation with PKC occurred in a stoichiometry of 2 mol/mol, and resulted in a 25% increase of activity by enhancing the affinity for the substrate⁵⁶. A more recent study by Hou and colleagues identified the mitogen-activated protein kinase (MAPK) p38 γ as the kinase responsible for activation of PTPN3 by phosphorylation of its S459 both *in vitro* and *in vivo*⁵⁹. Some relevant known phosphorylation sites in PTPN3 are shown in figure 1.7 and table 1.3.

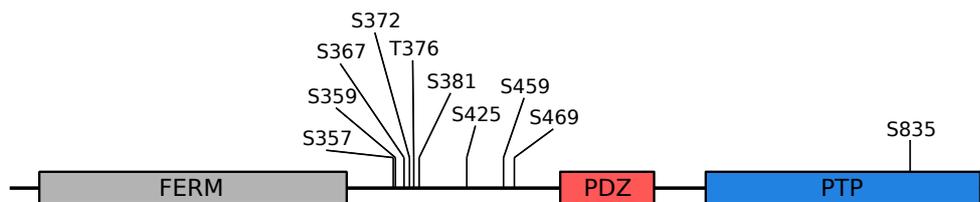


Figure 1.7: PTPN3 phosphorylation sites. Relevant phosphorylation sites are indicated with their respective position. Data obtained from PhosphoSite (www.phosphosite.org).

Table 1.3: PTPN3 phosphorylation sites. The phosphorylated residue in each case is underlined. The phosphorylations shown have been selected either because they have a known biological effect, or because they have been identified in at least 5 studies, including high- and low-throughput. Data obtained from PhosphoSite (www.phosphosite.org).

Sites	Flanking sequence	Known effects on biological processes
S357	WNPAMRRSL <u>S</u> VEHLE	
S359	PAMRRSL <u>S</u> VEHLETK	Regulation of molecular association ⁵⁸
S367	VEHLETK <u>S</u> LPSRSPP	
S372	TKSLPSR <u>S</u> PPITPNW	
T376	PSRSPPI <u>T</u> PNWRSPR	
S381	PITPNWR <u>S</u> PRLRHEI	
S425	GSLAPQD <u>S</u> DSEVSQN	
S459	QKSSSSV <u>S</u> PSSNAPG	Induced cell growth and carcinogenesis ^{59,60}
S469	SNAPGSC <u>S</u> PDGVDQQ	
S835	VRSLRVD <u>S</u> EPVLVHC	Regulation of molecular association ⁵⁸

1.5.3 PTPN3 cleavage sites

PTPN3 is more active *in vitro* after the proteolytic cleavage of its N-terminal segment. Treatment with trypsin cleaves the protein at the beginning of the FERM-PDZ linker, releasing a C-terminal fragment of around 50 kDa, corresponding to the PDZ-PTP bidomain. Further treatment releases a fragment of 39 kDa that corresponds to the isolated PTP domain, starting at residue 601 (the trypsin cleavage sites are indicated as red dotted lines in figure 1.6). The apparition of this PTP fragment coincides with a 5–10 fold increase of the phosphatase activity towards a synthetic substrate. These observations indicate that the N-terminal region of the protein negatively regulates the catalytic activity⁵⁶, and it suggests that the two alternatively spliced isoforms could be more active. It was also shown that the activity of the

full-length protein towards a synthetic substrate is stimulated *in vitro* by increasing ionic strength or addition of positively charged species such as MBP, suggesting that the N- and C-terminal segments of PTPN3 could interact through electrostatic interactions⁵⁶.

1.5.4 14-3-3 recognition sites

14-3-3s are a family of highly conserved signalling proteins that are involved in cell cycle control, transformation, mitogenic signaling, apoptosis, and learning⁶¹. Typically, they recognize different high-affinity binding short linear motifs containing phosphoserine. PTPN3 has been shown to interact *in vitro* and by a yeast two-hybrid screening with the protein 14-3-3 β , in a phosphorylation-dependent way⁵⁸. PTPN3 possesses two 14-3-3-binding sites identified by *in vitro* binding assays, RSLS³⁵⁹VE in the FERM-PDZ linker, and RVDS⁸⁵³EP in the PTP domain. These associations had no effect on the catalytic activity of PTPN3⁵⁸. 14-3-3s are adaptor proteins that could mediate the interaction of PTPN3 with other partners, and these interactions could be regulated by the serine phosphorylation state of PTPN3. Further research is needed to clarify the role of this association in cell physiology.

1.5.5 Structural features of PTPN3

1.5.5.1 The FERM domain

FERM domains are found in numerous membrane-associated signalling and cytoskeletal proteins. They have a molecular mass of around 30 kDa and are formed by three lobes, A, B, and C, which are arranged as the leaves of a clover (figure 1.8). These lobes are associated with ligand binding activities and confer to the FERM domain a varied potential of interaction, notably with phosphatidylinositol 4,5-bisphosphate, phosphatidylserine, calmodulin and p55^{62,63}. FERM domains are responsible for localizing proteins at the interface between the plasma membrane and the cytoskeleton by binding the cytosolic tails of transmembrane proteins, or by binding phosphoinositides⁶³. It has been shown that the FERM domain is determinant in the cellular localization of PTPN3 and PTPN4. In Jurkat cells, both phosphatases are found in the cytosol and enriched at the plasma membrane. Upon suppression of the FERM domain, the proteins become exclusively cytoplasmic (figure 1.9)⁵⁷. A similar distribution was observed for full-length PTPN3 in mouse neurons⁶⁴.

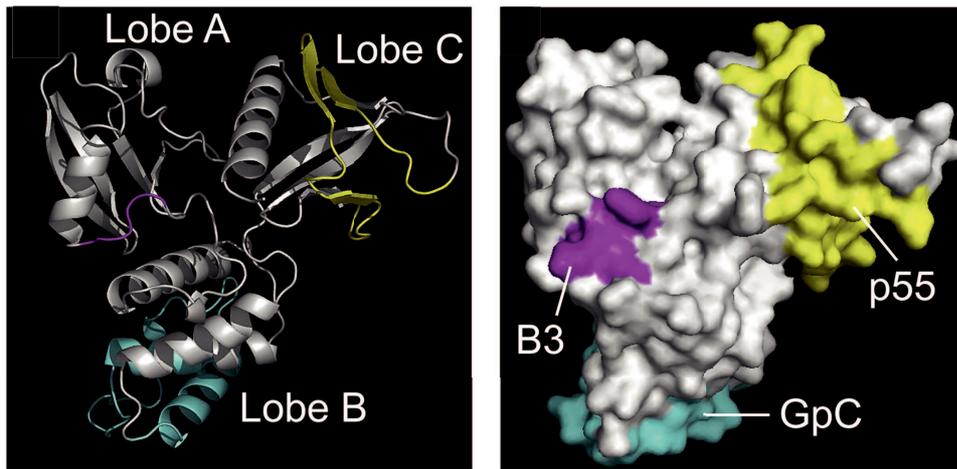


Figure 1.8: Structural organization of the FERM domain. Left: Cartoon representation of the fold of the FERM domain of the 4.1R protein (PDB ID 1GG3). The three lobes are indicated (A, B, C). Right: Space filling model of the FERM domain showing binding sites for transmembrane proteins Band 3 (B3), glycophorin C (GpC) and p55. The colours represent the same sequences. Adapted from Baines *et al.* 2014⁶³.

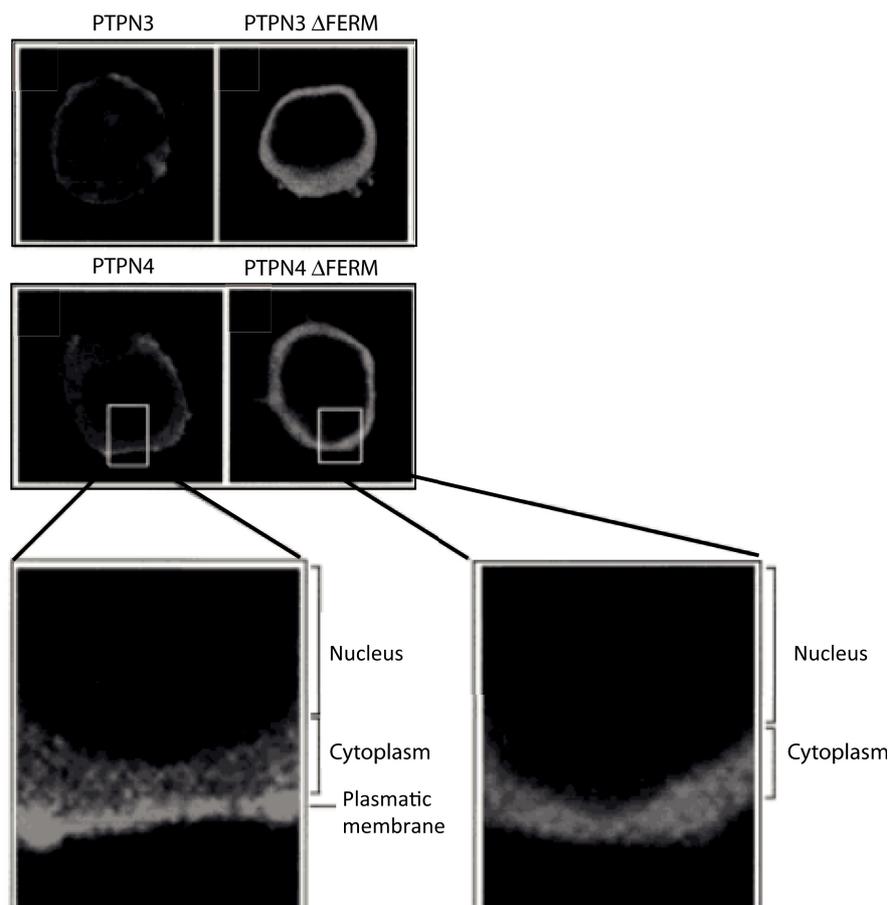


Figure 1.9: Subcellular location of PTPN3 and PTPN4 in Jurkat T cells. Confocal microscopy of Jurkat cells transfected with pEF/HA expression plasmids encoding the indicated phosphatases and stained with a FITC-conjugated anti-HA mAb. The selected regions are enlarged to highlight the difference between PTPN4 and PTPN4 Δ FERM. Adapted from Gjørloff-Wingren *et al.* 2000⁵⁷.

The FERM domain has been proposed to play a regulatory role in PTPN4, inhibiting catalytic activity by an unknown mechanism⁶⁵, although this was not observed for PTPN3. Besides PTPN3 and PTPN4, FERM domains are also found in other NTPTPs: PTPN13, PTPN21 and PTP14.

1.5.5.2 The PDZ domain

PDZ domains are protein-protein interaction modules that bind short linear motifs usually located at the extremity of the C-terminus of protein partners. These domains have a central role in the assembly of signalling complexes at cellular membranes. The PDZ domain of PTPN3 (PTPN3-PDZ) mediates its interactions with cellular partner proteins containing a PDZ-binding motif (PBM), such as the MAP kinase p38 γ ⁶⁶. Additionally, PTPN3-PDZ is targeted by PBM-containing proteins of the oncogenic viruses human papillomavirus (HPV) 16 and 18, and hepatitis B Virus (HBV)^{67,68}. However, at the start of this project, the structure of this domain was not solved, and the structural basis of its cellular and viral interactions remained unknown. PDZ domains are the topic of chapter II, where they will be discussed in detail.

1.5.5.3 The PTP domain

The catalytic domain of PTPN3 (PTPN3-PTP) is a classical, cysteine-based tyrosine phosphatase domain with a classic PTP fold (figure 1.10). Its structure was first solved in 2009 in the free form in the frame of the Structural Genomics Consortium (SGC)⁶⁹. Later, its structure was determined using a catalytically inactive mutant (D118A/C842S) in complex with a phosphopeptide derived from its substrate p38 γ (the substrates and ligands of PTPN3 are detailed below)⁷⁰. The kinase p38 γ is activated by dual phosphorylation in a Thr-Gly-Tyr (TGY) motif located in positions 183–185 of its activation loop, which can be dephosphorylated by PTPN3⁶⁶. Kinetics data showed that PTPN3 displays increased K_M for the dually phosphorylated (pT¹⁸³-G-pY¹⁸⁵) rather than for monophosphorylated (T-G-pY¹⁸⁵) p38 γ peptide as a substrate⁷⁰. This preference of PTPN3 for dually phosphorylated p38 γ as a substrate is explained by a salt bridge between the phosphate oxygen of pT¹⁸³ and the nitrogen atoms of the guanidine group of R⁷⁵¹ in the crystal structure of the complex with a dually phosphorylated peptide that mimics the p38 γ activation loop. The structure of the complex is in good agreement with the structure of the free PTP domain, with the pY¹⁸⁵ of the peptide inserted into the catalytic cleft and surrounded by the substrate-binding loops. The major

difference was found in the WPD loop, which switched from an open to a closed conformation upon complex formation (figure 1.10). Moreover, mutation of R751 to serine disrupts this interaction, and the K_M of the interaction is similar to the one of the wild type PTPN3-PTP for the monophosphorylated p38 γ peptide⁷⁰. This position is also conserved in PTPN4. Specific recognition of dually phosphorylated p38 γ has not been reported for other phosphatases capable of dephosphorylating MAPKs, such as DSPs or PTPs with ‘kinase interaction motifs’ (specific motifs that determine recognition of kinases)^{71,72}.

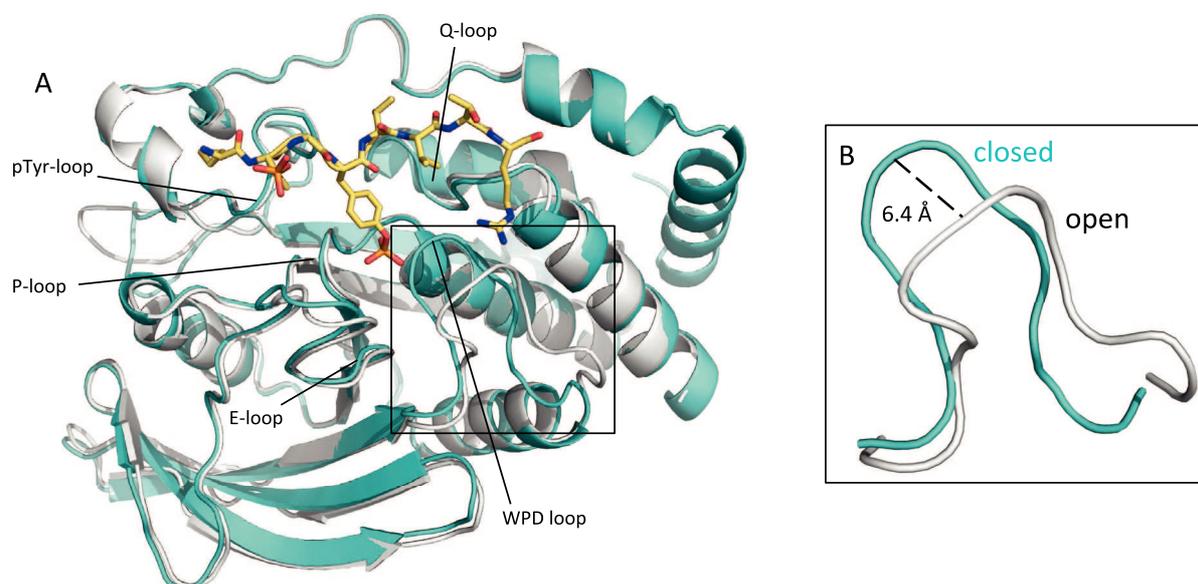


Figure 1.10: Comparison of the structures of the unbound, wild type PTPN3-PTP and its substrate-trapping mutant PTPN3-PTPD118A/C842S in complex with a phosphopeptide. (A) The phosphopeptide complex (cyan cartoon and yellow sticks, PDB ID 4QUM) superimposed with the wild-type PTPN3-PTP structure (white ribbon, PDB ID 2B49). **(B)** Enlarged view showing the open and closed conformations of the WPD loop. The measure shows the distance between the C α of H812 in each case.

The reactive cysteine of the catalytic site of phosphatases is susceptible to reversible oxidation by reactive oxygen species (ROS). Oxidation of the catalytic cysteine has an inhibitory effect upon the activity of PTPs and is becoming an increasingly appreciated regulatory mechanism of PTP catalysis⁷³. PTPN3 has been shown to be sensitive to oxidation by H₂O₂ in NIH3T3 (mouse embryo fibroblast) cells, albeit significantly less than PTPN4, and insensitive to treatment with the inductor of oxidative stress BSO⁷⁴. PTP activity can also be regulated by the action of nitric oxide, which causes reversible nitrosylation of their catalytic

cysteines. PTPN3 appeared to be insensitive to exposure to different nitrosylating agents in NIH3T3 cells, while PTPN4 was more susceptible to being nitrosylated⁷⁴. These differences in reactivity of PTPN3 and PTPN4 could be significant in the role of the phosphatases in signalling, although currently there is a lack of data on this subject. Cancer cells often produce an increased level of ROS and have unique PTP oxidation profiles. Similarly, multiple cancer cells and tissues show upregulated nitric oxide production⁷⁵. These alterations affect the activity of PTPs and can modulate the balance of tyrosine phosphorylation in the cell⁷⁴.

1.5.6 PTPN3 functions

PTPN3 is a signalling protein that has been implicated in regulation of cell growth and homeostasis^{60,76}. Its FERM and PDZ domains account for specificity in its functions: the FERM domain targets the protein to the plasma membrane providing spatial regulation, while the PDZ domain mediates interactions with specific partners and substrates, or anchors the phosphatase in multi-protein signalling complexes. However, its precise role in cell signalling has not yet been clearly established. In the following, ligands and substrates of PTPN3 are reported to exemplify the known functions of this protein.

1.5.6.1 Epidermal growth factor and estrogen signalling

The epidermal growth factor receptor (EGFR) is a cell surface receptor that regulates cellular processes including growth, proliferation and differentiation. EGFR ligand binding leads to auto-phosphorylation on cytosolic tyrosine residues and subsequent activation of downstream signalling pathways such as Ras/MAPKs and PIK3/AKT⁷⁷. Overexpression or constitutive activation of EGFR is found in breast, colon, pancreas, lung, and brain cancers. Several tyrosine phosphatases are known to regulate EGFR signalling, either by directly dephosphorylating EGFR^{36,78-81}, or by determining its cellular location^{82,83}. PTPN3 has been shown to play different roles in modulating EGFR function, which are reported below.

In breast cancer, there is a relationship between EGFR, estrogen receptor α (ER), and PTPN3. ER is expressed in about 70% of breast cancers, and promotes cancer growth upon activation by estrogen and growth factors. ER is the target of anti-estrogens such as tamoxifen (TAM), but about one quarter of ER-positive breast cancers are resistant to anti-hormone therapy⁸⁴. Phosphorylation of ER tyrosine 537 by the Src kinase regulates ER activity and decreases TAM sensitivity⁸⁵⁻⁸⁷. ER can bind EGFR, forming a complex that inhibits the

effects of anti-hormone therapy^{88,89}. PTPN3 was found to dephosphorylate ER/Y537 *in vitro* and in breast cancer cells, resulting in increased ER turnover and nuclear accumulation, as well as increased breast cancer sensitivity to anti-estrogen treatment in cells and in a mouse xenograft model⁹⁰. PTPN3 disrupts the ER-EGFR complex by dephosphorylation of EGFR/Y1173, increasing the sensitivity of breast cancer cells to EGFR-targeted therapies with tyrosine kinase inhibitors (TKIs) and stabilizing the inactive form of EGFR⁶⁰. Phosphorylation by p38 γ of the S459 of PTPN3, located in the linker between the FERM and PDZ domains (figure 1.7 PTPN3 phosphorylation sites), was required for this effect⁵⁹. Although phosphorylation of S459 did not alter the activity of PTPN3 towards p38 γ , it was required for EGFR dephosphorylation, indicating that this modification affects the interaction with EGFR. On the other hand, triple-negative breast cancers (TNBC) lack ER, progesterone receptor and human epidermal growth factor 2 receptor. They account for 15–20% of breast cancers, and represent the most aggressive subtype⁹¹. Diluvio and collaborators recently found that depletion of NOTCH3 in TKI-resistant TNBC cells induces downregulation of EGFR surface expression by promoting its dephosphorylation by PTPN3, increasing the sensitivity of the cells to TKI treatment⁹². In lung cancer cells, PTPN3 dephosphorylates the EGFR substrate Eps15, promoting the lysosomal degradation of activated EGFR and subsequently downregulating EGFR signalling⁹³. Notably, in K-Ras mutant colon cancer cells, EGFR dephosphorylation by PTPN3 confers resistance to TKIs⁹⁴, which highlights the cell type-dependent specificity of PTPN3 effects.

Thus, the destabilization of activated EGFR or the EGFR-ER complex by PTPN3-mediated dephosphorylation is a potentially attractive approach to enhance the sensitivity of breast cancer to therapies combining TKIs and TAM⁶⁰. However, it is clear that the role of PTPN3 in cancer is context-dependent, as shown by the opposite effect of EGFR dephosphorylation on colon cancer cells.

1.5.6.2 Growth hormone signalling

The growth hormone receptor (GHR) is a transmembrane protein that activates intra- and intercellular growth-promoting signal transduction pathways in response to growth hormone binding. The action of the GHR is mediated by the Janus kinase 2 (JAK-2), which is activated upon growth hormone binding to GHR. JAK-2 phosphorylates multiple tyrosines on the cytoplasmic intracellular domain of the receptor, resulting in subsequent binding and

activation of signal transducer and activator of transcription (STAT) proteins, which deliver cell proliferation responses⁹⁵.

Screening with substrate-trapping mutants showed that PTPN3 recognizes the growth hormone receptor as a substrate⁹⁶, and has been proposed to participate in GHR signalling modulation *in vivo*⁹⁷. PTPN3 knockout mice showed increased body size and expression of insulin-like growth factor 1 mRNA in plasma and liver. This is consistent with an increased sensitivity to growth hormone, and points towards an involvement of PTPN3 in regulation of growth⁹⁷.

1.5.6.3 Cell cycle regulation and transitional endoplasmic reticulum assembly

Another substrate of PTPN3 is the ATPase valosin-containing protein (VCP), also known as p97. VCP is involved in many cellular processes, including maintenance of protein homeostasis^{98,99}, modulation of signalling pathways^{100,101}, activation of transcription factors¹⁰², cell cycle progression¹⁰³ and response to DNA damage^{104,105}.

VCP was shown *in vitro* and *in vivo* to be a substrate of PTPN3. Ectopic PTPN3 expression in NIH3T3 cells inhibited cell cycle progression, coinciding with VCP dephosphorylation⁷⁶. Further studies linked PTPN3-dependent VCP dephosphorylation to a stabilization of the association of VCP with membranes, which promoted the dynamic assembly of transitional endoplasmic reticulum (tER). This suggests that PTPN3 could be a link between signal transduction events and the early secretory pathway¹⁰⁶. Interestingly, it was also found that the kinase that opposes the action of PTPN3 and inhibits tER assembly is JAK-2¹⁰⁶, the same kinase that mediates GHR-dependent signal transduction events.

1.5.6.4 T cell activation

T cell activation is a critical process for initiation and regulation of the immune response. This is a complex event that requires two associated signals. The first one is an antigen-specific signal triggered by binding of an antigenic peptide to the T cell receptor (TCR), and the second one is mediated by cytokines or other co-stimulatory molecules¹⁰⁷. The simultaneous occurrence of these events initiates an intra-cellular signalling cascade that leads to the development of immune mechanisms mediated by cytotoxic T cells, engagement of other immune cells such as macrophages, and increase of antibody effector function and antibody production¹¹.

Experiments performed in Jurkat T cells implicated PTPN3 as a negative modulator of T cell activation, strongly inhibiting signal transduction from the TCR¹⁰⁸. Moreover, *in vitro* experiments showed that PTPN3 dephosphorylates the TCR ζ subunit¹⁰⁹. However, *in vivo* observations were not consistent, as mice expressing catalytically inactive PTPN3 exhibited normal numbers and ratios of T cells, and grew normally without any signs of spontaneous T cell activation. Moreover, TCR-dependent signal transduction and cytokine production were unaffected by the lack of PTPN3 catalytic activity¹¹⁰. One possibility was that the apparent absence of function of PTPN3 in T cells could be due to redundancy with the functions of the FERM-containing phosphatases PTPN4 or PTPN13. To rule this out, Bauler and collaborators generated PTPN3/PTPN4 double mutant and PTPN3/PTPN4/PTPN13 triple mutant mice¹¹¹. They observed in both cases that the mutant mice exhibited normal T cell development, homeostasis and differentiation, as well a normal cytokine secretion and proliferation, showing that both PTPN3 and PTPN4 are dispensable for TCR signal transduction *in vivo*¹¹¹.

1.5.6.5 Tumor necrosis factor α -convertase regulation

The tumor necrosis factor α (TNF- α)-convertase (TACE, or ADAM17) is a metalloprotease involved in the proteolytic cleavage and subsequent release of extracellular domains (ectodomains) of transmembrane proteins, a process known as ectodomain shedding. This cleavage is a fundamental mechanism to control abundance, activation and inactivation of membrane proteins¹¹². TACE was first identified as the responsible of the release of the inflammatory cytokine TNF- α ¹¹³, and its current list of over 80 substrates includes cell adhesion proteins, cytokines, growth factors and their respective receptors¹¹⁴, implicating this protein in many pathophysiological processes and thus making it an attractive therapeutic target.

The interaction between PTPN3 and TACE was first detected in a yeast two-hybrid screening and validated using *in vitro* and *in vivo* methodologies¹¹⁵. The interaction occurs through a sequence in the C-terminus of TACE that targets PTPN3-PDZ. Interestingly, it was observed that TACE protein levels were downregulated when co-expressed with catalytically active forms of PTPN3, compared to phosphatase inactive mutants. Furthermore, PTPN3 activity decreased the TNF- α shedding activity of TACE in COS-7 cells. In agreement with these results, other studies indicate a possible role of phosphorylation of the cytoplasmic domain of TACE in regulation of its sheddase activity¹¹⁶⁻¹¹⁹, although this is still not fully

understood. Taken together, all these observations point towards an implication of PTPN3 in modulation of TACE-mediated ectodomain shedding.

1.5.6.6 14-3-3 β binding

14-3-3 proteins are small and highly conserved adaptor proteins that function as signalling hubs that regulate cellular processes such as signal transduction, cell cycle regulation, protein trafficking and apoptosis⁶¹.

PTPN3 was identified *in vitro* and by co-expression experiments as a partner of 14-3-3 β . 14-3-3 β targets two different binding sites in PTPN3, which contain serine residues that are phosphorylated *in vivo*⁵⁸ (see section 1.5.4). Although no functional effects of this association were evident on PTPN3 activity assays, it is possible that 14-3-3 β behaves as a cofactor that recruits additional regulators⁵⁸, or as a linker between PTPN3 and some of its cellular ligands to promote their dephosphorylation.

1.5.6.7 Cardiac voltage-gated sodium channel function modulation

Cardiac voltage-gated sodium channels (Nav1.5) mediate membrane depolarization and propagation of an electrical action potential that enables the heart to contract. Function and regulation of these channels are critical for the regularity of the contractions and optimal functioning of the heart. These channels are composed by α and β subunits. ‘Primary’ α subunits are large transmembrane proteins that form the pore and control gating and ion selection through voltage-sensing domains. The smaller ‘auxiliary’ β subunits have a regulatory role in gating, expression and oligomerization of channels¹²⁰. α subunits possess long C-terminal domains (NavCTD) that contain a series of protein-protein interaction motifs that allow interaction with cytosolic partners. The last residues of this domain comprise a PBM sequence that targets PDZ domain-containing proteins, such as the adaptor protein syntrophin¹²¹.

PTPN3 interacts with the NavCTD, and this interaction is dependent on the PDZ domain-binding motif. Moreover, it appears that the channel may be a substrate of PTPN3, and that tyrosine phosphorylation could have an effect in the functioning of the channel. When co-expressed with Nav1.5 in HEK293 cells, PTPN3 caused a shift of the availability curve towards more negative potentials¹²², which suggests that tyrosine dephosphorylation could stabilize the inactivated state of the channel, decreasing its excitability. Although PTPN3-dependent tyrosine dephosphorylation of the channel was not directly evaluated, the results

are consistent with previous observations that phosphorylation of Na_v1.5 by the Src family tyrosine kinase Fyn destabilizes the inactivated state, thus accelerating recovery from inactivation¹²³.

1.5.6.8 Ras transformation

MAPKs are serine/threonine kinases that are involved in regulation of crucial cellular functions, including proliferation, survival and apoptosis. A major MAPK family are the p38s, composed of 5 different isoforms: p38 α , p38 β 1, p38 β 2, p38 γ , p38 δ . p38 MAPKs are activated in response to extracellular stimuli such as pathogens or cytokines, by double phosphorylation in threonine and tyrosine residues¹²⁴. The p38 γ isoform (*MAPK12*) differs from the other members of the family in that it contains a C-terminal PBM, and is thus able to interact with PDZ-containing proteins such as α 1-syntrophin, PSD95 and hDlg¹²⁵⁻¹²⁷. Additionally, p38 γ is the only p38 whose expression is induced during cell differentiation and Ras transformation^{128,129}.

PTPN3 binds p38 γ in rat intestinal IEC-6 cells through a PDZ/PBM interaction. As a result of this interaction, PTPN3 dephosphorylates and inactivates p38 γ , which was observed to promote Ras transformation⁶⁶. This will be further developed in the next section.

1.5.6.9 Cognition and inflammatory response

Patrignani and colleagues examined the contribution of PTPN3 in the inflammatory response and in cognition using PTPN3 knockout (PTPN3-KO) mice^{64,130}. First, they found that PTPN3 is expressed in neural populations in the adult mouse brain that are involved in locomotor and cognitive functions. Their observations highlighted a gender bias in behavioural tests, where male PTPN3-KO mice displayed better short-term memory than the male WT mice. Female PTPN3-KO mice exhibited delayed learning and coordination capacities. They proposed that these effects could be related to the involvement of PTPN3 in growth hormone signalling⁶⁴. Indeed, GHR is highly expressed in the central nervous system, and the growth hormone axis has been shown to affect cognitive functions¹³¹. Interestingly, studies in *Drosophila* demonstrated a role of the PTPN3 and PTPN4 homolog Ptpmeg in the establishment and maintenance of normal axonal projections in the *Drosophila* central brain¹³². These observations highlight the relevance of PTPN3 in brain function and physiology.

Later, it was found that PTPN3-KO mice exhibited a slightly increased perception of inflammatory pain. In early phases of lipopolysaccharide-induced inflammation, the knockout mice exhibited some decrease in TNF α expression, and a significant delayed cytokine release, which is in agreement with the inhibitory role of PTPN3 on TACE activity. All together, their observations pointed towards a potential role of PTPN3 in the regulation of LPS-induced cytokine release *in vivo*¹³⁰, which contrasts with previous observations that PTPN3 and also PTPN4 are dispensable for the normal immune response¹¹¹.

1.5.7 PTPN3 in cancer

Altered function or expression of PTPN3 has been associated to many forms of cancer, either as an oncogenic protein or as a tumor suppressor, depending on the cellular context (table 1.4). Here, I will describe the known links between PTPN3 and cancer progression or suppression.

1.5.7.1 PTPN3 as an oncogenic protein

Increased levels of PTPN3 expression could favour oncogenic phenotypes. In ovarian cancer, overexpression of PTPN3 favours drug resistance, stemness and tumorigenicity¹³³. In breast cancer, PTPN3 promotes tumorigenicity through cytoplasmic accumulation and stabilization of vitamin D receptor, independently of its phosphatase activity. Upregulation of PTPN3 expression was found in almost 50% of the breast cancer samples studied, and its expression levels showed correlation with clinical metastasis¹³⁴. This drastically contrasts with the previously mentioned capacity of PTPN3 to enhance breast cancer sensitivity to therapies by modulating EGFR activity. PTPN3 expression levels were also correlated to the differentiation grade of gastric adenocarcinoma¹³⁵. In esophageal cancer lesions, PTPN3 mRNA levels were upregulated in comparison to normal mucosa, and the expression showed positive correlation with the differentiation grade of the cancer¹³⁶. PTPN3 is also linked to brain tumors. In glioblastoma it is proposed to promote cell proliferation and invasion, and high expression levels are associated with poor prognosis¹³⁷. In glioma, the most frequent form of brain tumor, PTPN3 expression is increased, and this is associated with tumor growth and metastasis. Depletion of PTPN3 in glioma cell lines caused the cell cycle to arrest in the S phase and limited tumor growth in mice¹³⁸. I previously mentioned that the p38 γ MAPK is a substrate of PTPN3, and that these proteins cooperate to enhance Ras-dependent oncogenesis. Indeed, the expression of p38 γ is induced by K-Ras (the oncogenic form of Ras) without

increasing its phosphorylation levels, and non-phosphorylated p38 γ promotes K-Ras oncogenesis¹³⁹. Furthermore, PTPN3 is phosphorylated by p38 γ dependently on their PDZ-PBM interaction, which also increases K-Ras oncogenesis⁵⁹. Notably, PTPN3 and p38 γ protein expression is increased in primary colon cancer tissues⁶⁶. In addition to overexpression, increased PTPN3 activity due to activating mutations was shown to be a factor promoting tumor progression and postoperative tumor recurrence in intrahepatic cholangiocarcinoma, a common form of liver cancer. Notably, the mutations were found in or near the FERM domain, not directly affecting the PTP and PDZ domains¹⁴⁰.

1.5.7.2 PTPN3 as a tumor suppressor

PTPN3 can also function as a tumor suppressor. In hepatocellular carcinoma (HCC), PTPN3 expression is decreased by epigenetic regulation, showing an inverse correlation with the aggressiveness of the tumor. This loss of PTPN3 favours cell migration and invasion both *in vitro* and in lung metastasis¹⁴¹. As previously described, one of the known physiological functions of PTPN3 is to promote degradation of EGFR by dephosphorylation of its substrate Eps15. In line with this, silencing of EGFR expression counteracts the pro-metastatic effect caused by PTPN3 downregulation, suggesting that PTPN3 achieves its role as a tumor growth and metastasis inhibitor in HCC by modulating EGFR signalling¹⁴¹. Endocytic degradation of EGFR seems to be as well the mechanism by which PTPN3 prevents cell proliferation and migration in non-small cell lung cancer (NSCLC)⁹³. Additionally, PTPN3 tumor suppressor function can also be affected by mutations. In a NSCLC cell line, PTPN3 was mutated by a translocation event resulting in a null PTPN3 allele. Re-expression of functional PTPN3 in NSCLC cell lines decreased cell survival, supporting its tumor-suppressive role¹⁴². A study on the mutational state of tyrosine phosphatase genes in colorectal cancers revealed a number of somatic mutations in *PTPN3*, almost all of which were missense and located in or near the FERM domain³¹. Mutations in the FERM domain could abolish PTPN3 tumor-suppressing role by either altering its subcellular localization or disrupting its normal interactions. In liver cancer, altered interactions were reported with a leucine-to-arginine substitution at position 232 in PTPN3 (PTPN3^{L232R}), frequently observed in intrahepatic cholangiocarcinoma. In normal conditions, PTPN3 enhances tumor growth factor β (TGF β) signalling by a mechanism that is independent of its phosphatase activity. PTPN3 binds the TGF β type I receptor (T β RI) through its FERM domain, and prevents its proteasomal degradation by blocking the binding of E3 ligases. PTPN3^{L232R} still interacts with T β RI but is unable to

prevent the binding of the ligases, and thus the TGF β -induced transcriptional responses that normally inhibit growth are attenuated¹⁴³.

Table 1.4: Implications of PTPN3 in cancer.

Role	Cancer type	Observations	References
Oncogenic	Ovarian	Overexpression favours drug resistance, stemness and tumorigenicity	133
	Breast	Promotes tumorigenicity through cytoplasmic accumulation and stabilization of vitamin D receptor, independently of its phosphatase activity. Increased expression levels correlated with clinical metastasis	134
	Gastric adenocarcinoma	Expression levels correlate to grade of differentiation of the cancer	135
	Esophageal	Expression levels correlate to grade of differentiation of the cancer	136
	Glioblastoma	Promotes cell proliferation and invasion. High expression levels associated with poor prognosis	137
	Glioma	Increased expression associated with tumor growth and metastasis	138
	Intrahepatic cholangiocarcinoma	Increased activity due to activating mutations promotes tumor progression and postoperative tumor recurrence	140
	Colon cancer	Dephosphorylates and cooperates with p38 γ to promote K-Ras oncogenesis	139
Tumor-suppressor	Hepatocellular carcinoma	Expression is decreased by epigenetic regulation, showing an inverse correlation with the aggressiveness of the tumor. PTPN3 loss favours cell migration and migration	141
	Non-small cell lung cancer	Prevents cell proliferation and migration by promoting EGFR degradation, and attenuating Src activity and Src-mediated DAAM1 activation	93,142,144
	Colorectal cancer	Frequently presents somatic missense mutations located in or near the FERM domain	31

2 PDZ domains

2.1 Role of PDZ domains in the cell

PDZ domains are protein-protein interaction modules that were discovered in the early 1990s. The acronym PDZ derives from the first three proteins in which they were initially identified – the postsynaptic density protein 95 (PSD-95), the *Drosophila* homologous Discs large protein (Dlg), and the mammalian tight junction protein zonula occludens 1 (ZO-1)^{145,146}. These are globular domains of about 90 residues. Their best-known function is the recognition of short C-terminal linear motifs known as PBMs in partner proteins. However, some PDZ domains can also bind internal motifs, lipids, and other PDZ domains¹⁴⁷. PDZ domains are highly prevalent in metazoan proteins – there are 152 PDZ-containing proteins in the human proteome, containing 266 different PDZ domains^{148,149}. The widespread occurrence of PDZ domains in metazoans and their absence in other kingdoms of life has led to the proposal that PDZ domains might have evolved together with multicellularity¹⁵⁰.

PDZ domains are often found in multiple copies in proteins. In many cases, they are found in combination with other protein-protein interaction domains, catalytic domains, or cellular localization domains (figure 2.1). Their high frequency of occurrence combined with their high promiscuity (second after PH domains) situates them at the core of cellular signalling networks¹⁵¹, where they play a key role in the assembly and function of multi-protein signal transduction complexes. PDZ proteins have been implicated in processes that include the establishment and maintenance of cell polarity and cell-cell junctions, cell migration, visual and auditory processes, signal transduction in neurons and regulation of cell homeostasis (the reactome participation of PDZ proteins is detailed in section 9.2). The following examples illustrate the biological roles of PDZ proteins: the organization of the *Drosophila* phototransduction cascade by INAD¹⁵² and the organization of the postsynaptic density by PSD-95¹⁵².

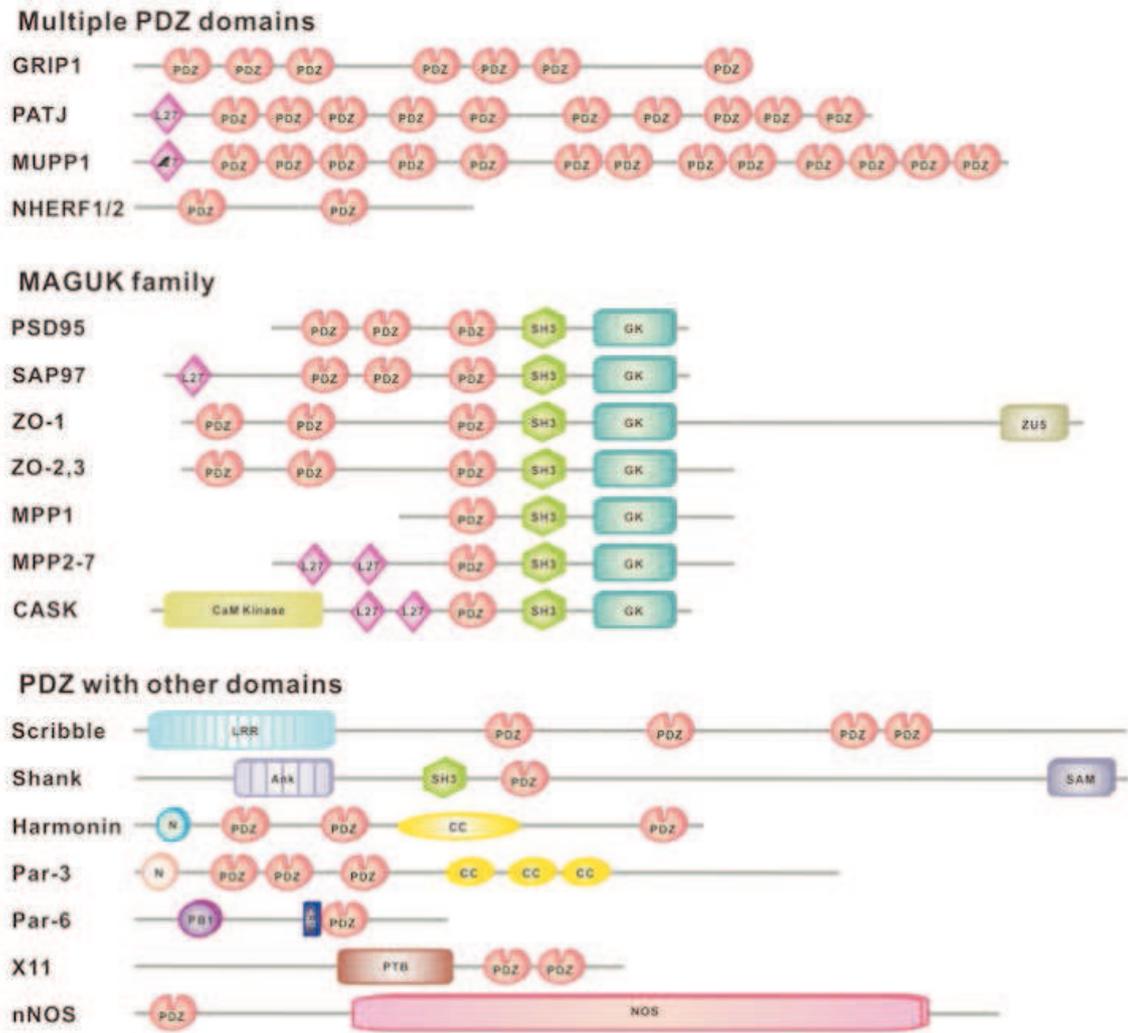


Figure 2.1: Domain organizations of a selected set of PDZ domain proteins. The proteins are grouped according to their domain organization patterns. PDZ domains are shown in orange. Ank, ankyrin repeats; CaM kinase, calmodulin-dependent kinase (CaMK)-like domain; CASK; calmodulin-associated serine/threonine kinase; CC, coiled-coil domain; CRIB, Cdc42/Rac-interactive binding domain; LRR, leucine-rich repeat; MPP, membrane protein, palmitoylated; N, N-terminal domain; Par, partitioning defective homologue; PATJ, PALS1-associated tight junction protein; PB1, Phox and Bem1p domain; SAP97, synapse-associated protein 97; ZU5, domain present in ZO-1 and UNC5-like netrin receptor. Figure from Ye and Zhang 2013¹⁵³.

2.1.1 Phototransduction in rhabdomeres

The phototransduction cascade in *Drosophila* occurs in specialized microvillar structures known as rhabdomeres that project from the photoreceptor cell surface (figure 2.2.A). When light hits the photoreceptor cells, it triggers a series of responses that ultimately open light-activated cation channels (TRP and TRPL), resulting in depolarization of the photoreceptor cell (described in detail by Venkatachalam, Luo, and Montell¹⁵⁴). After this stimulation, the light response is deactivated in a calcium-dependent process that involves a

PKC, calmodulin (CaM), and NINAC (an unconventional myosin III). Activation and deactivation are extremely rapid processes that take less than 100 ms. The close linking of the components of the phototransduction cascade is essential to achieve this level efficiency. The protein scaffold responsible for this multi-protein assembly is INAD, a 5 PDZ domain-containing protein (figure 2.2.B). TRP, PKC, and phospholipase C (PLC) are generally considered to be strict components of the INAD phototransduction complex, and it has been reported that these interactions are mediated by C-terminal and internal PBMs¹⁵². The association pattern with the other signalling molecules has been subject to debate and controversial reports and appears to be dynamic¹⁵⁵. Figure 2.2.B represents the current consensus on how the INAD complex might be assembled¹⁵⁵.

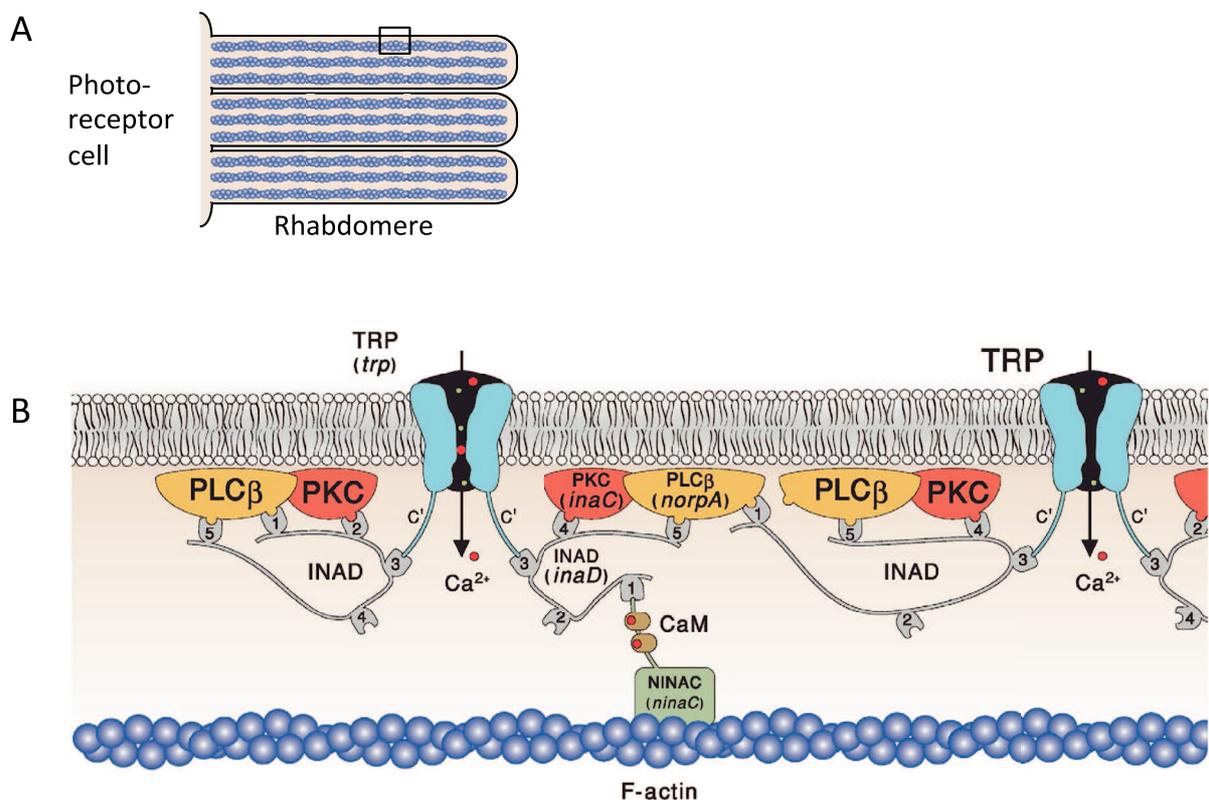


Figure 2.2: Organization of the INAD signaling complex. (A) Schematic of actin-filled microvilli of rhabdomeres protruding from the photoreceptor cell surface. Actin is represented as blue filaments. (B) Detail of the rhabdomere membrane. INAD contains 5 PDZ domains (indicated by numbers 1–5) and identified specific interactions between PKC (INAC, encoded by the *inaC* gene) and PDZ2 (or PDZ4), TRP and PDZ3, and PLC (NORPA, encoded by the *norpA* gene) with PDZ1 and PDZ5. This binding pattern is still in debate due to several contradictory reports. Adapted from Katz and Minke 2018¹⁵⁵.

The major physiological role of INAD is to organise its interacting phototransduction proteins into an integrated signalling unit, increasing the rate of interaction between them and

isolating them from potential interferences. It has been shown by mutational analysis that the capacity to assemble a multiprotein complex directly impacts the efficiency and activation and termination kinetics of the photoresponse¹⁵². In addition, INAD is responsible for the subcellular localization of its three main interacting proteins, TRP, PKC, and PLC below the plasma membrane and in close proximity to each other and to the light-activated ion channels¹⁵⁶. This enables the light response to occur with high efficiency, and promotes the formation of small microdomains where light response can be finely modulated by changes in the levels of intracellular calcium¹⁵⁷. Mutations that prevent the binding of any of these proteins to INAD result in mislocalization of the protein to the cell body. Moreover, INAD-TRP interaction is necessary for anchoring and retention of the signalling complex to the rhabdomere^{156,157}. This example highlights the relevance of PDZ proteins acting as a scaffold to enhance the functions of its partner signalling proteins in specialized cells.

2.1.2 Organization of the postsynaptic density by PSD-95

In excitatory glutamatergic synapses, the postsynaptic membrane is specialized in rapid response to glutamate released from the presynaptic terminal. Underneath the postsynaptic membrane, an electron-dense structure known as the postsynaptic density (PSD) is specialized in receiving, processing and propagating the glutamatergic signal. The PSD is formed by a high concentration of membrane receptors and ion channels, tethered to signalling proteins, cell-adhesion molecules and components of the cytoskeleton by scaffold and adaptor proteins (figure 2.3).

The most abundant scaffolding protein in the PSD is PSD-95 (also known as DLG4), a three PDZ domain-containing member of the membrane-associated guanylate kinase (MAGUK) protein family (see section 5.iii for a description of the MAGUK family). PSD-95 plays a central role in the organization of the PSD, as it binds N-methyl-D-aspartate (NMDA) receptors, the main glutamate receptor subtype in the PSD. NMDA receptors are composed of two subunits, NR1 and NR2, whose cytoplasmic domain contains PBMs that bind to the first two PDZ domains of PSD-95. Other transmembrane proteins, including receptor tyrosine kinases, potassium channels, calcium pumps, and the cytoskeletal protein cysteine-rich PDZ-binding protein (CRIPT) also target the PDZ domains of PSD-95 via C-terminal interactions^{152,158}.

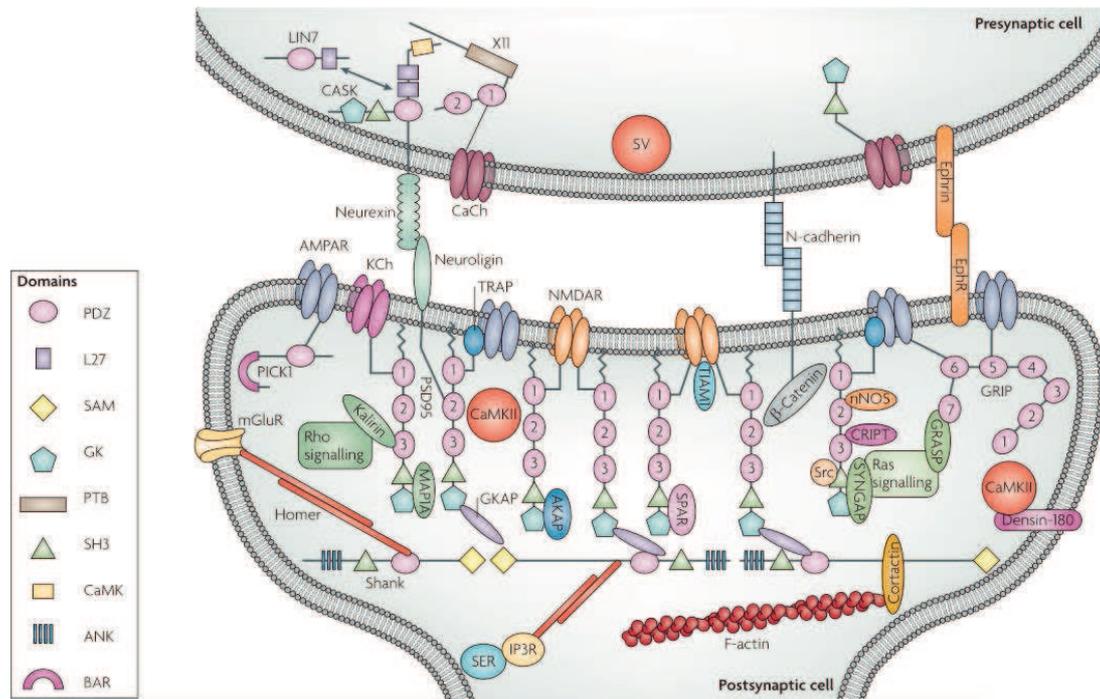


Figure 2.3: Protein complex organization in the postsynaptic density. The postsynaptic density is comprised of membrane receptors and ion channels, scaffold and adaptor proteins, signalling proteins, cell-adhesion molecules and components of the cytoskeleton. Glutamate receptors, such as NMDARs (N-methyl-d-aspartate receptors) and AMPARs (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptors), are located at the postsynaptic membrane. The PDZ-domain-containing scaffold proteins PSD95 and the Src-homology domain 3 (SH3) and multiple ankyrin repeat domains (Shank) family form a two-layer protein network below the postsynaptic membrane, which is bridged by guanylate kinase-associated protein (GKAP). Other signalling molecules occupy the spaces in the PSD95–GKAP–Shank protein web. Shank-family scaffolds are further linked to actin filaments. The domains of PSD95 and Shank (PDZ, SH3, guanylate kinase (GK), sterile-alpha motif (SAM) and ankyrin repeats (ANK) (see key)) are shown; other proteins are represented by simple shapes and are labelled. The presynaptic and postsynaptic membranes are connected by cell-adhesion molecules. AKAP, adenylate-kinase anchoring protein; CaCh, Ca²⁺channel; CaMKII, calcium/calmodulin-dependent protein kinase II; CRIPT, cysteine-rich PDZ-binding protein; EphR, ephrin receptor; GKAP, guanylate kinase-associated protein; GRASP, GRIP-associated protein; GRIP, glutamate receptor interacting protein; IP3R, inositol-1,4,5-trisphosphate receptor; KCh, K⁺channel; MAP1A, microtubule-associated protein 1A; mGluR, metabotropic glutamate receptor; nNOS, neuronal nitric oxide synthase; PICK1, protein interacting with PRKCA1; SER, smooth endoplasmic reticulum; SPAR, spine-associated RAPGAP; SV, synaptic vesicle; SyNGAP, synaptic Ras GTPase-activating protein; TIAM1, T-cell lymphoma invasion and metastasis 1; TRAP, C-terminal receptor-binding region. Figure from Feng and Zhang 2009¹⁵⁸.

The PSD is composed of two interconnected protein layers. The first layer is formed primarily by PSD-95, but other members of the MAGUK family are also present. Below, the PDZ-containing protein Shank forms the second protein layer and associates with the cytoskeleton. GKAP binds the Guanylate Kinase domain of PSD-95 and the PDZ domain of Shank, forming a network that features an array of binding sites to which other PSD proteins

can attach (figure 2.3). Thus, the main role of PSD-95 is the assembly of a multi-protein signalling complex that is physically and functionally coupled to postsynaptic receptors¹⁵².

2.2 Structure of PDZ domains

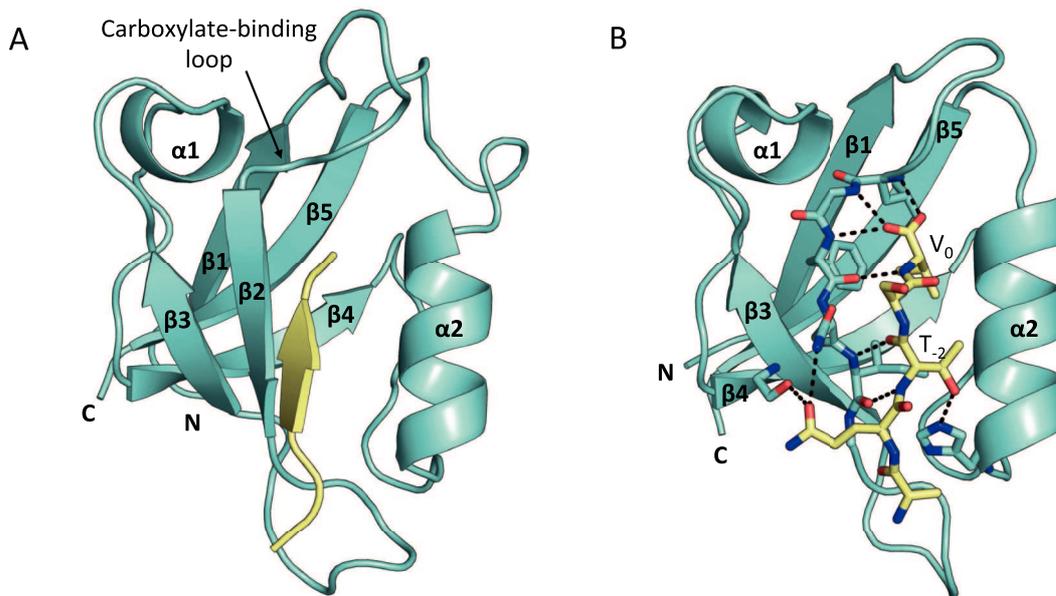


Figure 2.4: Structures of PDZ domains and their interaction with peptide ligands. PDZ domains (cyan) interacting with C-terminal peptide ligands (yellow). Secondary structure elements are indicated in bold letters. (A) complex between PTPN4-PDZ and the Cyto8-RETEV peptide ligand (PDB ID 5EYZ). (B) Detail of the binding mode of the PSD-95 PDZ3/CRIPT peptide ligand complex (PDB ID 1BE9). The $\beta 2$ -strand is shown as sticks. The N and C termini have been removed for clarity. Positions 0 and -2 of the PBM are indicated. Hydrogen bonds are represented as black dashes.

The first structures of free and liganded PDZ domains were published in 1996^{159,160}. Since then, a growing number of PDZ structures (about 530 structures in the PDB by August 2019) have shed light into the binding modes and determinants of recognition of these domains. The canonical PDZ fold consists of a β sandwich formed by 5 to 6 antiparallel β -strands flanked by 1 to 2 α -helices (figure 2.4.A). The N and C termini of the domain are usually located in close proximity, sometimes interacting with each other. This arrangement is common in modular protein interaction domains and reflects their structural independence from the rest of the molecule. The $\beta 2$ -strand and the $\alpha 2$ -helix form a groove where a short linear motif called PDZ-binding motif or PBM binds (developed below and in section 3). PDZ domains are capable of interacting with only one PBM at a time.

2.2.1 Peptide-binding groove

The peptide-binding groove is formed by the β 2-strand, the α 2-helix, and the β 1- β 2 loop, also known as the carboxylate-binding loop (figure 2.4.A). This loop contains a conserved Gly-Leu-Gly-Phe (GLGF) motif that is essential for coordination of the terminal carboxylate of the ligand.

The C-terminal residue of the PBM inserts into the binding groove and interacts with the GLGF motif of the carboxylate-binding loop. By convention, this position is known as position 0 (P_0); the subsequent residues in direction to the N terminus are numbered in reverse order (P_{-1} , P_{-2} , etc.) (figure 2.4.B). Backbone-backbone interactions between the PBM and the β 2-strand stabilize the PBM as an additional strand in the antiparallel β 2-sheet, a mechanism known as β -augmentation.

2.3 Classification of PDZ domains

The first classification system for PDZ domains arose from attempts to study the binding specificities of PDZ domains using a synthetic peptide library screening approach, which derived in the identification of two distinct consensus motifs¹⁶¹. PDZ domains were thus divided into two main specificity classes based on the consensus sequence of the PBM they recognise: class I (S/T-X- ϕ_{COOH}) and class II (ϕ -X- ϕ_{COOH}), where X is any amino acid and ϕ is a hydrophobic amino acid. Later, other that recognize different motifs were identified. These observations highlighted P_0 and P_{-2} as strong determinants for the selectivity of PDZ domains. However, it has been repeatedly observed that these two positions are not sufficient to account for the diversity and specificity of PDZ domains. For example, when the PDZ domain 2 (PDZ2) of membrane-associated guanylate kinase-related 3 (MAGI3) binding specificity was investigated against a library of randomized peptides, it was found that PDZ2 binds peptides with the consensus sequence S/T-W-V $_{COOH}$, which is consistent with a class I PDZ domain. In spite of this, it was clear from analysis of binding affinities that the W in P_{-1} (W_{-1}) was largely responsible in this case for the strong interaction between PDZ2 and this PBM. Indeed, mutation of W_{-1} to alanine dramatically decreased binding, and structural modelling indicated that the W side chain established Van der Waals contacts with the PDZ domain that stabilized the complex¹⁶².

The combined use of high-throughput screening techniques and binding affinity measurements has contributed to the definition of more classes of specificity, and has even led to the proposal that PDZ domain specificity might be more accurately described as a continuum, rather than classified into discrete specificity classes¹⁶³. Nonetheless, the original division into three classes is still in use, and for practical purposes will be used throughout this work to refer both to the PDZ domains and the peptides they recognise.

2.4 Structural determinants of ligand binding

2.4.1 Specificity at the C-terminal position

Hydrogen bonding of the terminal carboxyl oxygens to the backbone amide protons in the carboxylate binding loop stabilizes the residue in P₀ and causes its side chain to be oriented towards the interior of the binding groove (figure 2.4.B). The residues that form the carboxylate-binding pocket in P₀ are hydrophobic, therefore PDZ domains usually select for PBMs with hydrophobic residues in their C terminus (some exceptions will be reported in results part II). Depending on the size and configuration of the hydrophobic pocket, PDZ domains can display a preference for PBMs with V, L, I, F, or A at position 0, although this is usually not highly specific and variability in this position is often observed¹⁵².

2.4.2 Recognition of position -2

As discussed previously, P₋₂ is a major determinant of PBM recognition and has been taken as the basis of the first PDZ classification system. Specificity in this position is achieved by the interaction of the P₋₂ residue of the PBM with the first residue of the α 2 helix (position α 2-1, figure 2.5.A) in the PDZ domain. The nature of the α 2-1 residue determines the nature of the corresponding residue in P₋₂ in the interacting PBM. Class I PDZ domains have a conserved histidine in position α 2-1. Indeed, PBMs with a threonine or serine residue at this position can interact with the imidazole group of histidine by hydrogen bonding of their side chain hydroxyl group. This is seen in the archetypal example of class I PDZ domain-PBM interaction, the third PDZ domain of PSD-95 bound to the C-terminal PBM of CRIPT (Q-T-S-V_{COOH}), which features a T in P₋₂^{164,165} (figure 2.5.A).

In class II PDZ domains, a hydrophobic residue can be found in the place of the conserved histidine that determines preference for a hydrophobic residue in P₋₂. For example,

in the complex of PICK1 PDZ and GluR2 peptide, a lysine in position $\alpha 2-1$, through its methylen groups, determines preference for valine in P_{-2} of its ligand ligand (figure 2.5.B). However, in some cases the selectivity in this position can arise from interactions with other residues of the $\alpha 2$ helix. In the complex of Par3 PDZ3 and vascular endothelial cadherin peptide, L_{-2} is selected by hydrophobic interactions with arginine and methionine sidechains located in the middle section of the $\alpha 2$ helix (figure 2.5.C).

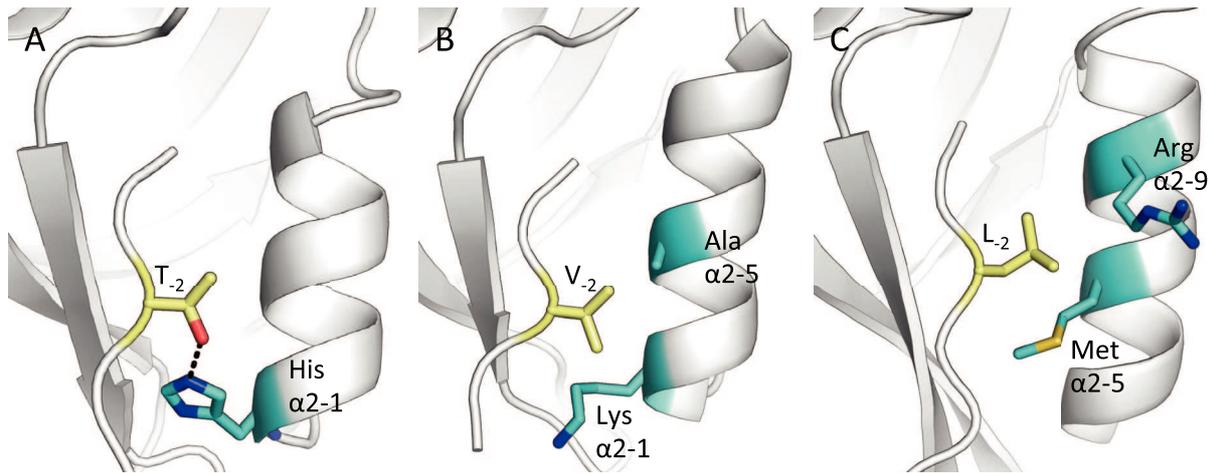


Figure 2.5: Recognition of position -2 by different subclasses of PDZ domains. Relevant domain and peptide residues are shown in cyan and yellow respectively. Hydrogen bonds are shown as black dashes. (A) Recognition of class I C-terminal peptides having a S or T at position -2 that establish a hydrogen bond with a conserved His in position $\alpha 2-1$ (PDZ of PSD-95 bound to CRIPT peptide, PDB ID 1BE9). (B and C) Recognition of class I C-terminal peptides. The hydrophobic residue at position -2 interacts with aliphatic side chains of the $\alpha 2$ -helix (B: PDZ of PICK1 bound to GluR2 peptide, PDB ID 2PKU; C: PDZ3 of Par3 complexed to vascular endothelial cadherin peptide, PDB ID 2KOH).

2.4.3 Recognition of position -3

In general, P_{-3} allows for more variability than P_{-2} , but residues at this position also make specific contacts with the PDZ domain. In the complex between the C terminus of CRIPT (Q-T-S- V_{COOH}) and PSD-95 PDZ3, the Q_{-3} establishes contacts with both the $\beta 2$ and the $\beta 3$ strands¹⁶⁵ (figure 2.4.B). Glutamine is conserved in known natural ligands of PSD-95 PDZ3¹⁶⁵, although glutamate is most commonly found at this position when screening peptide libraries for PSD-95 PDZ3 partners¹⁶¹. Both residues have the potential of establishing polar contacts, reinforcing the idea that position -3 has an impact, even if lesser, in determining selectivity¹⁶⁵.

Analysis of the Erbin PDZ and the zonula occludens 1 (ZO-1) PDZ1 showed that these two proteins differ markedly in their specificity in this position due to interactions with

residues in the $\beta 2$ and $\beta 3$ strands. Erbin selects for PBM with acidic residues in P_{-3} , because they can establish salt bridges with the arginine in position $\beta 3-5$ (figure 2.6). All the PDZ domains in the family of Erbin contain K or R in position $\beta 3-5$, and therefore preferentially bind peptides with D/E in P_{-3} . ZO-1 PDZ1 has a D in position $\beta 3-5$ that determines preference for residues with positively charge side chains in P_{-3} . Additionally, S and T are also accepted in this position P_{-3} , because they are able to form hydrogen bonds with the S in position $\beta 2-4$ (figure 2.6).

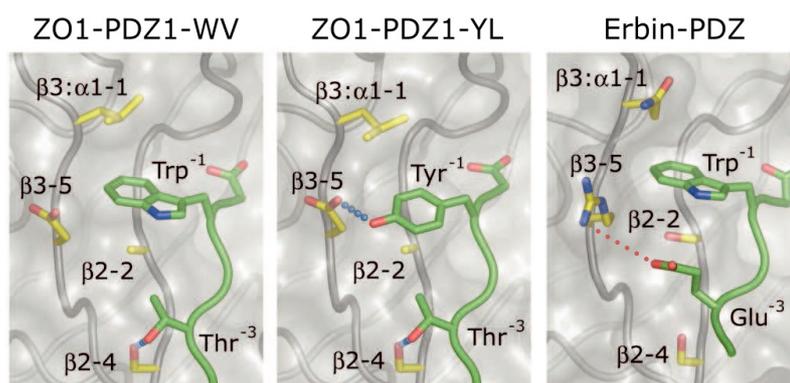


Figure 2.6: Specific interactions determine selectivity in positions -1 and -3 for ZO1 PDZ1 and Erbin PDZ. The PDZ domains are shown in grey with relevant side chains in yellow, and the peptide ligands are coloured in green. Hydrogen bonds and salt bridges are shown as blue and red dotted lines, respectively. Both Erbin-PDZ and ZO1-PDZ1 accommodate Trp-1, which extends over strand $\beta 2$. However, ZO1-PDZ1 also recognizes Tyr-1 because of a hydrogen bond formed with the carboxylate group of Asp($\beta 3-5$). In contrast, Erbin-PDZ contains an Arg at $\beta 3-5$, which instead mediates a salt bridge with Glu-3 (red dotted line). The ZO1-PDZ1 domain has a Ser in position ($\beta 2-4$) that forms hydrogen bonds with Thr in position -3 (Thr⁻³). Selectivity in Erbin-PDZ is defined by an Arg in position $\beta 3-5$, which mediates a salt bridge with Glu in position -3 (Glu⁻³).

2.4.4 Recognition of position -1

In the complex of CRIPT and PSD-95 PDZ3, the S_{-1} side chain does not contact the PDZ domain (figure 2.4.B). However, it would be erroneous to interpret this as an indication that this position plays no role in the binding properties. Indeed, replacement of S for D in this position (Q-T-S-V_{COOH} to Q-T-D-V_{COOH}) disrupted the interaction between CRIPT and PDZ3, while favouring its binding to PSD-95 PDZ1 and PDZ2¹⁶⁴.

This position is also determinant for the interaction of Erbin and ZO-1 PDZ1 domains and their ligands. For Erbin, tryptophan is strictly selected in P_{-1} . ZO-1, on the other hand, is able to accommodate either W or Y in this position¹⁶⁶. Selectivity for W in both cases arises from interactions with the main chains of the $\beta 2$ and $\beta 3$ strands¹⁶⁷. In ZO-1, a Y can occupy

the position of the W. The smaller Y side chain implies a loss of hydrophobic interactions, but it is able to form an H-bond with the carboxylate group of an aspartate in position β 3-5 that was found to be crucial for the affinity of the interaction. In Erbin, the arginine residue found in position β 3-5 is less suited to establish such an interaction, explaining why binding of Y is not possible in this case (figure 2.6)¹⁶⁶. Thus, main chain and side chain interactions are implicated in selectivity in P₋₁, showing that although in some cases it does not participate in binding, it cannot be disregarded as a specificity determinant.

2.4.5 Interactions beyond position -3

Accumulating evidence supports the contribution of interactions upstream of the core PBM in achieving selectivity or binding affinity^{161,168-170}. Further extending the PBM by using longer peptides has led to some unexpected results, such as interactions mediated by positions as distal as P₋₁₁¹⁷¹.

A good illustration is the complex between peptides derived from β -catenin and the Tax-interacting protein-1 PDZ domain. It was observed that a longer peptide bound the PDZ with higher affinity than a shorter one. The W₋₅ is responsible for this effect by establishing hydrophobic interactions with the β 3 strand and the β 2- β 3 loop¹⁷² (figure 2.7). This loop appears in many cases to be part of an extended binding pocket that participates in interaction with upstream residues in longer peptides, contributing to the diversity of binding modes and increasing the binding specificity and affinity of PDZ domains¹⁷³.

Upstream PBM positions have been observed to overlap with phosphorylation sites, suggesting that these interactions could be subjected to regulation. Indeed, it has been reported that phosphorylation can either increase the binding affinity, as in the case of S₋₈ of Cadherin bound to partitioning defective (Par)-3 PDZ¹⁷⁴; or decrease it, as seen for Y₋₇ of ErbB2 in complex with Erbin¹⁶⁹. Thus, upstream residues in some cases can account for the global affinity of the PDZ-PBM interaction, and provide a mechanism of regulation. This interesting topic has been reviewed by Luck, Charbonnier and Travé¹⁴⁸.

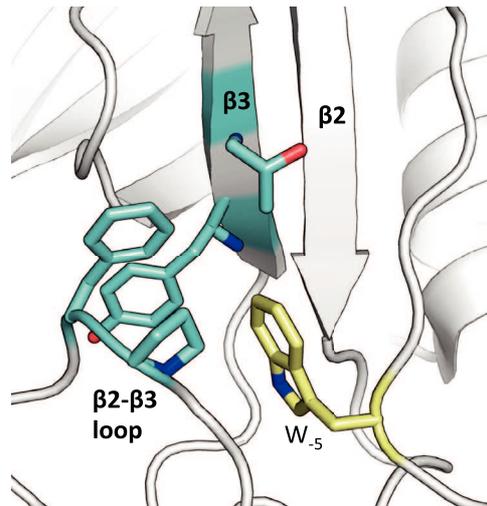


Figure 2.7: PDZ/PBM interactions beyond position -3 contribute to the binding affinity. Complex of Tax-interacting protein-1 PDZ domain and the β -catenin C-terminal peptide (PDB ID 3DIW). The W in position -5 (W_{-5}) establishes hydrophobic contacts with residues of the β 3-strand and β 2- β 3 loop, increasing the affinity of the interaction. Relevant domain and peptide residues are shown in cyan and yellow respectively.

2.5 The affinity of PDZ-mediated interactions

PDZ-mediated interactions need to be dynamic and transient to intervene in rapid and reversible responses to environmental changes, especially in signalling pathways. These characteristics correlate with moderate to low affinities between PDZ domains and their peptides¹⁷⁵. Indeed, the K_D s of PDZ-peptide interactions fall in the micromolar range¹⁷⁶, as do PDZ-lipid¹⁷⁷ and PDZ-PDZ interactions¹⁷⁸. Stiffler and collaborators set out to characterize interactions between 157 mouse PDZ domains and 217 peptides using protein microarrays and fluorescence polarization¹⁶³. From 1301 array positives, around 90% of the interactions had a K_D below 50 μ M, and around 60% of interactions had a K_D below 20 μ M, despite that the threshold they set for ‘positive’ interactions was rather high ($K_D < 100 \mu$ M). Similarly, 20% of the K_D s of the HPV E6 PBM peptide against a library containing all human PDZ domains, estimated from the binding intensities in a high-throughput assay, were below 250 μ M¹⁴⁹. The values are in agreement general observation that interactions mediated by linear motifs have affinities that range between 1–150 μ M¹⁷⁵. Moreover, it has been suggested that care should be taken when setting the threshold to determine positive interactions in these kinds of assays, since weak and promiscuous peptide-domain interactions are likely to become stronger and more specific in the context of the full-length interacting proteins¹⁷⁹.

2.6 Non-canonical binding modes and internal binding motifs

Besides the conventional C-terminal PBM binding, PDZ domains are capable of establishing other types of interactions. These non-canonical interactions could at least partially account for the high degree of specialization of the cellular functions of PDZ domains, which seems incompatible with their promiscuity and the rather low affinity of their interactions.

For example, some PDZ domains are capable of recognizing internal binding motifs. Comparison of the crystal structures of the PDZ domain of Par-6 in complex with an internal and a C-terminal PBM shows how the binding of the internal PBM is achieved thanks to the structural plasticity of the PDZ domain¹⁸⁰. Although the global fold of the domain is mostly conserved, the internal PBM adopts an extended conformation that causes the carboxylate-binding loop to reorient. This enables K165 to form a salt bridge with the PBM D₊₁ side chain that mimics the free carboxylate group. Backbone-backbone interactions replace the hydrogen bonds that normally are formed between the GLGF motif and the terminal carboxylate (figure 2.8).

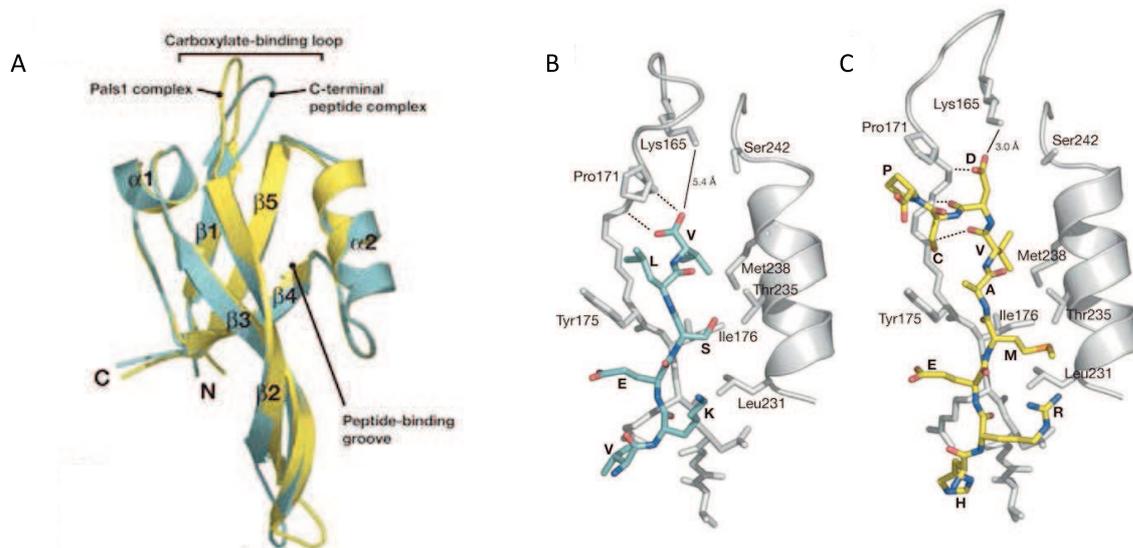


Figure 2.8: Structural Comparison of Par-6 PDZ internal and C-terminal binding. (A) Superimposition of Par-6 PDZ domains from the Pals1 complex and the C-terminal peptide complex (PDB code: 1RZX). The Pals1 complex is shown in yellow and the C-terminal complex in cyan with the secondary structure elements labelled. The ligands have been removed for clarity. (B) Par-6 PDZ - C-terminal ligand interactions. C-terminal peptide - Par-6 PDZ complex (PDB ID 1RZX). Peptide residues are labelled by their amino acid and PDZ domain residues are labelled by amino acid and sequence number. The distance between the C-terminus and lysine 165 is shown (solid line) along with interactions between the carboxylate and the PDZ backbone (dashed lines). (C) Par-6 PDZ - Pals1 internal ligand interactions. The interactions between the PDZ domain and peptide are shown as in (B). Adapted from Penkert, DiVittorio and Prehoda 2004¹⁸⁰.

Another interesting case of a non-canonical binding motif is found in the complex between the PDZ domains of nNOS and α 1-syntrophin. The C-terminal part of nNOS PDZ folds into a β -hairpin motif that inserts into the binding groove of syntrophin PDZ, without affecting the binding groove of nNOS PDZ (figure 2.9)^{181,182}. This interaction directs the subcellular localization of the complex, regulating NO production and signalling. Additionally, the free binding groove in nNOS PDZ domain mediates the interaction of syntrophin with other proteins¹⁸³.

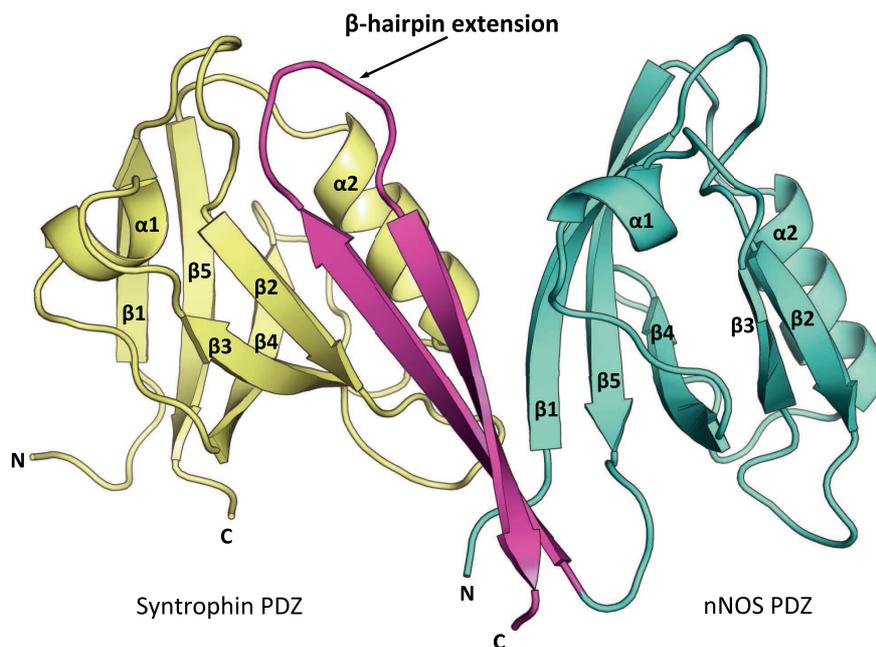


Figure 2.9: The structure of the nNOS PDZ–syntrophin PDZ heterodimer (PDB ID 1QAV). Ribbon representation of the nNOS PDZ (cyan)-syntrophin PDZ (yellow) heterodimer. The C-terminal part of nNOS PDZ folds into a β -hairpin motif (magenta) that inserts into the binding groove of syntrophin PDZ without affecting the binding groove of nNOS PDZ. The secondary structure elements of both PDZ domains are indicated.

Two studies suggest that at least 50% of PDZ domains can bind internal motifs. In one of them, protein partners of all *Caenorhabditis elegans* PDZ domains were searched for with high-throughput yeast two-hybrid screening coupled with validation by co-affinity purification¹⁸⁴. In another one, 24 PDZ domains were used as baits to screen a yeast two-hybrid library of random octapeptides¹⁸⁵. The significance and prevalence of these interactions have most likely been underestimated, partly because of the technical difficulties associated with the identification of internal PBMs that display a very short consensus of sequence. Other atypical binding modes have also been described, contributing to the diversity and growing complexity of PDZ domains-mediated interactions^{147,153}.

2.7 Other functions of PDZ domains

2.7.1 PDZ multimerization

PDZ domains can mediate dimerization and oligomerization of their parent PDZ-containing proteins¹⁸⁶. The capacity to form homo- and hetero-dimers and multimers is advantageous in the assembly of multi-protein complexes. Two cases illustrate different mechanisms by which PDZ dimerization can occur. First, the PDZ6 of GRIP1 can form back-to-back homodimers stabilized by interactions involving the β 1 strand and the α 1- β 4 loop of each protomer (figure 2.10.A). In this configuration, the PBM binding sites are located on distal sides of the interaction interface and are free to interact with ligands¹⁸⁷. Second, ZO-1 forms stable homodimers *in vitro* and *in vivo* via its second PDZ domain. Dimerization occurs via a β -strand-swapping mechanism, in which the β 2 and β 3 strands from one monomer are inserted into the PDZ fold of the other monomer, and *vice versa* (figure 2.10.B)^{188,189}. Here again, both binding grooves are free to interact with peptide ligands. A large-scale study was conducted to establish the frequency in which PDZ-PDZ dimerization occurs¹⁷⁸. Protein microarray technology was used to assess over 12400 potential dimerization events between 157 mouse PDZ domains. Positive interacting pairs were further validated using a solution-phase fluorescence polarization assay. 37 PDZ-PDZ interactions were confirmed *in vitro*, formed between 46 different PDZ domains. This suggests that at least 30% of mammalian PDZ domains could potentially be involved in PDZ-PDZ interactions. Further biochemical and bioinformatics analysis of the tighter binding pairs ($K_D < 5 \mu\text{M}$) led to the proposal that PDZ dimerization contributes to selective formation and strengthening of multi-protein complexes.

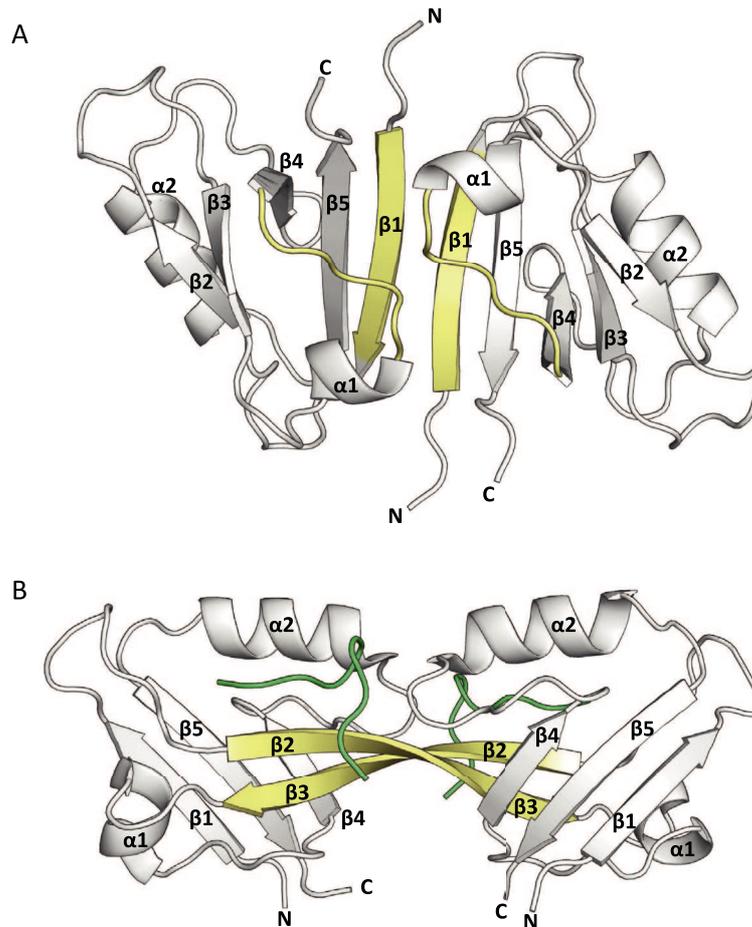


Figure 2.10: two PDZ dimerization mechanisms. (A) GRIP-1 PDZ6 homodimer (PDB ID 1N7E). The secondary structure elements involved in dimer formation are shown in yellow. (B) ZO-1 PDZ2 homodimer (PDB ID 3CYY). The β -strands that swap to stabilize the homodimer are coloured in yellow. The bound connexin43 ligands are represented in green.

Two examples of this phenomenon are found in the PSD. First, in PSD-95, a short, rigid, and highly conserved linker joins the first two PDZ domains. NMR structural analysis showed that the two domains establish contacts with each other, forming a tandem with restricted interdomain motions in which the PBM-binding grooves are oriented in an almost parallel manner. This arrangement results in synergistic interactions with dimeric ligands, which is thought to enhance the binding to multimeric membrane proteins such as the NMDA receptor channel (figure 2.11.A)¹⁹⁰. Second, the five PDZ domain-containing protein GRIP is also a component of the PSD (figure 2.3), where it binds the C-terminus of the AMPA receptor GluR2 subunit through its PDZ5 domain^{191,192}. However, covalently linked PDZ4 is an absolute requirement for this interaction to occur. Structural analysis of the supramodule revealed that PDZ4 acts as a chaperone, stabilizing the folding of PDZ5 and thus enabling the interaction (figure 2.11.B)^{193,194}.

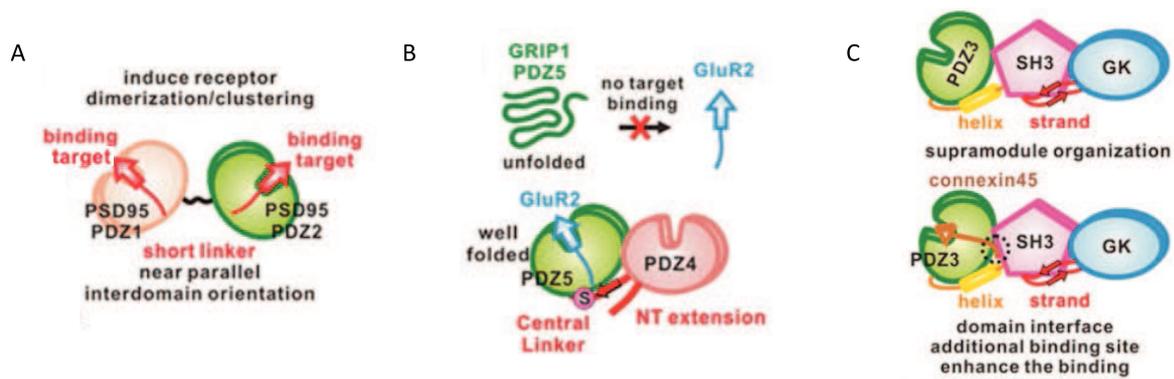


Figure 2.11: Schematic diagrams of the distinct target-binding properties conferred by the supramodular organization of GRIP1 and ZO-1 PDZ domains. (A) The protomers in the PSD-95 PDZ12 tandem have a limited conformational freedom, which results in synergistic interactions with dimeric ligands. (B) The supramodular organization of the PDZ tandem for the GRIP1 PDZ domains 4 and 5 generates distinct target-binding properties. The interdomain linker (red), which forms a β -strand (red arrow) antiparallel to β 1 of PDZ5 is highlighted. (C) ZO-1 PDZ3-SH3-GK (green-pink-blue) supramodule organization (top) and formation of the ZO-1 PDZ3-SH3-GK/connexin45 complex (bottom). The interaction surface and additional binding site generated by the formation of the supramodule is represented by the black dotted circle. Adapted from Ye and Zhang 2013¹⁵³.

2.7.2 Heterotypic supramodules

It is also common that PDZ domains lie in close proximity to other protein interaction domains. Similarly to what happens with homotypic supramodules, a growing number of cases is evidencing that PDZ domains also co-operatively interact with these neighbouring domains establishing higher-order units with distinct functional properties, termed heterotypic supramodules¹⁵³. ZO-1 and PTPN3 and 4 are two examples.

ZO-1 is a member of the MAGUK family. MAGUKs are scaffolding proteins known to be implicated in diverse functions, including cell polarity establishment and maintenance, and signal transduction¹⁹⁵. All the MAGUK family proteins, with the exception of MAGI, share the same core arrangement comprising a PDZ domain (PDZ3), a Src homology 3 (SH3) domain and a catalytically inactive guanylate kinase (GK) domain (figure 2.1). The PDZ3 domain of ZO-1 establishes extensive interdomain contacts with the SH3 domain, which is in turn coupled to the GK domain, forming between the three a heterotypic supramodule (figure 2.11.C). Contacts between the SH3 and PDZ3 domains significantly alter the binding properties of PDZ3 by creating an additional binding site at the inter-domain interface^{196,197}. Similar arrangements have been proposed to exist in other MAGUK family proteins, although the structural and functional details remain to be explored¹⁵³.

PTPN3 and its close homolog PTPN4 feature a supramodule in which their PDZ domain transiently interacts with the phosphatase domain, inhibiting its catalytic activity^{70,198}.

It was shown by our team that binding of a PBM ligand to the PDZ domain of PTPN4 released this auto-inhibition that operates by allosteric modulation and changes of the internal dynamics of the PDZ-PTP supramodule^{198,199}. These observations contribute to the notion that scaffolding domains such as PDZ can actively participate in the regulation of signalling events¹⁵⁸.

2.7.3 Phosphatidylinositol and cholesterol binding

Phosphorylated forms of phosphatidylinositol (PIPs) are important components of cell membranes and participate in cell signalling and regulation of cell polarization²⁰⁰. Despite being originally described as protein-binding modules, several PDZ domains are able to bind PIPs through electrostatic interactions between a patch of cationic residues in the PDZ domain and the negatively charged head of the phospholipid. This type of interaction has been observed for the PDZ domain of Par-3, the PICK-1 PDZ domain, and the ZO-1 PDZ2 domain (figure 2.12)²⁰¹. Depending on the binding mode, simultaneous binding of peptide ligands and lipids could be possible. For ZO-1 PDZ1 and PDZ2 and Par-3 PDZ, the peptide and lipid binding sites overlap; therefore the bindings of peptide and phospholipid are mutually exclusive^{202,203}. On the contrary, PICK-1 and -2 and CASK are able to bind peptides and PIPs simultaneously, and the interactions were found to be cooperative^{204,205}.

Large-scale studies encompassing *in vitro*, *in silico* and *in cell* assays with panels of human and *Drosophila* PDZ domains aimed to determine the occurrence of PDZ-PIP interactions^{177,203,205}. These indicated that 20-30% of human PDZ domains are capable of interacting with phospholipids, and highlighted the diversity of PIP binding motifs.

PDZ domains are also able to bind cholesterol, another major component of mammalian cell membranes. This function has been less investigated than PIP binding, but a recent study²⁰⁶ showed that 12 over a panel of 30 PDZ domains could bind cholesterol with an affinity in the range of micromolar or higher. The functional effects of this interaction were assessed for the NHERF-1 PDZ domain. Association with cholesterol is essential for NHERF-1 colocalization with its known binding partner CFTR ion channel, suggesting that cholesterol binding strengthens the PDZ-mediated interaction. Moreover, cholesterol binding seems to be required for CFTR activation²⁰⁶.

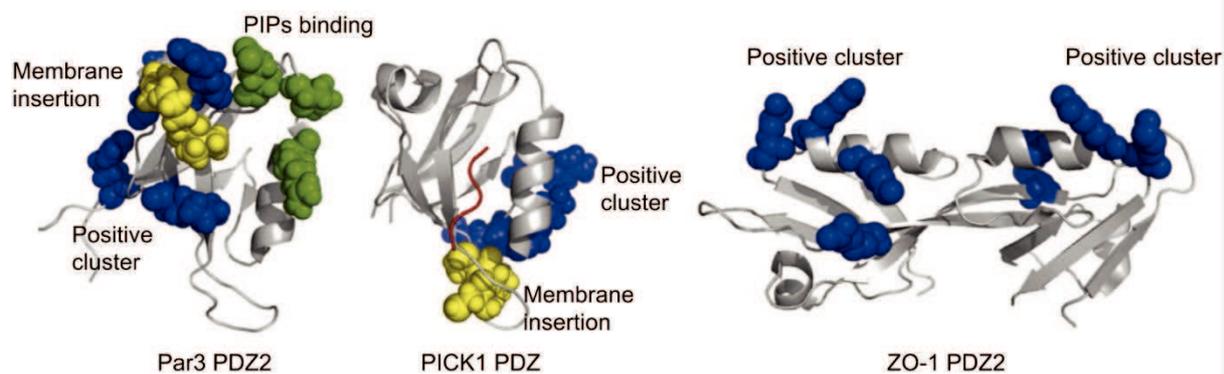


Figure 2.12: PDZ-phospholipid interactions. PDZ domains interact with phosphatidylinositol (PIP)-containing lipid membranes via different interaction surfaces. Par3 PDZ2 (left, PDB ID 2OGP) interacts with negatively charged lipid membranes through a combination of electrostatic interactions (indicated in blue spheres), membrane penetration (in yellow spheres), and a defined PIPs binding site (in green spheres). PICK1 PDZ (middle, PDB ID 2PKU) interacts with lipid membranes through a conserved membrane insertion motif (yellow spheres) and a positive charge cluster (in blue spheres). A bound peptide is indicated in red. In ZO-1 PDZ2 (right, PDB ID 2RCZ) a set of basic residues (blue spheres) participate in interactions with PIP-containing liposomes. Adapted from Ivarsson 2012²⁰¹.

Phospholipid and cholesterol binding appears to be a relatively frequent feature of PDZ domains. This could be correlated with the fact that a vast majority of PDZ proteins are associated with the plasma membrane²⁰⁷. Lipid binding can compete with or potentiate PBM binding, and recruit the PDZ domain to cellular membranes, determining in this way the interacting partners it is more likely to encounter. The fact that phosphoinositide binding modules such as PH, PTB or FERM are commonly present in signalling proteins highlights the importance of lipid binding in the organization of signalling complexes and targeting proteins to the membrane²⁰⁸. Further studies will be required to shed more light into the prevalence and physiological relevance of these interactions.

2.8 Regulation of PDZ-mediated interactions

2.8.1 Phosphorylation

Phosphorylation has been found to occur in both the PBM and the PDZ domain, and it can function either as a positive or as a negative modulator¹⁴⁷. PBMs often contain serine, threonine and tyrosine residues that can be phosphorylated. In the majority of cases, addition of a phosphate group to the PBM disrupts the interaction with the PDZ domain, but in few others the interaction can be strengthened, suggesting that this is a mechanism to increase selectivity in PDZ/PBM interactions¹⁴⁷. A well-characterized example of phosphorylation-

regulated PDZ/PBM interactions is the GluR2 subunit of AMPA receptors. GluR2 contains a C-terminal PBM (-IESVKI_{COOH}) that binds to the PDZ domain of PICK1 and PDZ4,5 tandem of GRIP1²⁰⁹. The protein kinase C (PKC) can phosphorylate S₋₃ of the GluR2 PBM. This phosphorylation differentially regulates the binding of GluR2 to GRIP1 and PICK1. S₋₃ phosphorylation greatly decreases the binding affinity of the GluR2 PBM to GRIP1 PDZ4,5, although it does not affect its binding to PICK1 PDZ²¹⁰ (figure 2.13.A:). The residue in position α 2-1 is responsible for this discrimination. PICK1 PDZ contains a lysine that favours binding to the phosphorylated GluR2 PBM through electrostatic interactions. Instead, GRIP1 PDZ5 contains a glutamic acid in this position, which is not suited to interact with the phosphorylated PBM, therefore binding affinity is diminished²⁰⁴.

Large-scale studies have highlighted the significance of P₋₁, and to a lesser extent P₋₂, in this type of regulation^{211,212}. Phosphorylation of the PDZ domain, although less studied, has been observed to have a much more moderate effect in binding than when it occurs on the peptide²¹³.

2.8.2 Auto-inhibition

Some PDZ-containing proteins contain a C-terminal PBM with the capacity to bind their own PDZ domain, competing with other peptide ligands. For example, X11 α has two PDZ domains that directly interact with each other forming a tandem PDZ supramodule (figure 2.1 Domain organizations of a selected set of PDZ domain proteins). The C-terminal tail of X11 α folds back and inserts into the target-binding groove of its own PDZ1 domain, resulting in a closed conformation of PDZ1²¹⁴ (Figure 2.13.B). The authors of this study proposed that the auto-inhibition of X11 α could be subject to regulation by phosphorylation in a highly conserved tyrosine residue at the -1 position, which is essential for binding to PDZ1. Replacing the Y₋₁ residue with E₋₁ to mimic phosphorylation disrupted the PBM binding to PDZ1, releasing the auto-inhibition. Curiously, the mutated PBM tail was found to bind the PDZ2, suggesting a potential phosphorylation-dependent dynamic regulation mechanism of the binding properties of the X11 α PDZ supramodule mediated by its Y₋₁²¹⁴ (Figure 2.13.B:)

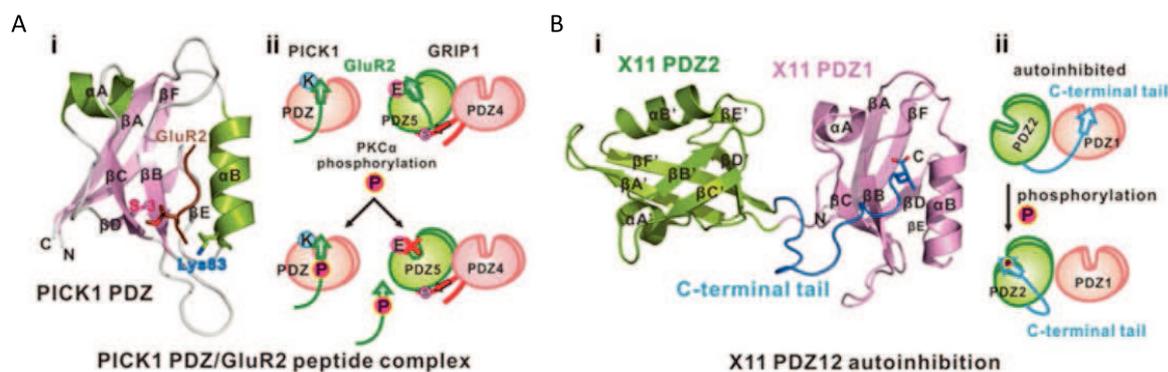


Figure 2.13: Regulation of PDZ–target interactions by phosphorylation and auto-inhibition. (A) Phosphorylation-mediated regulation of PDZ–target interactions. (A, panel i) Structure of PICK1 PDZ bound to the GluR2 peptide (PDB ID 2PKU). (A, panel ii) Schematic showing that the PKC phosphorylation of GluR2 differentially regulates its binding to GRIP1 PDZ45 and PICK1 PDZ. Lys83 at the α 2-1 position of PICK1 PDZ is highlighted. (B) Auto-inhibition of X11 PDZ1 by its own C-terminal PBM. (B, panel i) Structure of the auto-inhibited conformation of X11 α PDZ12 (PDB code 1U3B). (B, panel ii) Schematic diagram showing the phosphorylation-mediated switch of the inhibition of the two PDZ domains by the C-terminal PBM. The C-terminal tail of X11 α inserts into the peptide-binding groove of PDZ1, resulting in a closed conformation of PDZ1. Phosphorylation of Y₁ releases the PBM from PDZ1 and promotes its binding to PDZ2. Adapted from Ye and Zhang 2013¹⁵³.

2.8.3 Allosteric conformational regulation

The binding properties of PDZ domains can be modulated by long-range allosteric conformational changes. These changes can be produced by interactions with structural elements in the adjacent sequence of the PDZ domains, or by interactions that occur in homo- or heterotypical supramodules¹⁵³. For example, both PSD-95 PDZ3²¹⁵ and NHERF1 PDZ2²¹⁶ have helical extensions that make contacts in sites distal to the peptide-binding cleft (figure 2.14). If these helices are removed, the truncated PDZ domains are in a more dynamic conformation. The binding of a peptide to the truncated PDZ domains causes a penalty in conformational entropy, which is translated into a lower binding affinity. When the helices are present, their contacts with the core PDZ domain allosterically modulate the internal dynamics and side-chain conformations in the PDZ domain. This results in more structurally stable PDZ domains, which lowers the entropic penalty of peptide binding and reflects in higher binding affinities. The reader can refer to a review by Jemth and Gianni²¹⁷ for a comprehensive analysis of conserved networks of interacting residues in PDZ domains, and the relationship between binding, stability, and allostery; and to one by Gautier and collaborators²¹⁸ describing different computational and experimental approaches to characterize allosteric networks in PDZ domains.

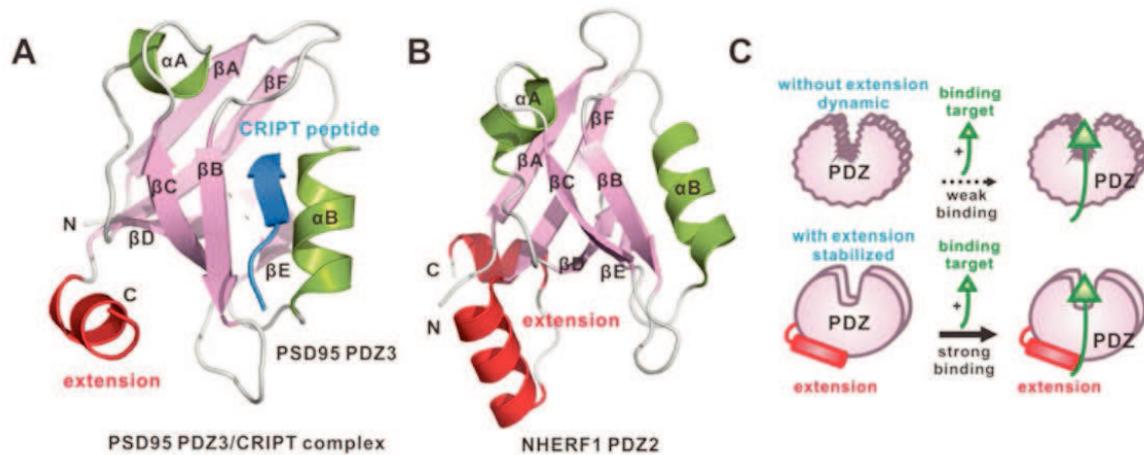


Figure 2.14: Allosteric conformational regulation of PDZ domains. (A) Structure of the complex of PSD-95 PDZ3 with the CRIP1 peptide (blue) (PDB code 1BE9). (B) Structure of the unbound NHERF1 PDZ2 (PDB code 2KJD). The C-terminal α -helix extension of PSD-95 PDZ3 and that of NHERF1 PDZ2 are coloured in red. (C) Schematic showing the truncated PDZ domains are dynamic. The inclusion of the C-terminal α -helical extension reduces the conformational entropy of both PDZ domains and enhances their target-binding affinities. Adapted from Ye and Zhang 2013¹⁵³.

2.8.4 Alternative splicing

PDZ domains or PBMs can be deleted from different protein isoforms by alternative splicing, consequently affecting their subcellular localizations and functions. For example, in the murine inner ear, the protein usherin exists in several isoforms. One of these isoforms is a transmembrane protein that features a PBM that can bind the PDZ protein whirlin²¹⁹. In turn, there are two isoforms of whirlin in the inner ear, one long and one short. The full-length isoform has three PDZ domains and interacts with usherin, while the short isoform only has one PDZ domain and is unable to interact with usherin²²⁰. Mice carrying a mutation that abolishes the long isoform of whirlin and renders the protein unable to interact with usherin display hearing loss, simulating the effects found in humans with Usher syndrome type II, caused by similar mutations²²¹.

3 Targeting of PDZ domains by viruses

3.1 Viral interactions with PDZ-containing proteins

Viruses are obligate intracellular parasites that require their host cell functions to replicate and disseminate. One of the strategies that viruses have developed for this is molecular mimicry. This is the adoption or imitation of a host's characteristics to successfully interact with its cellular machinery²²². Frequently, viruses target interactions mediated by short linear motifs. These are short sequences of 3 to 10 amino acids that mediate transient, low-affinity protein-protein contacts, and determine subcellular targeting and also encode post-translational modification sites²²³. They constitute hotspots for convergent evolution due to their short length and lack of structural constraints. Indeed, they may arise from a single point mutation in an otherwise inert sequence²²². Considering the prevalence and importance of PDZ proteins in cellular functions and the minimal 3 residues-long motif required for recognition, it is hardly surprising that many viruses display PBMs in their proteins to hijack host cell functions²²⁴.

The first viral PBMs were identified in the E4-ORF1 protein of the human adenovirus (AdV) 9, the Tax protein of the human T-lymphotropic virus type 1 (HTLV-1) and the HPV E6 protein²²⁵⁻²²⁸. These three proteins have a tumorigenic activity and target DLG1, the mammalian homolog of the *Drosophila* tumour suppressor Discs large. Later, PBMs were found as well in non-oncogenic viruses such as rabies and influenza A. It is striking how different families of viruses with distinct replication cycles have evolved to target the same set of cellular PDZ proteins, although this can be explained by the fact that all these pathogens are faced with similar host environments and pathways. Table 3.1 details known PDZ protein targets of viral PBM-containing proteins in the virus families that are discussed in this work.

Table 3.1: Known cellular PDZ protein targets of viral proteins. * Although PDLIM2 is a PDZ protein, the interaction with Tax is not mediated by this domain, but rather by an α -helical region in PDLIM2 and multiple regions in Tax²²⁹. AdV9, human adenovirus 9; HTLV-1, human T-lymphotropic virus type 1; HPV, human papillomavirus; HBV, hepatitis B virus. Adapted from Javier and Rice 2011, and Ganti et al 2015^{229,230}.

Virus	AdV 9	HTLV-1	Rabies	HPV	HBV
Protein	E4-ORF1	Tax	G	E6	HBc
PBM	-ATLV _{COOH}	-ETEV _{COOH}	-ETRL _{COOH}	16 -ETQL _{COOH} 18 -ETQV _{COOH}	-ESQC _{COOH}
PDZ targets	DLG1 MAGI-1 MUPP1 PATJ ZO-2	β 1-syntrophin DLG1 DLG4 ERBIN MAGI-3 pro-IL-16 SCRIB TIP-1 TIP-2/GIPC PDLIM2*	DLG1 MAST2 MUPP1 PTPN4	CAL/GOPC DLG1 DLG4 MAGI-1 MAGI-2 MAGI-3 MUPP1 PATJ PTPN3 PTPN13 SCRIB TIP-1 TIP-2/GIPC PDZRN3 14-3-3 PAR3 SNX27 ARGGEF12 FRMPD2 LRRC7	PTPN3 TIP-2/GIPC

3.1.1 Cell polarity proteins: a common target

The large majority of PDZ-containing proteins targeted by viruses are involved in cell polarity. Cell polarity originates from the asymmetric segregation and distribution of proteins within the cell and is responsible for the ordered orientation of cells within a tissue. Apicobasal polarity is defined by the existence of distinct apical and basolateral membrane domains, each with its particular lipid and protein composition. In epithelial cells, apicobasal polarity is defined and regulated mainly by three spatially segregated protein complexes: the PAR complex, which associates with tight junctions (TJ) and promotes the establishment of the apical-basal membrane border; the CRUMBS complex, required for the establishment of the apical membrane; and the SCRIBBLE complex, associated to adherens junctions (AJ) and responsible for the definition of the basolateral domain^{231,232}. The spatial organization and the molecular composition of these complexes are illustrated in figure 3.1.

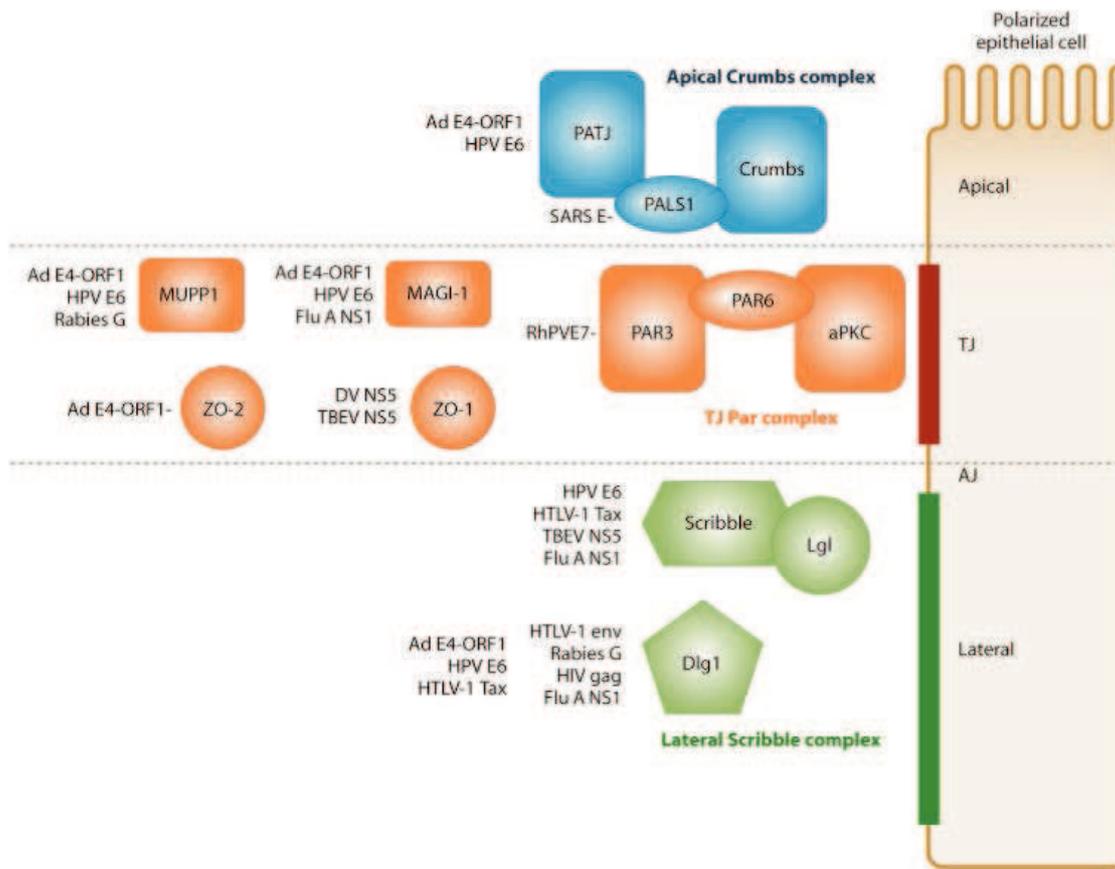


Figure 3.1: Cellular PDZ proteins that function in cell polarity establishment and cell junction formation are commonly targeted by viruses. On the right, a polarized epithelial cell divided into its apical, tight junction (TJ) and lateral regions. On the left, the cellular PDZ proteins that localize to these regions and function to regulate the formation of TJ or adherens junction (AJ), and/or establishing the cell polarity. Examples of viral proteins known to interact with these proteins are indicated. Figure from Javier and Rice 2011²³³.

Most of the proteins that form these complexes are scaffold proteins that feature at least one PDZ domain. Targeting PDZ domains in these proteins usually disrupts their normal interactions, likely affecting cellular functions to the benefit of viral dissemination^{229,234}. The viral hijacking of PDZ polarity proteins has been reviewed with a focus on its consequences in tumorigenesis^{224,229,234} and as a mechanism of modulation of the immune system functions²³⁵. Indeed, the proteins that form the polarity complexes are conserved in mammals, where they are involved in cell-cell adhesions and epithelial cell polarity, but also play different roles in other types of cells. In T and B lymphocytes, for example, polarity is responsible for the effects that occur during migration, antigen recognition, the establishment of the immunological synapse, and during effector functions^{231,236}. Importantly, the loss of cell polarity has been related to tumorigenesis. Polarization determines the correct localization of receptors and enables cells to interact with their surroundings in an anisotropic manner,

sensing cues from neighbouring cells or the surrounding microenvironment and responding appropriately. Therefore, disruption of polarity prevents cells from responding to growth-inhibitory signals and can deregulate senescence, differentiation, and apoptosis²³². Altered polarity can also lead to incorrect distribution of degradative enzymes such as matrix proteases at the cell surface, affecting cell migration and stimulating cell invasion²³⁷. Moreover, many of the polarity proteins function as tumour suppressors or proto-oncoproteins and are linked to cell growth and proliferation signalling pathways.

Three remarkable examples reported below were chosen to illustrate the variety of viruses targeting PDZ domains through a PBM present in one of their proteins, and how this deregulates cell functions to promote viral replication and dissemination, or to escape immune detection.

3.1.2 Human adenovirus

Human adenovirus (AdV) is the causative agent of acute and self-limiting infections in the respiratory tracts, conjunctivitis, hemorrhagic cystitis and gastroenteritis. Although AdVs are not tumorigenic for humans, human adenovirus 9 (AdV9) can produce estrogen-dependent mammary tumors in rodents under experimental conditions²³⁸. The main oncogenic determinant of AdV9 is the E4-ORF1 protein, which contains a class I PBM (-ATLV_{COOH}), the first viral PBM to be identified^{226,228}. This sequence can bind the MAGUK family members DLG1, ZO-2, and MAGI-1, as well as MUPP1 and PATJ, which belong to the multi-PDZ family (figure 2.1). These scaffolding proteins participate in the assembly of signal transduction complexes and in the formation of cell-cell contacts. In epithelial cells, MUPP1, PATJ, MAGI-1, and ZO-2 are part of the TJ, while DLG1 participates in AJ assembly (figure 3.1). AJ are responsible for cell-cell adhesion, while TJ contribute to determining cell polarity by acting as selective permeability barriers that divide cells into distinct apical and basolateral membrane domains²³⁹.

The E4-ORF1 PBM has different effects in the cell. First, it targets the TJ proteins and inactivates them by sequestering them in insoluble complexes in the cytoplasm. This disrupts TJ and deregulates TJ-associated polarity, two characteristic features of cancer cells²³⁹. A second function of the E4-ORF1 PBM is the interaction with an oncogenic, membrane-associated isoform of Dlg1, resulting in a complex that hijacks cytoplasmic phosphatidylinositol 3-kinase (PI3K) and relocates it to the membrane, resulting in constitutive, growth factor-independent activation of the PI3K signalling pathway²⁴⁰. The E4-

ORF1 PBM is also present in other subgroups of adenovirus that lack a transforming activity, such as the subgroup C adenovirus type 5 and subgroup A adenovirus type 12. These PBMs are able to interact with DLG1, MUPP1, and MAGI-1, but fail to bind ZO-2²⁴¹. ZO-2 has been proposed to act as a tumour suppressor in breast cancer, as its expression was lost or significantly decreased in at least 80% of breast cancer lines and primary breast adenocarcinomas examined²⁴². The transforming and tumorigenic potential of Ad9 E4-ORF1 in cells depends on its ability to bind and sequester ZO-2 through its PBM.

3.1.3 Human T-lymphotropic virus type 1

The human T-lymphotropic virus type 1 (HTLV-1) is another well-studied case of targeting of PDZ domains through a PDZ-PBM interaction during viral infection. HTLV-1 is a retrovirus that infects CD4+ T lymphocytes and causes adult T-cell leukemia and lymphoma. The viral oncoprotein Tax is required for viral replication and transformation in mouse, rat, and human cells^{243–245}. Tax has a C-terminal class I PBM (-ETEV_{COOH}) that plays a major role in the oncogenic potential of the virus²²⁴. This is supported by the absence of the PBM in the Tax protein of the non-oncogenic HTLV-2 virus. The Tax PBM targets the Scribble cell polarity complex proteins DLG1 and SCRIB (figure 3.1). DLG1 and SCRIB participate in the formation of signalling complexes at specific sites of T-cells, and they are required for normal T-cell receptor-induced activation and inhibition of the proliferative PI3K/AKT pathway. Similarly to the Adv9 E4-ORF1 protein, the PBM-dependent interactions of DLG1 and SCRIB with Tax result in their inactivation by re-localizing them to insoluble complexes. In T cells, DLG1 is required to inhibit AKT-mediated signalling, and both DLG1 and SCRIB modulate TCR-induced activation via the transcription factor nuclear factor of activated T cells. Tax sequestration of DLG1 and SCRIB via its PBM may disrupt their scaffolding functions, deregulating normal T-cell signalling²²⁴. Other PDZ proteins such as MAGI-1 and MAGI-3 have been shown to interact with the Tax PBM, but the relevance of these interactions has not been explored.

The HTLV-1 envelope glycoprotein Env also targets DLG1 through a PBM. It is thought that this interaction favours the dissemination of the virus in the host by recruiting DLG1 to stabilize the viral synapses (cellular junctions induced by some viruses to allow cell-to-cell transmission) formed between neighbouring T-cells²⁴⁶. Tax enhances this process by hijacking the microtubule-organizing center, a major site of microtubule nucleation, to form

the viral synapse²⁴⁷. Thus, targeting host PDZ proteins allows HTLV-1 to manipulate cell polarity to enhance viral transmission while evading detection by the immune system²⁴⁸.

3.1.4 Rabies virus

Finally, the last example concerns the Rabies virus (RABV). RABV is a single negative-strand RNA virus that exclusively infects neurons and is the causative agent of rabies. Until the development by Louis Pasteur and Émile Roux in 1885 of the rabies vaccine made of an attenuated RABV strain, infection with rabies virus was almost invariably fatal. Virulent RABV strains promote neuron survival to favour their propagation, while attenuated RABV strains cause neuron apoptosis, and triggers an immune response following the vaccination²⁴⁹.

RABV glycoprotein (G) is the only protein exposed at the surface of the virion and acts as a ligand for cellular receptors, triggering virus endocytosis. Variations in the sequence of this protein correlate with the virulence of the virus strain and determine the fate of the infected cells. Our team and collaborators determined that a single mutation located in the C-terminal cytoplasmic tail of the G protein in position -3 of a class I PBM (-QTRL_{COOH} for virulent strains, -ETRL_{COOH} for attenuated strains,) alters the motif's specificity (figure 3.2.A) and is responsible for the switch between cell survival and cell death of infected neurons. The PBM of virulent strains binds mainly the MAST2 kinase, an inhibitor of neurite outgrowth in human neuroblastoma^{250,251}. MAST2 exerts its anti-survival function by phosphorylating its endogenous partner, Phosphatase and tensin homolog deleted on chromosome 10 (PTEN), a tumor suppressor and modulator of the PI3K/AKT proliferative pathway. Phosphorylation decreases PTEN activity by regulating its subcellular localization, stability, and conformation²⁵². MAST2 PDZ domain binds the C-terminal PBM of PTEN, increasing PTEN phosphorylation and maintaining cellular homeostasis. During infection with a virulent strain, the PBM in G protein competes with PTEN for MAST2-PDZ binding, preventing phosphorylation and disrupting the normal function of PTEN, which results in activation of the proliferative PI3K/AKT pathway^{171,250} (figure 3.2.B). The attenuated strain of RABV possesses a Q to E mutation in P₋₃ of the G protein PBM that enables G to interact with additional PDZ proteins, including DLG2, MUPP1, and particularly PTPN4, which plays an anti-apoptotic role in neurons^{250,253}. In normal conditions, PTPN4 likely maintains homeostasis by dephosphorylating its endogenous partners. During infection with an

attenuated RABV strain, the G PBM targets the PDZ domain of PTPN4 triggering apoptosis, possibly by disrupting the interaction between PTPN4 and its substrates²⁵⁴ (figure 3.2.B).

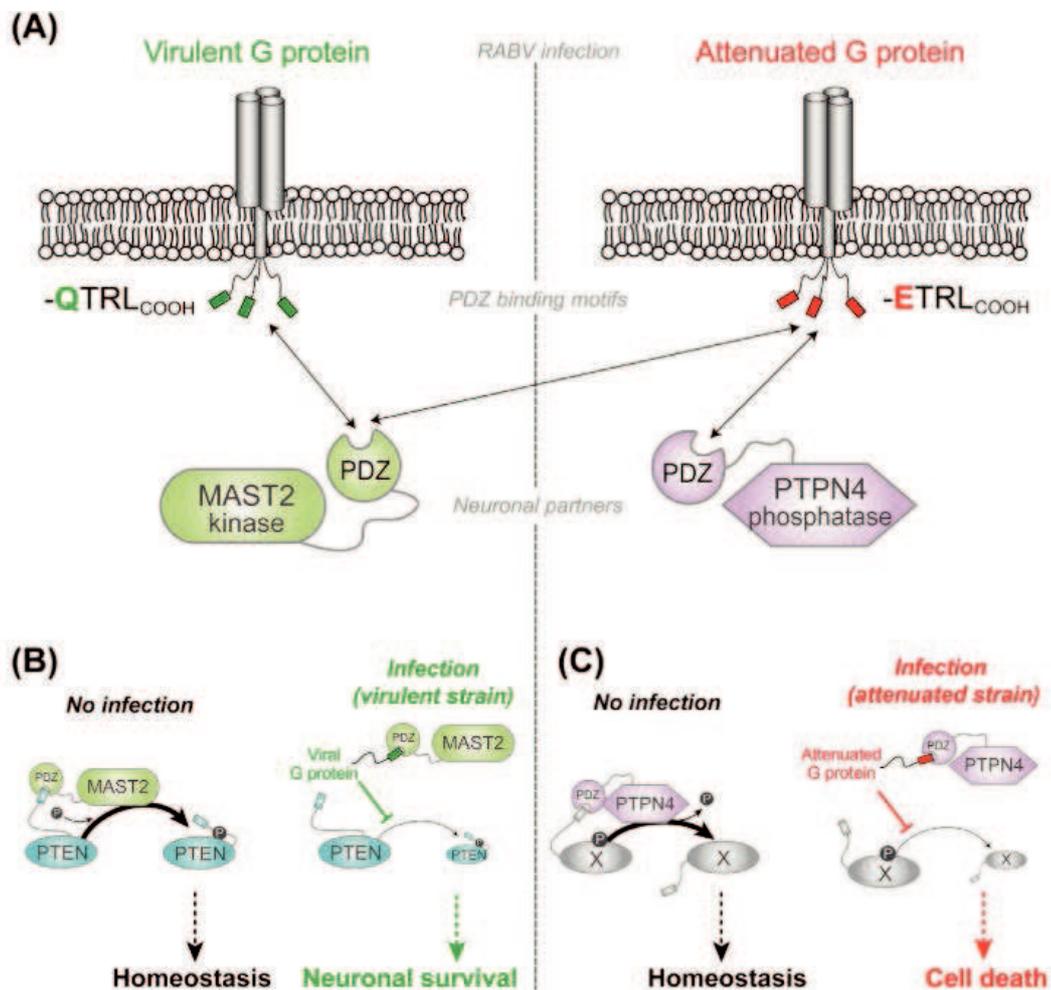


Figure 3.2: Model of infection by a virulent or an attenuated strain of RABV and associated proteins. (A) Upon infection by the virulent strain (left), the viral G protein with the -QTRL_{COOH} PBM interacts with the PDZ domain of MAST2 kinase, whereas upon infection by the attenuated strain (right), the viral G protein with the -ETRL_{COOH} PBM interacts not only with MAST2 but also with other cellular partners, particularly PTPN4. (B) In the absence of infection, PTEN is phosphorylated, notably by MAST2. Upon infection by the virulent strain, the viral G protein interacts via its PBM with the PDZ domain of the MAST2. This interaction could prevent MAST2-controlled phosphorylation of PTEN, which alters the homeostasis of the infected cell and leads to its survival. (C) In the absence of infection, the hypothetical cellular partner X is dephosphorylated by PTPN4. Upon infection by the attenuated strain, the viral PBM interacts with the PDZ domain of PTPN4. This interaction prevents efficient dephosphorylation of X, which alters the homeostasis of the infected cell and triggers apoptosis. Figure from Caillet-Saguy *et al.* 2015²⁵⁴.

In this particular case, understanding how the rabies virus interferes with cell homeostasis opened the way for the development of new therapies. Our team envisioned the rational design of optimized peptides that bind with high affinity the PDZ domains of PTPN4 or MAST2, promoting neuron death or survival respectively^{253,255}. Combining structural and

biophysical data derived from peptides of viral origin, they designed peptides that efficiently trigger cell death or induce neurite outgrowth in various cell lines in an affinity-correlated manner. Neuro-regenerative peptides targeting the MAST2 PDZ domain are currently undergoing trials in animal models and showed successful results in regenerating the optic nerve after lesion.

3.2 Viruses targeting PTPN3

I previously described the involvement of the human phosphatase PTPN3 in various cellular processes (section 1.5.6). Although the precise role of this protein in cellular signalling has not been clearly established, it has been linked to the regulation of cell growth and homeostasis^{76,96,97,106}. Moreover, deregulation of PTPN3 activity is associated with various forms of cancer (section 1.5.7). It is therefore not surprising that oncoviruses target the PDZ domain of PTPN3 (PTPN3-PDZ). Here, I present an overview of the human papillomavirus (HPV) and the hepatitis B virus (HBV), which encode PBM-containing proteins that bind PTPN3-PDZ^{68,256}.

3.2.1 Human papillomavirus

3.2.1.1 Overview of the HPV life cycle

HPV is a small, non-enveloped and double-stranded DNA virus that belongs to the Papillomaviridae family (genre papillomavirus). HPV infections are transmitted through sexual contact and are the most common viral infection of the reproductive tract. Although most infections with HPV cause no symptoms, persistent genital HPV infection can cause cervical cancer in women. HPV can also trigger other types of anogenital cancer, head and neck cancers, and genital warts in both men and women. HPV is accountable for one-third of all tumours induced by viruses, which represents 5% of all human cancers²⁵⁷. Cervical cancer is the second most common cancer in women living in less developed regions with an estimated 570,000 new cases with approximately 311,000 women died from cervical cancer in 2018 according to the World Health Organization (WHO).

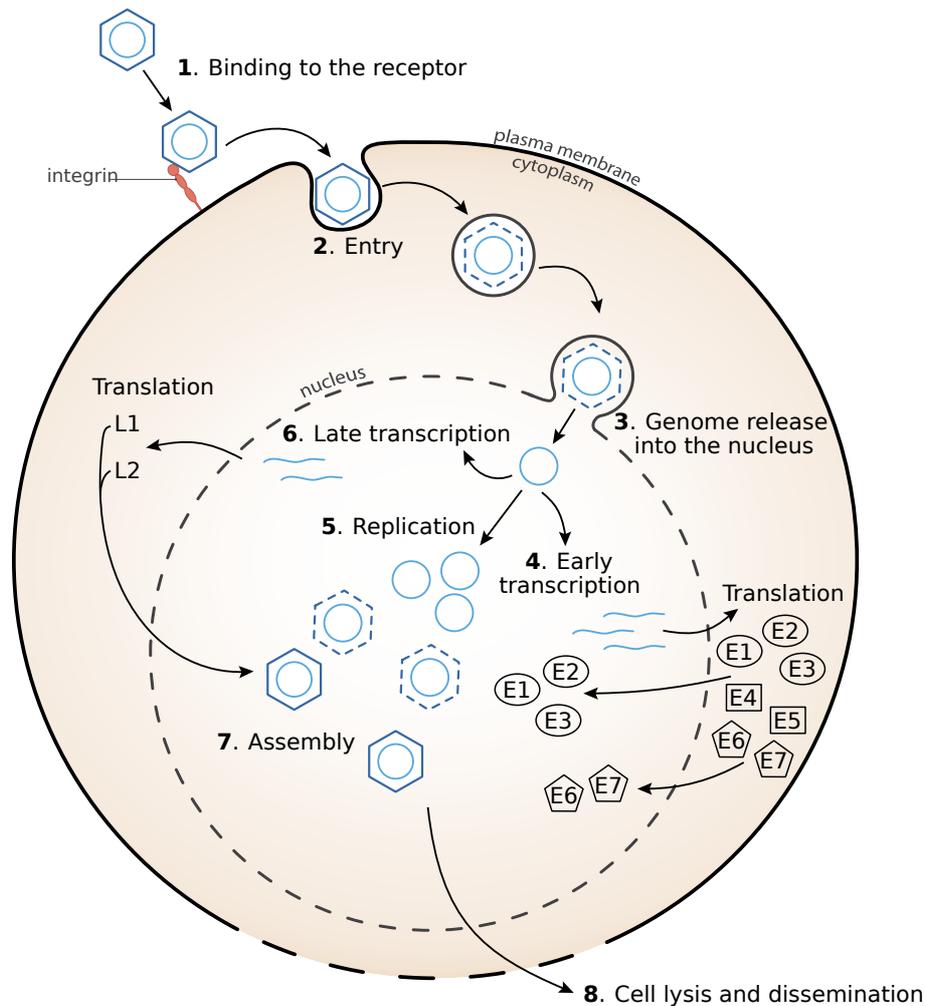


Figure 3.3: The essential steps in the HPV life cycle. The basic steps of the HPV life cycle are represented. The viral capsid is shown as a dark blue hexagon, and the viral genome as a light blue circle. E1, E2, E3, E4, E5, E6, viral ‘early’ proteins; L1, L2, viral ‘late’ proteins.

HPV presents a strict species specificity with an epithelial tropism. The HPV life cycle is intimately linked to the differentiation programme of its host tissue. The main steps of the HPV life cycle are illustrated in figure 3.3. HPV particles are composed of viral DNA and two proteins, L1 and L2, which form the icosahedral capsid. HPV infects basal keratinocytes via micro-abrasions. The attachment of the virus is mediated by interaction with components of the extracellular matrix, mainly heparin sulphate proteoglycans and possibly also laminin-5²⁵⁸. Structural changes in the virion capsid allow binding to a secondary receptor on the basal keratinocyte, which is required for internalization of the virus. The receptor used may vary according to the HPV genotype and the infected cell type²⁵⁹. Then, the virus is internalized through an endocytic mechanism (step 2), and subsequently undergoes endosomal transport,

uncoating, and cellular sorting. The L2 protein remains in complex with the viral DNA, while the L1 protein is targeted for lysosomal degradation²⁶⁰. The viral genome is then transported to the nucleus, where it may remain as an episome or, in the case of pre-cancerous and cancerous lesions, integrate into the host genome. The HPV genome encodes 8 open reading frames whose expression is closely coupled to the differentiation program of the host epithelium. As cells divide and progress to the upper strata of the epithelium, the viral ‘early’ E1, E2, E4, E5, E6, and E7 proteins, and ‘late’ L1 and L2 capsid proteins are expressed and the viral DNA is amplified (steps 4-7)²⁶¹. Eventually, the virus is released when the upper layers of the epithelial cells desquamate (figure 3.4).

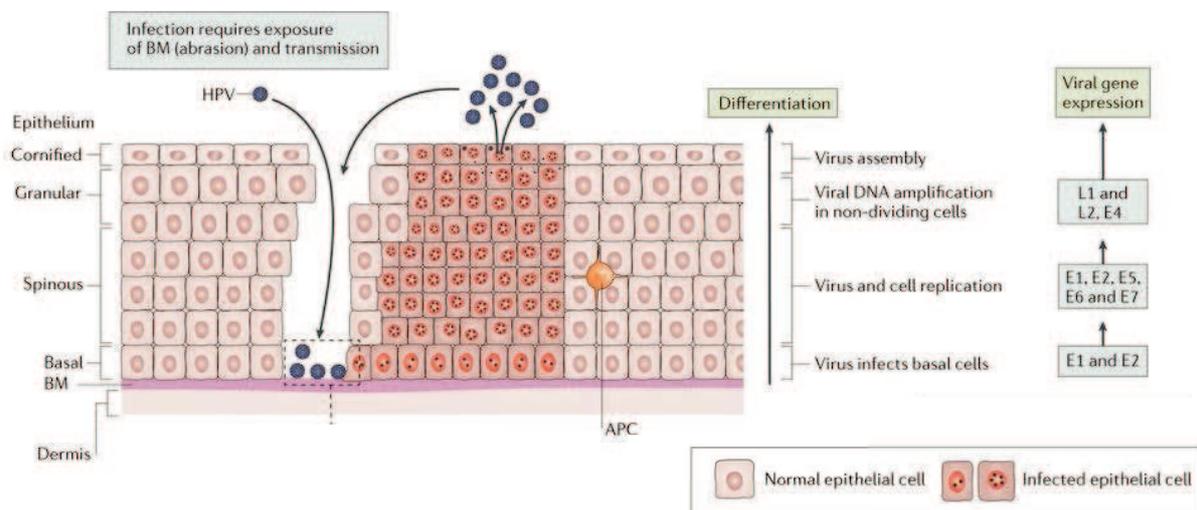


Figure 3.4: The HPV life cycle is coupled to the differentiation program of the host epithelia. HPV reaches the basement membrane (BM) through an abrasion, where it binds to surface receptors and is uptaken into the target basal keratinocyte. In the infected basal cells (which might include stem cells), the viral genome replicates and establishes episomes. The early viral proteins E6 and E7 stimulate the continued cell proliferation, while E1 and E2 drive viral genome replication to a very high copy number. Terminal differentiation of infected cells in the upper epithelial layers triggers the expression of E4 and then L1 and L2 to package the very high copy numbers of the viral genome. The virions are released as E4 disintegrates the cytokeatin filaments, and the keratinocyte remnants are sloughed off the epithelial surface. The viral life cycle is completed without directly causing cell death and without systemic viraemia or apparent inflammation to avoid alerting the local immune responses. APC, antigen-presenting cell. Adapted from Roden and Stern, 2018²⁶².

3.2.1.2 HPV-induced carcinogenesis

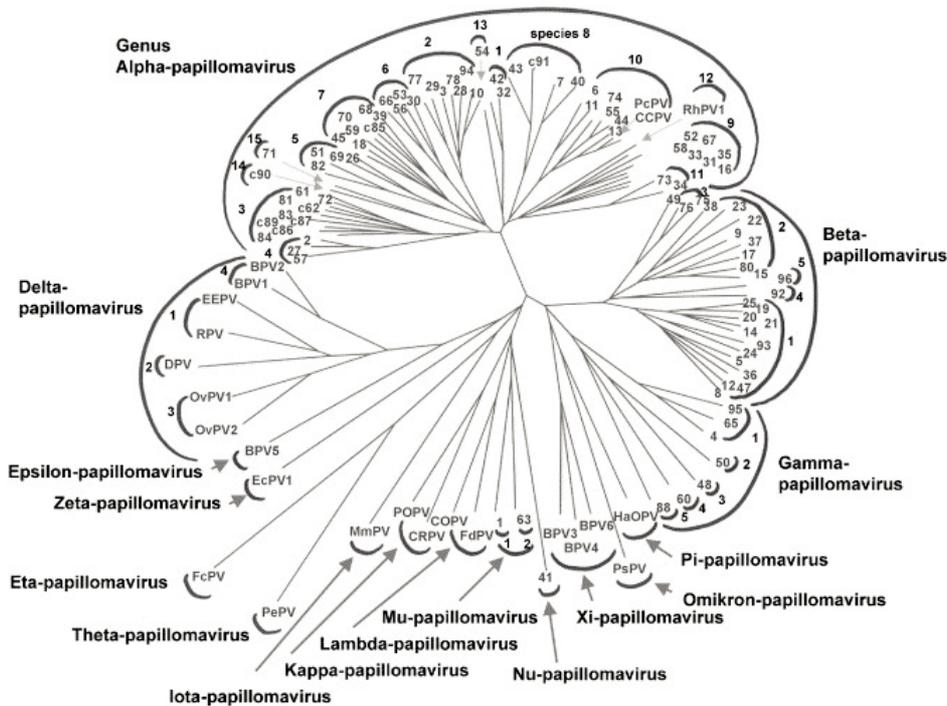


Figure 3.5: Phylogenetic tree of 118 papillomavirus types, built based on the L1 ORF sequences. The numbers at the ends of each of the branches identify an HPV type; c-numbers refer to candidate HPV types. All other abbreviations refer to animal papillomavirus types. The outermost semicircles identify papillomavirus genera, e.g. the genus alpha-papillomavirus. The number at the inner semicircles refers to papillomavirus species. From de Villiers *et al.* 2004²⁶³.

HPVs are highly heterogeneous and are broadly classified in five evolutionary groups: the α , β , γ , μ , and ν papillomaviruses (figure 3.5). α -HPV is the largest group and will be the subject of this description. The members of this group are able to specifically infect mucosae, including the cervical, oral and penile mucosae, and can be further divided into two subgroups depending on their pathogenicity. Low-risk HPVs (types 6, 11, 40, 42, 43, 44, 53, 54, 61, 72, 73 and 81) do not cause malignancies and are normally associated with benign epithelial condylomas, while high-risk HPVs (types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68) cause upper-respiratory and ano-genital lesions that can progress to carcinomas. The two principal high-risk types (HPV16 and 18) are together responsible for over 80% of all HPV-related cervical cancers²⁵⁷. The early proteins E5, E6, and E7 encoded by the virus stimulate host cell proliferation and survival, and modulate cell differentiation; therefore, these are

considered as oncoproteins. Sustained expression of E6 and E7 is one of the hallmarks of HPV-induced carcinogenesis²⁶⁴, and thus the transforming activities of these two proteins have been more characterised than that of E5. Recent studies, however, highlight the relevance of E5 in cell transformation and immune modulation^{265,266}. E6 and E7 are expressed from a common promoter and work individually and cooperatively to exert their oncogenic activity. The main functions of E6, E7, and the two acting together, as well as the cellular targets implicated in these functions, are detailed in section 9.3.

3.2.1.3 The HPV E7 oncoprotein

E7 (98 residues) was the first HPV oncoprotein to be identified. This protein recognises a LXCXE motif (where X is any amino acid) in cellular partners that mediates association to its targets. E7 has been involved in cell cycle deregulation, immune system modulation, cell invasion and genomic instability²⁶⁷. The main cellular partner of E7 is the cell cycle regulator retinoblastoma protein (pRb). E7 induces pRb ubiquitination and targets it for proteasomal degradation, which ultimately promotes G1/S phase transition. E7 activity maintains cells in an S-phase-competent state that is crucial for the viral life cycle²⁶⁸. Moreover, E7 can interact with histone deacetylases to modulate transcription of host genes, and regulates the activity of cyclin dependent kinases to alter the cell cycle^{267,268}.

3.2.1.4 The HPV E6 oncoprotein

E6 is a multifunctional protein that interacts with multiple host proteins to deregulate cellular functions. Similarly to E7, E6 recognises acidic leucine-rich peptides with the consensus sequence LXXLL. In α -HPV, recognition of this motif primarily mediates the association of the ubiquitin ligase E6 associated protein (E6AP), which is required for one of E6's major transforming activities: degradation of the host tumor suppressor p53²⁶⁹. Other partners of the high-risk E6-E6AP complex include proteasome subunits and the p300 and CBP coactivators²⁶⁴. In most cases, but not always, interaction with E6 results in degradation of the interacting partners by the proteasome.

This way, E6 promotes the development of cancer phenotypes such as escaping cell death, deregulating the cell cycle, modulating the host immune system, promoting cell immortalization, increasing genomic instability and stimulating cell invasion²⁶⁴. Other functions, such as promoting inflammation and angiogenic switch, are achieved by coordinated action with E7²⁶⁷.

One of the most evident differences between high- and low-risk HPV strains is the presence of a class I PBM (S/T-X- ϕ_{COOH}) on the C terminus of E6 in the high-risk strains. This PBM is absent in low-risk strains, and therefore it is considered as a molecular signature of the carcinogenic potential of an HPV strain²³⁰. Indeed, a diagnostic assay based on the binding of high-risk HPV-E6 oncoproteins to PDZ domains has been developed accordingly (OncoE6TM Cervical Test, Arbor Vita Corporation). Multiple studies in human cell lines, primary keratinocytes and mice highlight the relevance of the high-risk E6 PBM in the development of cancer phenotypes such as altered differentiation, epithelial to mesenchymal transitions and anchorage independent colony formation^{230,264}. Known host PDZ partners of the high-risk E6 PBM, among which we find PTPN3, are listed in table 3.2 and will be discussed in section 3.2.1.7.

Table 3.2: Known PDZ domain-containing targets of E6. Adapted from Ganti *et al.* 2015²³⁰.

Protein Name	Function
DLG1	Polarity/tumour suppressor
SCRIB	Polarity/tumour suppressor
MAGI1/2/3	Polarity/tumour suppressor
PSD95	Signaling complex scaffold
TIP2/GIPC	TGF- β signaling
NHERF1	PI3K/AKT signaling
MUPP1	Signaling complex scaffold
INADL/PATJ	Tight junction assembly
PTPH1/PTPN3	Protein tyrosine phosphatase
PTPN13	Non-receptor phosphatase
PDZRN3/LNX3	RING-containing ubiquitin ligase
14-3-3 ζ	Signaling complex adapter
PAR3	Polarity/tumour suppressor
SNX27	Endosomal trafficking/signaling
ARHGEF12	RhoGEF
FRMPD2	Tight junction formation
LRRC7	Normal synaptic spine function
GOPC	Vesicular trafficking

3.2.1.5 Structure of E6

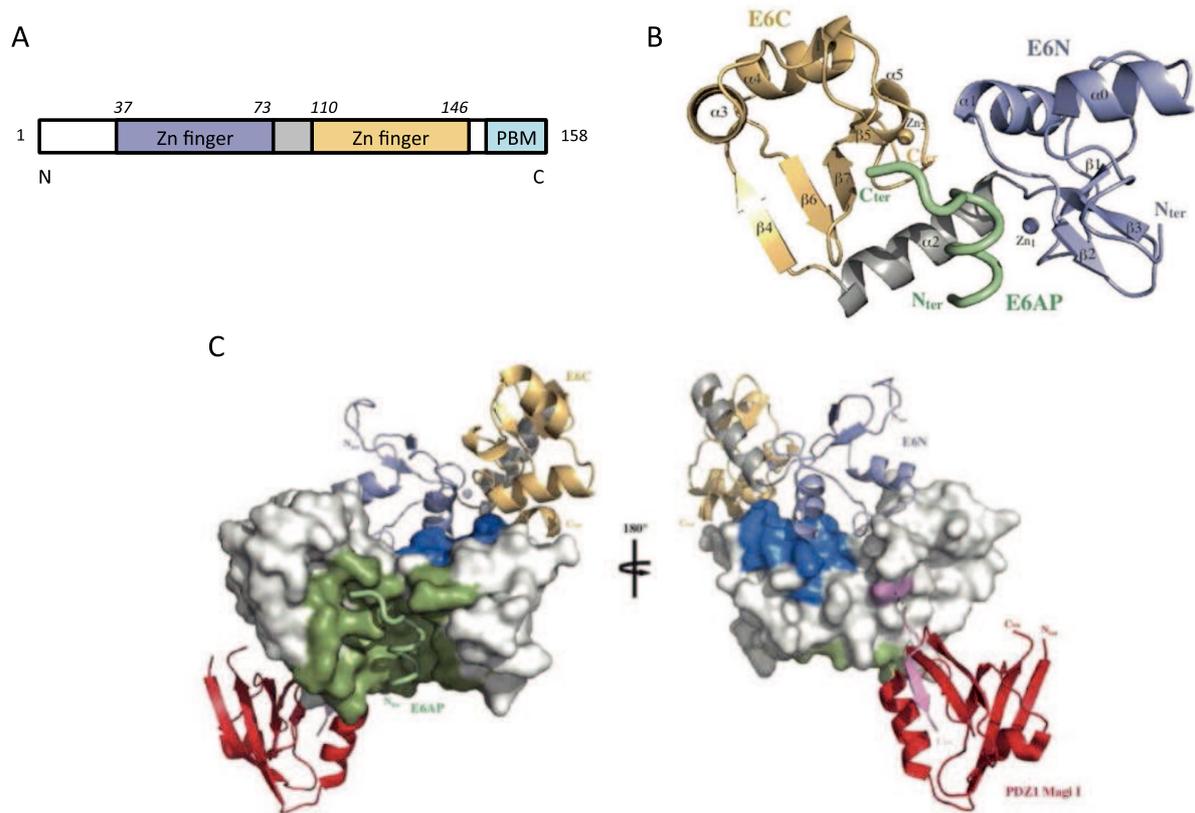


Figure 3.6: Structure of the high risk HPV E6 protein. (A) Schematic of the functional domains of high-risk HPV E6. Numbers in *italic* above the schematic indicate the boundary residues of each protein domain. The zinc finger domains mediate interactions with LXXLL motif-containing proteins. The PDZ-binding motif is located at the extreme C terminus of high-risk E6. (B) X-ray structure of HPV16 E6 bound to the LXXLL motif of E6AP. The interdomain linker helps form a binding pocket for the association with cellular targets, such as E6AP, which subsequently targets p53 for degradation. Violet: N-terminal domain (E6N); grey: linker helix; gold: C-terminal domain (E6C); green: E6AP LXXLL peptide. (C) Mapping HPV16 E6 functional regions. LxxLL binding and E6 self-association residues are colored green and blue respectively. The relative orientation of E6 and the PDZ domain shown is arbitrary. Surfaces colored white are potentially available for binding to p53. Adapted from Zanier *et al.* 2013²⁷⁰.

E6 proteins are cysteine-rich proteins of approximately 150 residues that feature two zinc finger domains joined by a short helical linker (figure 3.6.A). Although the structure of the isolated zinc finger domains was determined by solution NMR^{271,272}, full-length E6 self-associates into inclusion bodies when expressed in bacteria²⁷³, which prevented its structural studies. Zanier *et al* circumvented this problem by using an HPV16 E6 mutant with a decreased tendency to oligomerize, and further increased its solubility by forming a complex with an E6-binding LXXLL sequence fused to a crystallization-prone mutant of the bacterial Maltose Binding Protein²⁷⁰. The resulting structure shows that the LXXLL ligand binds in a

pocket formed between the two zinc-binding domains and the helical linker. The two domains and the linker establish contacts with the peptide, which also adopts a helical conformation (figure 3.6.B). The global fold of the complex appears to be maintained by the peptide, as only a few interactions connect directly the two domains and the linker. The functional regions of E6 do not overlap: the C-terminal PBM and the self-association surface, which is required for p53 binding²⁷¹, are located on the opposite side of the peptide binding pocket (figure 3.6.C)²⁷⁰. The multiple interactions in which E6 can engage define the role of this protein and its activities in the viral life cycle.

3.2.1.6 Relevance of E6 for the viral life cycle

Studies of the HPV life cycle in human cell lines and mice have highlighted the key role of E6, and in particular of its PBM, in the viral life cycle. In Human Foreskin Keratinocytes (HFKs) transfected with the high-risk HPV types 31²⁷⁴ or 18²⁷⁵, it was observed that presence of E6 was required for stable replication of the viral genome. Extended passaging of cells containing viral genomes with a deletion of the E6 PBM (E6ΔPBM) resulted in a loss of these mutant genomes, indicating that the PBM is relevant for the episomal maintenance of the viral DNA²⁷⁵. Similar effects were observed in Normal Immortal Keratinocytes containing HPV16 E6ΔPBM genomes²⁷⁶. The levels of E6ΔPBM expression correlated with defects in the viral life cycle. Interestingly, the loss of the capacity of HPV16 E6 to interact with PDZ domain-containing proteins results in enhanced proteasomal degradation of E6, and the PBM-mediated interaction with SCRIBBLE, MAGI and DLG stabilized E6, indicating that some of E6's PDZ partners might be involved in regulating E6 expression or stability²⁷⁶. These observations support the essential role of the E6 PBM in the viral life cycle and genome maintenance.

3.2.1.7 Role of the E6 PBM in transformation and cancer

The relevance of PDZ protein targeting by HPV E6 oncoproteins in the transformation capacity of HPV has been demonstrated both *in vivo* and *in vitro*. For example, HPV18 E6ΔPBM transfected keratinocyte cell lines showed decreased epithelial to mesenchymal transition, changes in the actin cytoskeleton, and aberrant adherens junctions, compared to those expressing wild type E6²⁷⁷. Moreover, anchorage independent growth²⁷⁸ and cell immortalization²⁷⁹ have also been linked to the presence of the E6 PBM. In transgenic mice models, those expressing HPV16 E6 in the epidermis develop epithelial hyperplasia that

evolves to squamous cell carcinoma, while those expressing a PBM deletion mutant do not progress to hyperplasia²⁸⁰. This indicates that a functional E6 PBM is required for cell transformation. Further analyses highlighted the cooperative activity of E6 and E7 for cancer development *in vivo*, and suggested that E7 could be primarily responsible for the development of tumors, while E6 could play a more important role in later stages of tumor progression²³⁰.

3.2.1.8 E6 targets cell adhesion and polarity proteins

Many proteins that regulate cell polarity and adhesion are PDZ proteins. High-risk HPV E6 targets major components of the three cell polarity control complexes, SCRIBBLE, PAR and CRUMBS, as well as the MAGI group of cell junction proteins²³⁰ (figure 3.1).

E6 has been shown to bind SCRIBBLE and DLG1 at lateral adherens junctions and subsequently targets them for degradation *in vitro* and *in vivo*²⁸¹⁻²⁸³. Interestingly, E6 appears to exclusively target a cellular subset of the pool of SCRIBBLE and DLG1 for degradation. For example, HPV18 E6 only degrades phosphorylated and nuclear DLG1, while the membrane bound forms are not affected²⁸⁴. This could be related to the pro-oncogenic activity that DLG1 and SCRIBBLE exhibit under certain conditions. DLG1 re-localization by Adv9 to the plasma membrane constitutively activates the PI3K proliferative pathway²⁴⁰, while SCRIBBLE can act as an oncoprotein in breast cancer, where it is overexpressed and mislocalized²⁸⁵.

MAGI proteins are found at subapical tight junctions, and are also targeted for degradation by HPV E6 oncoproteins^{283,286}, which results in metastatic phenotypes such as loss of TJ integrity and stability, increased proliferation and deregulation of apoptosis²³⁰. By contrast, restoration of MAGI-1 levels in HPV-positive cells by blocking its targeting for degradation by E6 restored TJ assembly, triggered a reduction in cell proliferation, and induced apoptosis²⁸⁷. These observations place MAGI-1 within the *bona fide* targets of E6.

Finally, E6 also targets for degradation the CRUMBS complex protein PATJ, which is essential for the localization of TJ proteins and cell polarity^{288,289}. The *in vivo* association of PATJ with full-length E6, assessed by pull down experiments on 293T cells, was dependent on the E6 PBM and caused a decrease in PATJ levels. Notably, it was also observed that an alternatively spliced isoform of HPV 18 E6* that lacks the PBM is also able to associate with PATJ and promote its degradation, possibly through a different interaction surface of E6, or by involving another cellular protein to mediate the interaction²⁹⁰. The fact that two isoforms

of E6 can target PATJ through different mechanisms, both resulting in a decrease in PATJ levels, indicates that this is a relevant target of HPV. However, more studies will be needed to understand the physiological effects of PATJ downregulation by E6.

3.2.1.9 Other targets of the E6 PBM

Several other PDZ proteins have been reported to bind the E6 PBM, although less is known about their role in the HPV life cycle and induced oncogenesis (table 3.2).

MUPP1 (multiple PDZ-containing protein) is a 13 PDZ domain-containing protein that is thought to be involved in the assembly of signalling complexes and is concentrated in TJs in polarized epithelial cells²⁹¹. MUPP1 is a common target of two viral oncoproteins, AdV9 E4-ORF1 and HPV E6²⁹². In both cases, the interaction of MUPP1 with the viral protein results in its functional inactivation: E4-ORF1 abnormally sequesters MUPP1 in the cytoplasm, while E6 targets it for degradation²⁹².

TIP-1 and TIP-2/GIPC1 were both identified as partners of E6 by yeast two-hybrid assays and by co-immunoprecipitation. These E6 binders are also shared with another oncovirus: they are partners of the HTLV-1 Tax protein. Interaction with E6 was proposed to produce a gain-of-function effect on TIP-1, which was in turn thought to increase cell motility by increasing GTP RhoA activity²⁹³. On the contrary, TIP-2/GIPC1 is targeted for proteasomal degradation by E6, which could benefit HPV by rendering cells less sensitive to cytostatic TGF- β signalling. GIPC1 is also known to interact with the HBV capsid protein, but the functional effects of this association are unknown²⁹⁴.

PTPN3 is also a target of HBV and HPV. Binding the E6 PBM targets PTPN3 for proteasomal degradation, and endogenous levels of PTPN3 are low in HPV-positive cervical carcinoma cell lines. siRNA-mediated depletion of PTPN3 reduced growth factor requirements of immortalized keratinocytes²⁵⁶, so it was proposed that targeting PTPN3 may promote the growth of HPV-positive cells, but further studies will be necessary to elucidate the effects of E6-dependent PTPN3 degradation.

The remaining E6 PBM partners have not been described as targets of other oncoviruses. The scaffolding protein PSD95 is targeted for degradation by E6²⁹⁵. PSD95 is mainly found in postsynaptic densities in neuronal tissues, but is also found in lower levels in other tissues, including cervix and skin normal keratinocytes. Then, the non-receptor PTPN13 is also degraded by interaction with E6, which was correlated to an increased capacity of keratinocytes for anchorage-independent growth²⁷⁸. The last target of the E6 PBM, also

subjected to degradation, is the cystic fibrosis transmembrane regulator-associated ligand (CAL), which is supposed to take part in vesicular trafficking processes²⁹⁶.

3.2.1.10 Therapeutic value of HPV oncoproteins

E6 and E7 are constitutively expressed in HPV-derived cancer and are required for the onset and maintenance of the malignant phenotype. Several therapeutic strategies to decrease HPV-mediated tumorigenesis have been envisioned to functionally inhibit E6 and E7. E6 and E7 gene expression have been blocked by specific siRNAs, significantly reducing HPV-positive cells growth and colony formation, and increasing apoptosis²⁹⁷. Additionally, E6 activity can be blocked by inhibitors such as small peptides, peptide aptamers, intracellular antibodies or small molecules²⁹⁸. Several of these approaches to target E6 function resulted in induction of apoptosis in HPV-positive cancer cell lines, in agreement with E6 anti-apoptotic role²⁹⁸. A significant advancement came from solving the structure of E6 E6/E6AP/p53 complex^{270,299}. This allowed identification of a E6AP binding pocket that can be targeted with high-affinity peptides to inhibit p53 degradation and induce apoptosis in HPV16-infected cells^{300,301}. Targeting E7 activity by using antagonist peptides to reactivate the tumor suppressor pRb also showed antitumor effects in cell lines and animal tumor models³⁰². Other approaches that have been proposed imply inhibition of E6 and E7 transcription or gene expression²⁹⁸.

E6 and E7 are also well-defined tumor-specific antigens that have been considered as the theoretical basis for the development of immunotherapy³⁰³. However, design of such therapies faces some obstacles, as HPV is able to modulate the host immune system and induce immune evasion mechanisms, such as inhibiting the antigen processing and presentation machinery³⁰⁴.

3.2.2 Hepatitis B Virus

HBV is a global public health problem, with 257 million chronic HBV carriers and almost 1 million deaths estimated in 2015 (WHO). Despite available prophylactic vaccines, chronic hepatitis B is still incurable; existing treatments focus on stopping the spread of the virus to prevent the progression of the disease. The infection is characterized by a usually asymptomatic acute phase, which can lead to chronicity in certain cases, resulting in the risk of advancing slowly to cirrhosis and hepatocellular carcinoma (HCC).

3.2.2.1 Overview of the HBV life cycle

HBV is a small, enveloped DNA virus that packages a circular double strand DNA genome with gaps in both strands, hence termed relaxed circular DNA (rcDNA). Its icosahedral nucleocapsid is composed by the HBc antigen or core protein (in the following, HBc). The viral life cycle, which has been reviewed in detail by several groups^{305–307}, is represented in figure 3.7. Briefly, HBV enters the cell endocytically by association of its surface glycoproteins with the Sodium Taurocholate Co-transporter Polypeptide (NTCP). In the cytoplasm, the virus core is released from the endosome and is directed to the nuclear pore complex by nuclear localization signals (NLS) present in HBc³⁰⁸. Association with the nuclear pore complex is thought to be linked to HBc dissociation and genome release into the nucleus, where it is turned into covalently closed circular DNA (cccDNA) by host enzymes and decorated with nucleosomes. This episome contains four open reading frames that are transcribed by the host RNA polymerase II. A full-length mRNA, the pre-genomic RNA (pgRNA), encodes the capsid protein HBc and the reverse transcriptase P. Smaller ‘sub-genomic’ mRNAs encode the surface proteins HBsAg, and the regulatory HBx protein, which is thought to affect cccDNA persistence and regulate its transcription. In the cytoplasm, P binds to the pgRNA to form a complex that initiates RNA encapsidation by HBc. Inside the capsids, the pgRNA is reverse transcribed to rcDNA. The mature capsids may then associate with HBsAg bound to endoplasmic reticulum membranes to be enveloped and secreted, or bind importin molecules to be transported back into the nucleus to maintain the levels of cccDNA. The HBc capsid protein is a keystone with functions that extend well beyond its structural role. HBc has been implicated in most stages of the viral life cycle, including subcellular trafficking and release of the viral genome, pgRNA encapsidation, and reverse transcription³⁰⁹

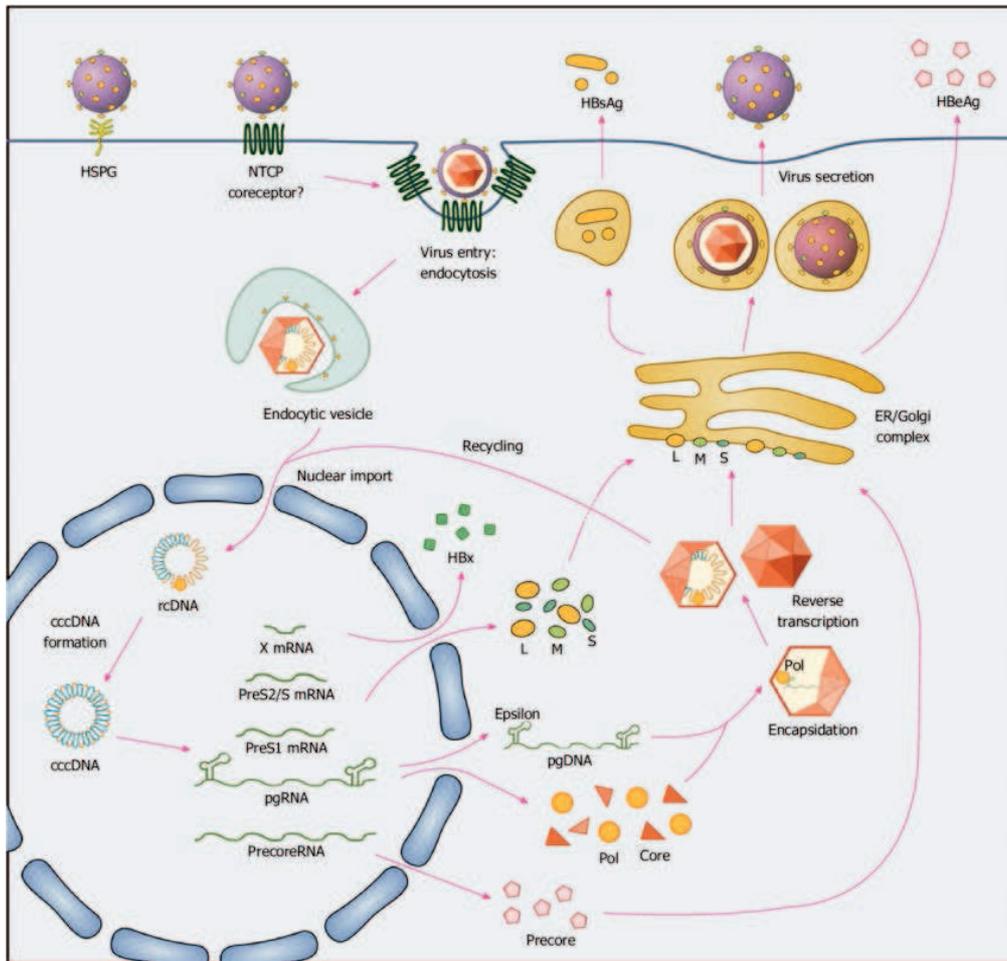


Figure 3.7: The HBV life cycle. Hepatitis B virus (HBV) binds to the surface receptor NTCP and enters hepatocytes, and its genome is released into the nucleus. The relaxed circular HBV genome (rcDNA) is transformed into covalently closed circular DNA (cccDNA), which serves as a template for the transcription of viral mRNAs (pregenomic (pg), precore, HBx, PreS1, and PreS2/S RNA). The HBV mRNAs are translated into the large (L), middle (M), and small (S) surface, precore, core, polymerase (Pol), and HBx proteins. pgRNA and Pol are encapsidated, and viral DNA is reverse-transcribed inside the capsid. Assembled HBV virions are secreted from hepatocytes. Some capsids are redirected to the nucleus to maintain the levels of cccDNA. Adapted from Kim, Kang and Kim, 2016³¹⁰.

3.2.2.2 Viral capsid assembly and HBc structure

HBc presents distinct N- and C-terminal domains (figure 3.8 HBc protein domain organization, phosphorylation sites and sequence features). The first 149 residues fold into a predominantly α -helical conformation that is known as the assembly domain (in the following, HBc₁₄₉), as this region is capable of self-assembling into morphologically normal capsids³¹¹. The remaining portion of the protein forms the arginine-rich RNA-binding C-terminal domain (CTD), which is dispensable for capsid assembly *in vitro* but is required for viral genome replication in cells³¹¹. The CTD also contains nuclear localization signals and

cytoplasmic retention signals that regulate nuclear import and export of the HBV capsid³¹², and multiple serine and threonine residues that are susceptible of being phosphorylated³¹³. Finally, the C-terminal residues of the CTD comprise a non-canonical class I PBM featuring a C-terminal cysteine (-RESQC_{COOH}, figure 3.8), that has been shown to interact with the PDZ domains of PTPN3⁶⁸ and GIPC1²⁹⁴.

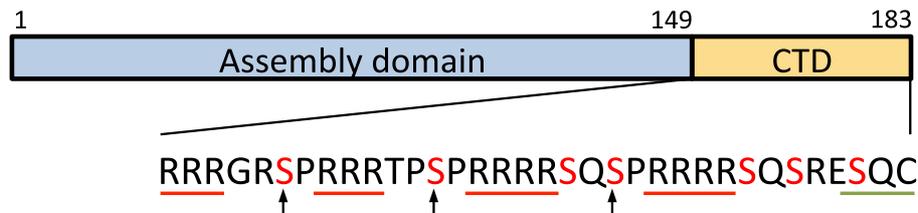


Figure 3.8: HBc protein domain organization, phosphorylation sites and sequence features. Schematic of the core protein sequence showing the assembly domain (residues 1-149) and the C-terminal domain (CTD, residues 150-183). The CTD contains 7 serines which are potential phosphorylation sites by SRPK (shown in red). *In vivo*, phosphorylation of S155, S162, and S170 (indicated by arrows) confers specificity for pgRNA packaging. The arginine clusters that are required for pgRNA encapsidation and reverse transcription are underlined in red. The C-terminal PBM is underlined in green. Adapted from Chen, Wang and Zlotnick 2011³¹⁴.

The crystal structure of the HBV capsid was first solved by Wynne and colleagues in 1999³¹⁵. HBc, the building block of the capsid, is a 183-residue, 21 kDa helical protein with 5 α -helices and an unfolded C-terminal region enriched in Arg residues (figure 3.8.A). Helices 1, 2 and 5 form the hydrophobic base of the capsid. Helices 3 and 4 form an alpha helical hairpin that acts as the dimerization surface. The hairpins of two monomers assemble into a 4-helix bundle, in an interaction driven by burial of hydrophobic surface, and stabilized by salt bridges and H-bonds³¹⁶. Dimers can then self-assemble into icosahedral capsids. The major assembly product ($\approx 95\%$) is a 120-dimer 22 nm capsid (figure 3.8.C), although a small number of dimers assemble into 90-dimer capsids³⁰⁶. The capsid assembly is entropically driven, and the contacts between dimers mainly involve hydrophobic interactions³¹⁷. Tyrosine 132 is particularly important at this interface. In wild type HBc dimers its side chain is completely exposed, and it gets buried when dimers assemble into a capsid, contributing about 10% of the buried surface³¹⁵. In fact, mutating this residue to alanine (Y132A) yields the protein assembly-deficient but still able to form dimers³¹⁸, which has been the basis of HBc dimer structural studies³¹⁹⁻³²¹. Crystal structure comparison of HBc₁₄₉ (construct from aminoacid 1 to 149) and HBc₁₄₉-Y132A dimers revealed globally conserved tertiary

structures, with minor changes occurring at the spike tips and a more compact conformation for the mutant dimer³⁰⁶ (figure 3.9.B X-ray structures of the HBc assembly domain).

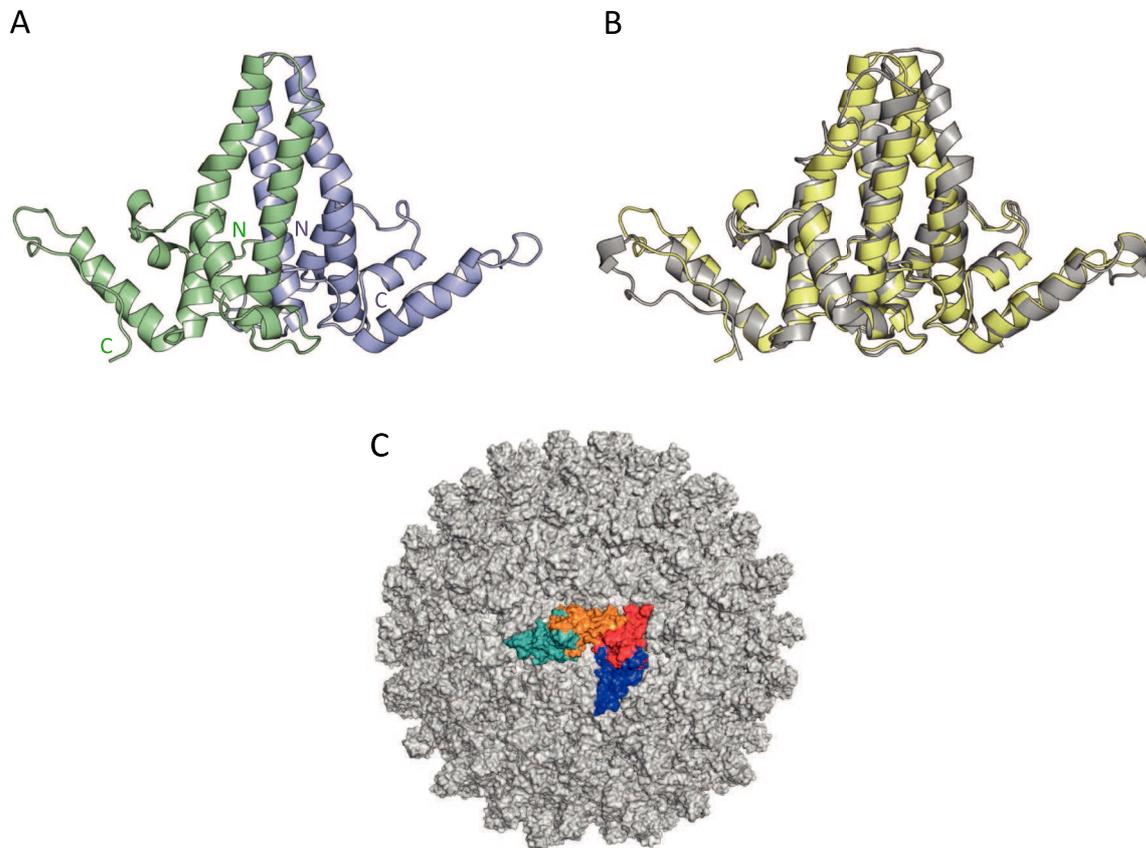


Figure 3.9: X-ray structures of the HBc assembly domain. (A) HBc₁₄₉ dimer in the context of the viral capsid (PDB ID 1QGT). The N and C terminus of each monomer are indicated. (B) Superimposition of capsid HBc₁₄₉ dimer (yellow) and a free HBc₁₄₉ Y132A dimer (grey). (C) 120-dimer HBc₁₄₉ capsid, with the asymmetric unit in colour. The individual subunits are A-B (teal-orange) and CD (blue-red) dimers.

HBc plays a central role in the viral life cycle, being required for viral genome encapsidation, maturation, and subcellular trafficking of the capsid. Its involvement in these diverse processes is described below.

3.2.2.3 Nucleic acid packaging and CTD accessibility

The HBc CTD is responsible for binding and encapsidating viral pgRNA. As previously mentioned, once the viral genome is matured to rcDNA inside the capsids, a fraction of the rcDNA-filled capsids are shuttled back to the nucleus to maintain the levels of cccDNA (figure 3.7). This is achieved thanks to NLSs present in the HBc CTD. To perform

these functions, the CTD should be able to shuttle between the interior and the exterior of the capsids. Exposure of the CTD would provide a mechanism for HBc to interact with host proteins involved in capsid trafficking, maturation or disassembly. Although the accessibility of the CTD was initially questioned, accumulating evidences support this model. In a recent study in which the high-resolution structure of HBV capsids was determined by cryo-EM scanning³²², some capsids were found to display the CTD in their exterior, while in others it was in the interior, bound to RNA (figure 3.10). *In vitro* studies suggested that the CTD exposure could be dependent on the nucleic acid content of the capsid, which reflects its maturation level³¹⁴. Specifically, it was shown a host protein that binds the CTD of empty full-length HBc capsids and cannot bind HBc₁₄₉ capsids did not bind RNA-filled full-length capsids, indicating that the nucleic acids inside the capsid regulate the dynamic exposure of the CTD³¹⁴. This is interesting since it indicates a correlation between CTD exposure and viral core maturation, which could provide a mechanism to distinguish mature cores from empty ones and to regulate HBV intracellular transport. Moreover, it suggests that a host protein could bind the CTD and acts as a chaperone to regulate capsid assembly³¹⁴.

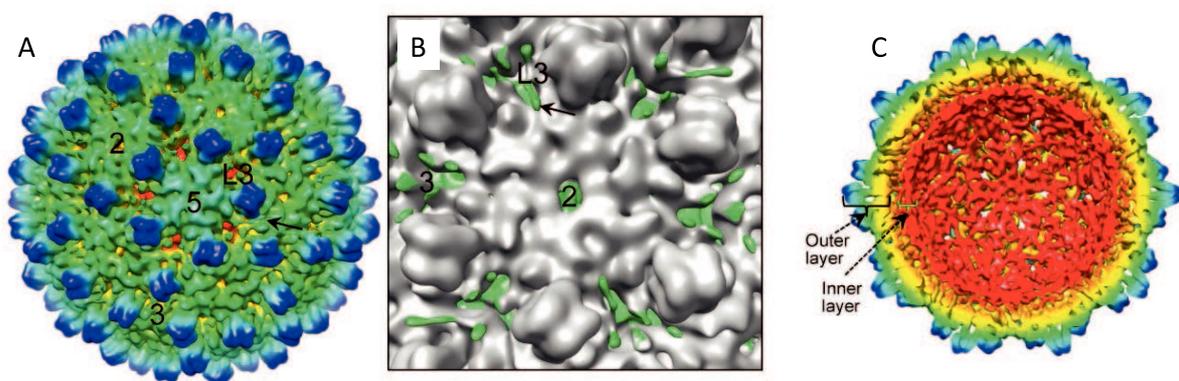


Figure 3.10: Direct observation of the HBc CTD segment protruding outside the core. (A) Radially-colored, shaded surface representation of the HBV reconstruction filtered to 10Å resolution as viewed along a five-fold axis, showing some densities of the CTD tails on the exterior of core (arrow). The two-fold, three-fold, local three-fold and five-fold axes are indicated by the labels 2, 3, L3, and 5, respectively. The density map is contoured at 1σ above the mean. (B) An enlarged view around a two-fold axis, showing a density protruding outside the core through the local three-fold axis (arrow). The ordered outer layer is colored in gray using atomic model, while the rest of density map is colored in green. The map is contoured at 1σ above the mean. (C) Inside view of the radially colored 3D reconstruction map, showing the full-length HBV core has a double-layer structure: an outer layer composed of the N-terminal core assembly domain (residues 1-149) and an inner layer composed of the basic CTD (residues 150-185) and its bound RNAs. The map is contoured at 2σ above the mean. Adapted from Yu *et al.* 2013³²².

3.2.2.4 HBc and reverse transcription

The HBV genome is reverse transcribed from pgRNA into rcDNA in a complex, multi-step process that involves genome rearrangements and template switches³²³. It has been shown that the HBc CTD is required at multiple stages of this process, and HBc participates as well in the context of the formed capsid.

HBc₁₄₉ can form capsids but fails to encapsidate pgRNA³¹¹. Mutational and mechanistic analyses showed that the arginine clusters located in the CTD (figure 3.8) are required for both pgRNA encapsidation and all steps of reverse transcription³²⁴⁻³²⁶, indicating that the CTD has pleiotropic roles in viral replication. Similar observations were made when mutating phospho-acceptor sites in the CTD, which suggests that the CTD must be phosphorylated to correctly encapsidate pgRNA and reverse transcribe it to rcDNA³¹³. Lewellyn and Loeb proposed that the CTD of HBc participates in reverse transcription by acting as a nucleic acid chaperone³²⁵, which would explain how this small region mediates complex template switches and genome rearrangements. The nucleic acid chaperone activity of the CTD was later confirmed *in vitro*³²⁷.

The assembly domain HBc₁₄₉ is also believed to contribute to reverse transcription. Affecting capsid assembly by introducing mutations in the assembly domain resulted in capsids that support minus strand DNA synthesis but fail to support second strand synthesis, demonstrating that the capsid is more than just an inert container for the viral genome³²⁸.

3.2.2.5 HBc CTD interactions with host proteins GIPC1 and PTPN3

Although it is clear that HBc can interact with host proteins, there is a significant gap of knowledge in this topic, and only a few of these interactions have been described. The most evident of these interactions are mediated by the NLS in the CTD of HBc, which can bind importin $\alpha\beta$ complexes and target the capsid to the nuclear pores³¹².

It is remarkable that despite the known relevance of viral PBMs, the one of HBc has received such little attention. Only two human PDZ proteins are known to date to bind this PBM: GIPC1 and PTPN3^{68,294}. GIPC1 is thought to be an adaptor protein involved in G protein signalling³²⁹. Its PDZ domain is targeted by at least two other viral proteins – Tax protein of HTLV-1³³⁰, and E6 protein of HPV18³³¹. It is still not clear which role GIPC1 plays in the HBV life cycle.

PTPN3 was found to be a partner of HBc in a yeast two-hybrid screening. The interaction was confirmed to be mediated by the PDZ domain of PTPN3 by pull-down

experiments⁶⁸. To assess possible effects of PTPN3 in HBV gene expression, Hsu and colleagues overexpressed PTPN3 in human hepatoma HuH-7 cells that were transiently co-expressing a plasmid with more than a unit of the HBV genome. They observed that PTPN3 overexpression caused a decrease in viral RNA levels and that siRNA-mediated endogenous PTPN3 silencing resulted in an increase of viral RNA levels in human embryonic kidney 293T cells⁶⁸. The decrease of the viral genome expression was independent of the PDZ-mediated interaction, but dependent on the correct subcellular localization of PTPN3 and its catalytic activity, as evidenced in experiments with deletions of PTPN3 FERM domain or catalytically inactive PTPN3 mutants⁶⁸. It remains unclear how these effects can be of benefit for the HBV life cycle.

3.2.2.6 HBc directed antivirals

Given its deep involvement in the HBV life cycle, HBc has attracted much attention as a potential target for the development of antiviral agents. In the early 2000s, molecules targeting the viral capsid assembly were discovered^{332,333}. The mechanism of action of these inhibitors is based mainly on allosteric modulation, and were thus termed core allosteric modulators (CAMs)³³⁴. Despite their activity against nucleoside analogue resistant strains and various HBV genotypes, they have not reached the clinic phase, probably due to safety or pharmacological concerns³⁰⁹. Since their discovery, many other CAMs have been obtained, and they are known to prevent transcription of pgRNA by preventing encapsidation or by promoting formation of morphologically normal, nucleic acid-free capsids or by promoting formation of aberrant capsids. CAMs have also been shown to prevent cccDNA formation, possibly by inhibiting rcDNA transport to the nucleus³⁰⁹. Currently, several CAMs are undergoing clinical trials, and others are being subjected to pre-clinical testing.

4 Project objectives

Protein phosphatases play a major role in cell signalling. In conjunction with kinases, they regulate the levels of phosphorylated species in signal transduction pathways that control cellular processes such as growth, differentiation, migration, survival and apoptosis. Large-scale genetic analyses of human tumours have highlighted the relevance of protein tyrosine phosphatases either as putative tumour suppressors or as candidate oncoproteins³³⁵.

The human protein tyrosine phosphatase PTPN3 is a modular protein comprising a N-terminal FERM domain that determines subcellular localization, a central PDZ domain involved in protein-protein interactions and a C-terminal PTP domain able to dephosphorylate protein substrates. PTPN3 is known to be implicated in regulation of cell cycle and growth, and dephosphorylates key regulators heavily involved in the progression of human cancers⁶⁶. Defects of its expression and/or mutations have been suggested to play a role in many cancers³³⁶. However, only a few of its cellular partners and substrates have been identified and the role of PTPN3 in cell signalling remains unclear.

In addition, viral proteins of oncoviruses, such as the capsid protein of HBV⁶⁸ and the E6 protein of high-risk HPV types 16 and 18⁶⁷, are able to interact with PTPN3 through a PDZ-binding motif. Targeting PDZ domains is a strategy developed by many viruses to hijack cellular machinery to their advantage²²⁴. It has been reported that the PDZ-mediated interaction of E6 with PTPN3 results in the proteasomal degradation of the phosphatase^{67,256}. In the case of HBV, PTPN3 has been shown to have a suppressive role in HBV gene expression, but the functional relevance of this interaction has not been established⁶⁸.

This project had two main objectives. The first one was to study the determinants of PTPN3 specificity for its PBM ligands, while the second one focused on studying the role of PTPN3 in signaling and understanding the disruptive effects of viral infections in the function of this protein. At the start of this project, the interactions of PTPN3 with cellular and viral proteins had not been characterized and the structure of its PDZ domain was not known. The few biophysical and structural studies on PTPN3 have focused on the PTP domain in complex with phospho-peptide substrates derived either from the kinase MAPK p38 γ ⁷⁰ or the Epidermal growth factor receptor substrate 15³³⁷. Thus, the interactions mediated by the PDZ domain of PTPN3 had not been deeply studied, and hence the role of this protein in cellular signalling was not well understood. I aimed to bridge this gap by performing an integrative structural and functional study of PTPN3 and its PDZ-mediated interactions. The goal of this

thesis project was to gain insights about fundamental aspects of PTPN3 function and signalling, and to understand in particular its interactions with PBM-containing partners.

I first performed a biophysical characterization of PTPN3-PDZ, which allowed me to explore the stability of this domain in solution and study the effects of PBM ligand binding. Then, I determined the affinities of the interactions between PTPN3-PDZ and its cellular and viral partners and obtained the first crystal structures of PTPN3-PDZ complexed to PBM peptides. This way, I highlighted the atomic determinants of PTPN3-PDZ/PBM interactions and identified the crucial positions that define the selectivity of this domain. By combining these observations with data previously obtained for the PTPN3 close homolog PTPN4, I explored the structural basis of the selectivity of the NT5 subfamily of phosphatases.

The second part of this project focused on HBV, the targeting of PTPN3 by the HBc PBM, and the effects of PTPN3 on the viral life cycle. Despite the known relevance of viral PBMs for the life cycle of various families of pathogenic and non-pathogenic viruses²²⁴, little was known about the HBc PBM. Thus, I set out to investigate the interactome of the HBc PBM by searching for new potential cellular PDZ-containing protein targets of HBc that could be relevant for the viral life cycle. I identified various potential PDZ protein partners of HBc, many of which are involved in cell polarity, and are also targeted by other viruses. Among these partners, I also found PTPN3. I showed that PTPN3 bound HBc in the context of the viral capsid, suggesting that this interaction could be physiologically relevant. To gain more insights on this, I analysed the impact of PTPN3 overexpression in HBV-infected cells. Finally, to gain a more comprehensive understanding of the physiological role of PTPN3, I searched for cellular partners that interact with PTPN3 through its PDZ domain and that could be displaced by HBc, potentially affecting the cellular homeostasis.

Results

5 Structural and functional characterization of the PDZ domain of PTPN3 and its interaction with the high-risk HPV type 16 E6 oncoprotein

5.1 Summary

Most structural and biophysical studies of PTPN3 have focused on its phosphatase domain, whose structure has been determined in complex with phosphopeptides derived from its substrates Eps15³³⁷ and p38 γ ⁷⁰. However, at the start of this project, the structure of the PTPN3 PDZ domain was not known, and thus there was a lack of data on the PDZ-mediated interactions with cellular and viral partners.

To address this, we first performed a biophysical characterization of our PTPN3-PDZ construct, and optimized its delimitation to increase its stability for further structural studies. Then, we solved the crystal structure of the PDZ domain of PTPN3 in complex with the PBM of HPV16 E6. We combined the information derived from the X-ray structure with data on the complex formation obtained by NMR spectroscopy to identify the molecular basis of the recognition of the C-terminal sequence of E6 by PTPN3 and the perturbations that spread from the binding groove upon ligand binding. We also measured the affinities of the interactions between PTPN3-PDZ and the PBM of cellular and viral partners: p38 γ , the HBV core protein and the HPV types 16 and 18 E6 proteins. We observed that the affinities of the viral and cellular partners of PTPN3-PDZ fall in the same range and are comparable. We report thus the first structural data on the PDZ domain of PTPN3, and we provide insights on the mechanism of binding of its PBM ligands.

5.2 Article: Structural and functional characterization of the PDZ domain of the human phosphatase PTPN3 and its interaction with the human papillomavirus E6 oncoprotein

SCIENTIFIC REPORTS

OPEN

Structural and functional characterization of the PDZ domain of the human phosphatase PTPN3 and its interaction with the human papillomavirus E6 oncoprotein

Mariano Genera^{1,2}, Damien Samson³, Bertrand Raynal⁴, Ahmed Haouz⁵, Bruno Baron⁴, Catherine Simenel³, Raphael Guerois⁶, Nicolas Wolff¹ & Célia Caillet-Saguy¹

The human protein tyrosine phosphatase non-receptor type 3 (PTPN3) is a PDZ (PSD-95/Dlg/ZO-1) domain-containing phosphatase with a tumor-suppressive or a tumor-promoting role in many cancers. Interestingly, the high-risk genital human papillomavirus (HPV) types 16 and 18 target the PDZ domain of PTPN3. The presence of a PDZ binding motif (PBM) on E6 confers interaction with a number of different cellular PDZ domain-containing proteins and is a marker of high oncogenic potential. Here, we report the molecular basis of interaction between the PDZ domain of PTPN3 and the PBM of the HPV E6 protein. We combined biophysical, NMR and X-ray experiments to investigate the structural and functional properties of the PDZ domain of PTPN3. We showed that the C-terminal sequences from viral proteins encompassing a PBM interact with PTPN3-PDZ with similar affinities to the endogenous PTPN3 ligand MAP kinase p38 γ . PBM binding stabilizes the PDZ domain of PTPN3. We solved the X-ray structure of the PDZ domain of PTPN3 in complex with the PBM of the HPV E6 protein. The crystal structure and the NMR chemical shift mapping of the PTPN3-PDZ/peptide complex allowed us to pinpoint the main structural determinants of recognition of the C-terminal sequence of the E6 protein and the long-range perturbations induced upon PBM binding.

Protein tyrosine phosphatases (PTPs) play critical roles in cell signaling pathways. They are known to regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. Together with kinases, they control the balance of phosphorylated species, enabling specific and varied signaling responses.

The human protein tyrosine phosphatase non-receptor type 3 (PTPN3) and the protein tyrosine phosphatase non-receptor type 4 (PTPN4) compose the NT5 type of non-transmembrane PTPs. These large modular proteins feature a N-terminal FERM (Band 4.1, Ezrin, Radixin, and Moesin) domain that is responsible for the localization of these enzymes to submembranous sites in the cell, a PDZ (PSD-95/Dlg/ZO-1) domain that allows specific binding to C-terminal recognition sequences in target proteins, and a C-terminal catalytic tyrosine phosphatase domain (Fig. 1). Three different isoforms of PTPN3 are known to be produced by alternative splicing, one being the full-length protein (UniProt P26045-1), whereas the FERM domain is partially or totally missing in the other two (UniProt P26045-2 and P26045-3). PTPN3 is localized in the cytoplasm and enriched at the plasma membrane, although in absence of the FERM domain it is exclusively cytoplasmic¹.

PTPN3 has been linked to many cancers either with a tumor-suppressive or a tumor-promoting role². Interestingly, PTPN3 is targeted by oncoviruses such as the high-risk genital human papillomaviruses (HPV)

¹Récepteurs-Canaux, Institut Pasteur, UMR 3571, CNRS, F-75724, Paris, France. ²Sorbonne Université, Complexité du Vivant, F-75005, Paris, France. ³RMN des biomolécules, Institut Pasteur, UMR 3528, CNRS, F-75724, Paris, France. ⁴Plate-forme de Biophysique Moléculaire, Institut Pasteur, UMR 3528, CNRS, F-75724, Paris, France. ⁵Plate-forme de Cristallographie, Institut Pasteur UMR 3528, CNRS, F-75724, Paris, France. ⁶Institut de Biologie Intégrative de la Cellule (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, 91190, Gif-sur-Yvette, Cedex, France. Correspondence and requests for materials should be addressed to C.C.-S. (email: celia.caillet-saguy@pasteur.fr)



Figure 1. Schematic representation of the PTPN3 constructs. Numbers on both extremities indicate the boundary residues for each construct; Numbers in italic above the schematic construct of full-length PTPN3 correspond to the boundary residues of each protein domain from Uniprot P26045.

Peptides	Sequences	K_D (μ M)	T_m ($^{\circ}$ C)
p38 γ PBM	SWARVSKETPL	26 \pm 25	45
HBVc PBM	RRRRSQSRESQC	29 \pm 24	47
HPV16E6 PBM	RSSRTRRETQL	53 \pm 31	52
HPV18E6 PBM	RQERLQRRETQV	37 \pm 20	49

Table 1. Peptide sequences, affinities for PTPN3-PDZ and melting temperature of complexes. The errors are the standard deviations of all the K_D values derived from 8 to 14 curves fitted for each titration.

types 16 and 18³ and the hepatitis B virus (HBV)⁴. HPV types 16 (HPV16) and 18 (HPV18) are responsible of 70% of cervical cancers and precancerous cervical lesions⁵.

The E6 proteins of HPV16 and HPV18 possess a conserved C-terminal PDZ binding motif (PBM), which mediates interaction with some cellular PDZ domain-containing proteins. The ability of E6 to bind to PDZ domain-containing proteins correlates with its oncogenic potential⁶. The PDZ domain of PTPN3 (PTPN3-PDZ) has been identified as a target of the E6 PBM of HPV16 and 18. This PDZ-PBM interaction results in the ubiquitin-mediated proteasomal degradation of PTPN3. Therefore, the endogenous levels of PTPN3 are particularly low in HPV-positive cervical carcinoma cell lines³.

To date, the structure of PTPN3-PDZ is not known and the PDZ-mediated interactions between PTPN3 and viral partners have not been described. Indeed, the few biophysical and structural studies on PTPN3 have focused on the PTP domain in complex with phospho-peptide substrates derived either from the mitogen-activated protein kinase (MAPK) p38 γ (also known as MAPK12)⁷ or the Epidermal growth factor receptor substrate 15 (Eps15)⁸. We have previously shown that the PDZ domain of the closely related PTPN4 inhibits its phosphatase activity^{9,10}. Similarly, the inhibitory role of the PDZ domain of PTPN3 on the catalytic activity has been reported using a phospho-p38 γ peptide as a substrate for kinetics experiments⁷.

In this study, we report the structural and functional analyses of the PDZ domain of PTPN3. We investigated the molecular mechanism of interaction between the PDZ domain of PTPN3 and the PBM of cellular and viral partners (Table 1), the MAP kinase p38 γ , the HBV core protein and the HPV E6 protein. We employed an integrative approach based on biophysical, nuclear magnetic resonance (NMR) and X-ray crystallography experiments. We showed that the C-terminal PBM sequences of the HPV E6 protein interact with PTPN3-PDZ with similar affinities to those of the endogenous PTPN3 ligand p38 γ and the HBV Core protein⁴. By solving the X-ray structure and mapping the NMR chemical shift changes of PTPN3-PDZ in complex with the PBM of HPV16 E6, we identified the molecular basis of recognition of the C-terminal sequence of E6 protein and the induced perturbations that spread from the binding groove.

Results

Quality control of the recombinant PTPN3-PDZ domain. Presently, there is no structural information on PTPN3-PDZ other than that provided by homology modelling⁷. To bridge this gap, we performed different structural and biophysical studies on the recombinant PTPN3-PDZ construct (Fig. 1) expressed in *Escherichia coli*.

To estimate the secondary structure content of the PDZ construct, we performed circular dichroism (CD) measurements in the far-UV (195–240 nm). The PTPN3-PDZ spectrum (Fig. 2A) appeared to be of mixed α -helix and β -sheet content, which upon deconvolution indicated 16% α -helix and 32% β -sheet. The secondary structure content extracted from the X-ray structure of the PDZ domain of its close homologue PTPN4 (PTPN4-PDZ) (PDB id 3NFK)¹¹ is 17% α -helix and 31% β -sheet. Thus, the PTPN3-PDZ construct presents a secondary structure content similar to that of PTPN4-PDZ.

We assessed the tertiary folding of PTPN3-PDZ by NMR by performing ¹⁵N heteronuclear single quantum correlation (HSQC) experiment. The high dispersion of resonances in the HSQC spectrum of PTPN3-PDZ (Fig. 2B) is indicative of a mainly well-folded construct in solution.

Stability and oligomeric state of PTPN3-PDZ. To evaluate the stability of PTPN3-PDZ, we performed differential scanning calorimetry (DSC) (Fig. 2C). A single endothermic peak was observed in the DSC thermogram, with a thermal transition midpoint (T_m) of 41 $^{\circ}$ C. As a comparison, we performed DSC on PTPN4-PDZ. We obtained an endothermic peak with a T_m of 47 $^{\circ}$ C (Fig. 2C). The data show that PTPN3-PDZ is less stable in

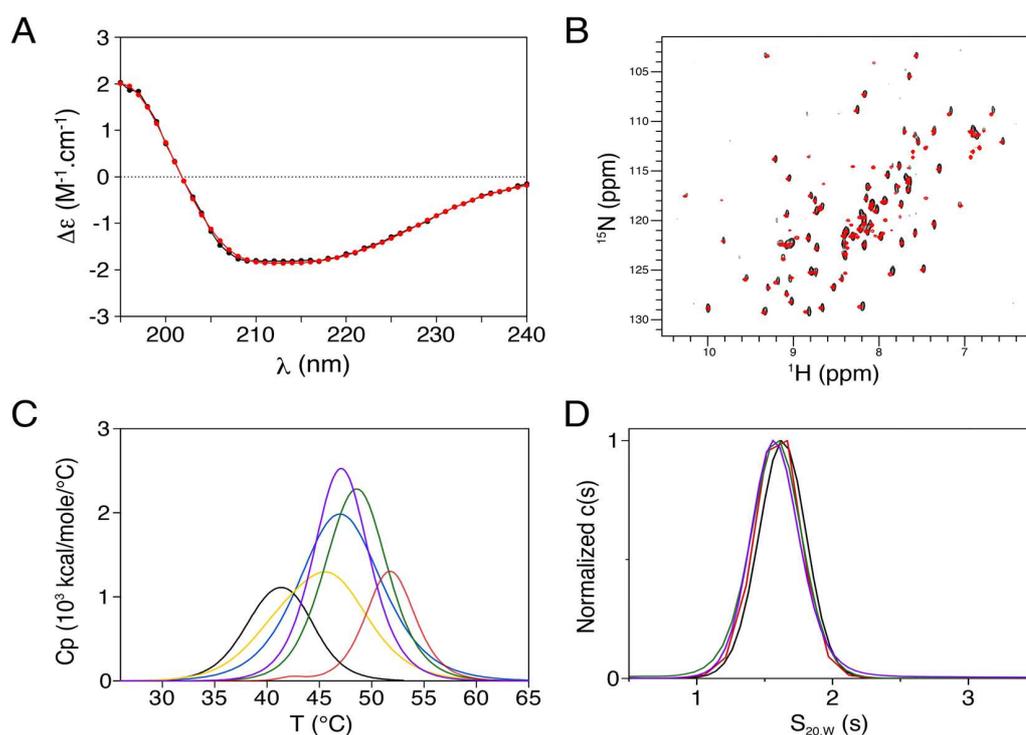


Figure 2. Stability, oligomerization and structural characterization of PTPN3-PDZ. **(A)** Far-UV CD spectrum (195–240 nm) of PTPN3-PDZ. The experimental spectrum of PTPN3-PDZ and the best fit obtained with the CONTINLL algorithm are represented as black and red lines, respectively. **(B)** Superimposed ^1H , ^{15}N HSQC spectra of PTPN3-PDZ (black) and PTPN3-PDZ_{Next} (red). **(C)** DSC thermograms of PTPN3-PDZ unbound and complexed to PBM peptides. Free PTPN3-PDZ curve is shown as a black line, and PTPN3-PDZ complexed to p38 γ PBM, HBVc PBM and HPV16E6 PBM and HPV18E6 PBM curves are shown as yellow, purple, green and red lines, respectively. The DSC thermogram of free PTPN4-PDZ is colored in blue. **(D)** Sedimentation coefficient distributions of PTPN3-PDZ_{Next} unbound and complexed to viral PBM peptides. The black, red, green and purple lines correspond to PTPN3-PDZ_{Next} unbound and complexed to HPV18E6PBM, HPV16E6 PBM and HBVc PBM, respectively.

PTPN3-PDZ						
	10 μM	25 μM	70 μM	70 μM + HBVc PBM		
MW theoretical (kDa)	10.5	10.5	10.5	12.0		
MW estimated (kDa)	12.4	11.3–25.9 ^a	nc	13.5		
Sed coef ($S_{0,w,20}$) (S)	1.5	1.5–2.6 ^a	nc	1.6		
PTPN3-PDZ _{Next}						
	17.5 μM	35 μM	70 μM	70 μM + HPV16E6 PBM	70 μM + HPV18E6 PBM	70 μM + HBVc PBM
MW theoretical (kDa)	12.7	12.7	12.7	14.1	14.1	14.2
MW estimated (kDa)	12.7	12.5	15	13.7	14.7	15.6
Sed coef ($S_{0,w,20}$) (S)	1.6	1.7	1.6	1.6	1.6	1.6

Table 2. Hydrodynamic parameters of PTPN3-PDZ derived from the analysis of analytical ultracentrifugation. ^aTwo peaks are present; nc: not calculated because of a continuum of species.

solution than PTPN4-PDZ. This thermal instability of the PTPN3-PDZ construct could be responsible for its aggregation previously detected by NMR. Indeed, we observed a significant loss of peak intensities in the HSQC spectrum of PTPN3-PDZ related to aggregation after 10 h at 20 °C (data not shown).

To estimate the oligomeric state of the PTPN3-PDZ construct, analytical ultracentrifugation (AUC) experiments were performed (Table 2). Three different concentrations of the protein were tested, 10, 25 and 70 μM . Only one single peak was detected at 10 μM with a sedimentation coefficient of 1.5S in agreement with a monomeric form of 10.5 kDa. At 25 μM , an additional peak was detected with a sedimentation coefficient of 2.6S corresponding to a dimeric form. At a concentration of 70 μM , higher oligomeric species could be detected. Thus, the AUC data of the PTPN3-PDZ construct reveal a concentration-dependent equilibrium between monomer, dimer and higher oligomeric states with only monomers detected at a concentration of 10 μM .

PBM-mediated stabilization of the PTPN3-PDZ. To check whether the PBM binding to the PDZ domain could affect the stability of PTPN3-PDZ, we performed DSC and AUC on PTPN3-PDZ in complex with the p38 γ PBM and the viral PBMs of HBV Core and HPV 16 and 18 E6 proteins, namely HBVc PBM, HPV16E6 PBM and HPV18E6 PBM respectively (Table 1). These PBMs have been shown to interact with the PTPN3-PDZ^{3,4,7}. A single endothermic peak was observed in the DSC thermograms of all complexes (Fig. 2C) with a T_m of 45 °C, 47 °C, 52 °C and 49 °C for PTPN3-PDZ complexed to p38 γ PBM, HBVc PBM, HPV16E6 PBM and HPV18E6 PBM, respectively (Table 1). For all the PBMs tested, an increase of 4 to 11 °C in the T_m was observed, showing that the PBM binding onto PTPN3-PDZ stabilizes the domain, whether the PBM is of cellular or viral origin.

In addition, AUC experiments of the PTPN3-PDZ construct at 70 μ M complexed to HBVc PBM show only one species with a sedimentation coefficient of 1.6S corresponding to a one-to-one complex with the peptide, without higher oligomeric species (Table 2). Thus, the PBM binding onto PTPN3-PDZ prevents oligomerization at high concentration.

Affinities of the viral and the cellular PBMs for PTPN3-PDZ. The affinities of HPV16E6 PBM, HPV18E6 PBM, p38 γ PBM and HBVc PBM for PTPN3-PDZ were measured by NMR titration. We followed ¹H, ¹⁵N chemical shift perturbations of PTPN3-PDZ signals in the ¹H-¹⁵N HSQC spectra as a function of increasing concentrations of the PBM peptide to determine the dissociation constant (K_D) (Fig. S1). The p38 γ PBM peptide binds to PTPN3-PDZ with a K_D value of 26 μ M (Table 1). This affinity is 16-fold lower than the one previously reported for PTPN4-PDZ (K_D of 1.6 μ M)¹⁰. We obtained K_D values of 29 μ M, 53 μ M and 37 μ M for PTPN3-PDZ with HBVc PBM, HPV16E6 PBM and HPV18E6 PBM, respectively (Table 1). The K_D values are all in the same tenth-of-micromolar range for the viral PBMs and are close to the one of the cellular partner p38 γ . The measured affinities fall in the standard 0.1–100 μ M range¹² for PDZ-PBM interactions. To rule out the possibility of an effect of self-association of PTPN3-PDZ on the K_D values, we compared the K_D values with the HPV16E6 and HPV18E6 PBM peptides fitted from the NMR data with the ones estimated from the binding intensities (BIs) obtained for the HPV16E6 and HPV18E6 PBM peptides measured from a quantitative screening assay against the human PDZome library, in which each PDZ domain is at a concentration of 4 μ M and in excess of PBM peptides¹². The BIs are directly related to the K_D. For PTPN3-PDZ, the estimated K_D values are 46 and 96 μ M for the HPV16E6 and HPV18E6 PBMs respectively, in good agreement with our experimental results.

Stabilization of PTPN3-PDZ for structural studies. The instability and tendency to oligomerize of PTPN3-PDZ made this construct unsuitable for crystallogenes. To gain in stability and protein expression yield, we extended the N-terminal extremity of PTPN3-PDZ by 15-residues with its wild-type upstream sequence (PTPN3-PDZ_{Next} in Fig. 1). These extensions have been reported to provide structural stability to some PDZ containing proteins^{13,14}. The PTPN3-PDZ production yield was increased two-fold after extension from 0.9 mg to 1.8 mg of purified protein per liter of culture. The stability of PTPN3-PDZ_{Next} was then evaluated by NMR HSQC spectra recorded during a week at 20 °C. Spectra show a similar pattern of resonances than in the HSQC spectrum of PTPN3-PDZ without any signs of aggregation, such as a loss of peak intensity (Fig. 2B). The additional peaks of the N-terminal extension residues fall in the spectral region characteristic of unfolded protein and no chemical shift differences between PTPN3-PDZ_{Next} and the shorter form are detected. Thus, the N-terminal extension is not structured in the PTPN3-PDZ_{Next} and does not interact with the PDZ domain.

In addition, one single peak was detected by AUC for PTPN3-PDZ_{Next} at concentrations between 17.5 μ M and 70 μ M with a sedimentation coefficient of 1.6–1.7S, in agreement with a monomeric form of 12.7 kDa (Table 2) and devoid of higher oligomeric species. All these data indicate that the extension of PTPN3-PDZ resulted in a stable and folded PDZ domain in solution.

AUC experiments were also performed on the PTPN3-PDZ_{Next} construct at 70 μ M complexed to HBVc PBM, HPV16E6 PBM and HPV18E6 PBM (Table 2 and Fig. 2D). As expected, only one species with a sedimentation coefficient of 1.6S is detected in all cases, corresponding to a one-to-one complex.

Crystal structure of the PDZ domain of PTPN3 in complex with the viral HPV16E6 PBM. To investigate the molecular determinants of the interaction of PTPN3 with the HPV16 E6 protein, we solved the crystal structure of the complex formed by PTPN3-PDZ_{Next} and HPV16E6 PBM by molecular replacement at 2.19 Å resolution (Table 3). The structure factors and coordinates have been deposited in the Protein data Bank under accession code 6HKS.

PTPN3-PDZ_{Next} adopts a typical PDZ fold, with a β -sandwich comprising five β strands and two α helices (Fig. 3A). The PBM ligand binds in a hydrophobic cleft formed by the β 2-strand, the α 2-helix and the “GLGF” loop (Figs 3A,B). HPV16E6 PBM binds to the PDZ domain as an anti-parallel extension of the β 2-strand domain in a conventional mode. There is no electron density corresponding to the N-terminal extension of PTPN3-PDZ_{Next} in the crystal, which indicates a disordered region, in agreement with the NMR results that show an unfolded extension.

We also determined a 3D structural model with the CS-ROSETTA approach¹⁵ using the ¹⁵N, ¹³C α , ¹³C β NMR backbone and ¹³C β resonances of PTPN3-PDZ_{Next} complexed to HPV16E6 PBM (BMRB accession number 27645). The CS-Rosetta modeling using solution NMR data revealed a very similar conformation to the crystal structure, with a low root mean square deviation (rmsd) of 1.11 Å for the backbone atoms between the crystal structure and the averaged solution model originated from the 10 lowest energy models (Fig. 4C). Thus, 3D models of PDZ-PTPN3 determined by CS-Rosetta reveal a conformation in solution that matches the crystal structure of the PDZ domain of the complex PTPN3-PDZ_{Next} and HPV16E6 PBM (Fig. 3B). The AUC sedimentation velocity was also back-calculated using the crystal structure¹⁶. We obtained a calculated sedimentation coefficient value

PTPN3-PDZ _{Next} /HPV16E6 PBM	
Data collection	
Space group	P 1 2 1
Unit cell (a, b, c) (Å)	46.62, 77.43, 130.03
α, β, γ (°)	90.00, 90.14, 90.00
Resolution (Å)	46.62–2.19 (2.23–2.19)
No. of reflections (total/unique)	324034/47024
Redundancy	6.9 (6.5)
Completeness (%)	99.1 (95.7)
$I/\sigma(I)$	8.17 (1.11)
R-meas [†]	0.14 (1.71)
CC(1/2)	99.6 (72.6)
Refinement	
Resolution (Å)	43.34–2.19
No. reflections	46869
$R_{\text{work}}^{\ddagger}/R_{\text{free}}^{\ddagger}$	0.195/0.245
No. of protein atoms/ligand atoms	4402/350
No. of solvent/hetero-atoms	235/9
Rmsd bond lengths (Å)	0.008
Rmsd bond angles (°)	0.949
Wilson B-factors	47.6
Ramachandran plot (favored/disallowed)*	97.4/0.4
PDB code	6HKS

Table 3. Data collection and refinement statistics. Values in parenthesis correspond to the highest resolution shell. [†]Rmeas = $\sum h(n/n - 1)^{1/2} \sum_i |I_i(h) - \langle I(h) \rangle| / \sum h \sum_i I_i(h)$, where $I_i(h)$ and $\langle I(h) \rangle$ are the i th and the mean measurement, respectively, of the intensity of reflection h . [‡]Rwork = $\sum h ||F_{\text{obs}}(h)| - |F_{\text{calc}}(h)|| / \sum h |F_{\text{obs}}(h)|$, where $F_{\text{obs}}(h)$ and $F_{\text{calc}}(h)$ are the observed and calculated structure factors, respectively. No I/σ cutoff was applied. [‡]Rfree is the R value obtained for a test set of reflections consisting of a randomly selected 5% subset of the data set excluded from refinement. *Categories were defined by MolProbity.

of 1.59S, which is in good agreement with the experimental measurement of 1.6S. Thus, the overall structure of the domain in the crystal is consistent with its arrangement observed in solution.

Then, we compared the PDZ domains of PTPN3 and PTPN4, which share 71% of sequence identity (Fig. 4A). Both structures of PTPN3-PDZ and PTPN4-PDZ are highly similar, with a very low rmsd value of 0.34 Å for the backbone atoms (PTPN4 PDB ID 5EYZ)¹⁰ (Fig. 4B).

Structural insights on the recognition of the PBM of HPV16 E6 protein. The PDZ domain of PTPN3 possesses the interaction network specific to class I PDZ domains and recognizes the consensus sequence S/T-X- Φ -COOH, where X is any residue, and Φ is a hydrophobic residue¹⁷. A clear electron density map was seen for only the last seven C-terminal residues of HPV16E6 PBM (-TRRETQL_{COOH}) that are inserted into the binding groove (Fig. 3A,B). Similarly, only the last 5 or 6 C-terminal residues of PBM peptides are visible in the crystal PTPN4-PDZ structures in complex with PBM peptides (PDB IDs 3NFL, 3NFK, 5EZO, 5EYZ in references^{10,11}).

The interactions of the PBM residues at positions 0 and -2 with PTPN3-PDZ are quite similar to the bonding patterns already observed in the complex between PDZ-PTPN4 and the viral peptide Cyto13-att (-GETRL_{COOH}) derived from the attenuated rabies virus, or the optimized pro-apoptotic 13-amino acids peptide Cyto8-RETEV (-RETEV_{COOH}) (PDB IDs 3NFK and 5EYZ respectively)^{10,11}. Indeed, the C-terminal carboxylate of the leucine (L₀) of the HPV16E6 PBM forms three H-bonds with the amide nitrogens of F521, G522 and F523 of the “GLGF motif” on PTPN3-PDZ (Fig. 3B) as observed for F528, G529 and F530 of PTPN4-PDZ with the L₀ of the Cyto13-att (-GETRL_{COOH}) PBM¹¹. The hydroxyl group of threonine at position -2 forms a hydrogen bond with the N ϵ 2 of the conserved histidine H572 from the α 2-helix of PTPN3-PDZ (Fig. 3B). Electron acceptors such as serine and threonine are therefore preferred at this position for the class I PDZ domains.

At position -1 of HPV16E6 PBM, the glutamine (Q₋₁) side chain well defined and forms a H-bond with a water molecule that is also H-bonded to the N δ 2 of N524 of PTPN3-PDZ (Fig. 3B), whereas it is exposed to the solvent. Interestingly, in the complex of PTPN4-PDZ with the PBM of the glutamate receptor subunit GluN2A (PDB ID 3NFL), which presents a D in position -1¹¹, the D₋₁ forms a H-bond with the N531 side-chain amine group of PTPN4-PDZ, the equivalent of N524 of PTPN3-PDZ.

A glutamate at position -3 (E₋₃) is conserved in all viral PBMs targeting PTPN3 and is also present in the one of p38 γ . The side-chain carboxyl of E₋₃ forms a bifurcated H-bond with the hydroxyl of the conserved S538 (S545 for PTPN4) and N524 (N531 for PTPN4) amine group (Figs 3B and 4A). E₋₃ is also stabilized by hydrophobic contacts involving its C β -C γ carbon chain and the long aliphatic side chain of K526 as observed in PTPN4 (K533) bound to Cyto8retev (PDB 5EYZ).

Finally, the guanidinium nitrogens of arginine at position -4 (R₋₄) form ionic bonds with the carboxylate oxygens of D573 (D580 in PTPN4)(Fig. 3B).

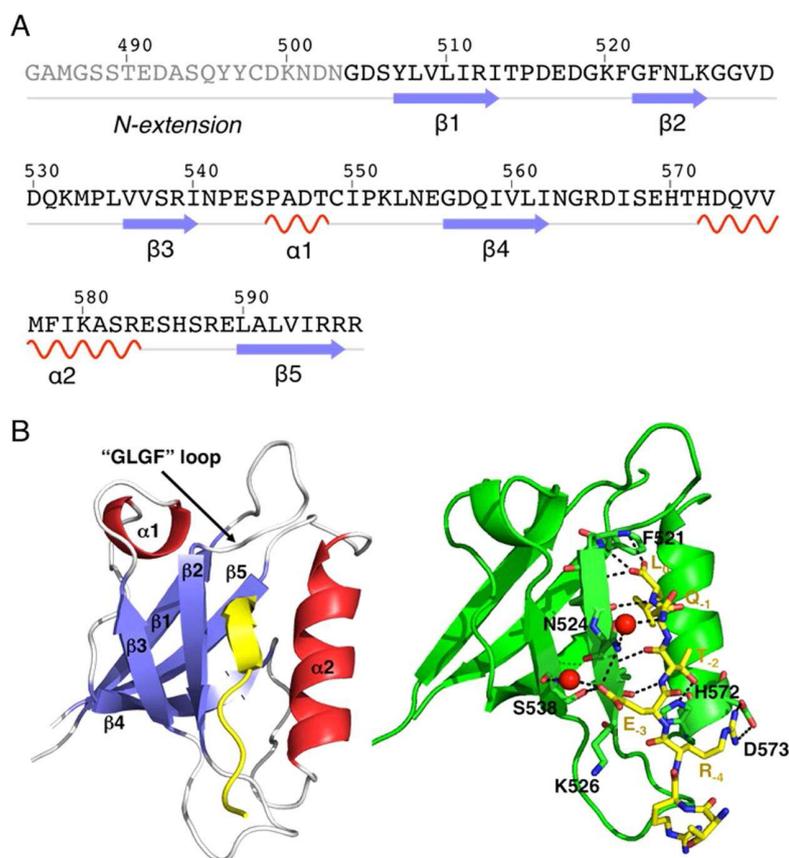


Figure 3. X-ray structure of PTPN3-PDZ_{Next} bound to HPV16E6 PBM. **(A)** Sequence and secondary-structure elements in PTPN3-PDZ_{Next}. No electron density was observed for the residues shown in gray. **(B)** Left panel: Overall structure of PTPN3-PDZ_{Next} bound to HPV16E6 PBM displayed as a ribbon. The structure is colored according to secondary structure elements (β -strands in blue, helices in red, and loops in white). The peptide (in yellow) enters the PDZ-binding pocket delimited by the β 2 strand, the α 2 helix, and the "GLGF" loop⁵¹⁹GKFGF⁵²³. Right panel: Detailed view of the PTPN3-PDZ bound to HPV16E6 PBM. Important residues are shown as sticks in CPK colors, molecules of water are represented as red spheres, and intermolecular H-bonds and polar contacts are shown as black dashed lines.

To determine short- and long-range perturbations in PTPN3-PDZ upon PBM binding, we analyzed by solution NMR the chemical shifts of the free PTPN3-PDZ in solution and its complex with peptide HPV16E6 PBM, comparing their ¹H, ¹⁵N HSQC spectra. We assigned 97% of the HN resonances, 99% of the C α resonances, 90% of the C β resonances and 89% of the CO resonances (89 non-Proline residues over 94) of PTPN3-PDZ complexed to HPV16E6 PBM (BMRB accession number 27645). Chemical shift changes ($\Delta\delta$) in PTPN3-PDZ spectra upon complex formation provide insights at atomic level on the residues involved in the interaction with the HPV16E6 PBM and on potential distal effects. Two types of signals were useful for the analysis of the ¹H, ¹⁵N chemical shift mapping (Fig. 5A): (1) signals experiencing significant chemical shift changes ($\Delta\delta > 0.15$ ppm); (2) signals that undoubtedly disappear from their original well-resolved position in the spectrum upon complex formation and are therefore severely affected when binding to the PBM (either due to large chemical shift changes or to severe line broadening effects caused by exchange). Residues corresponding to type 1 are shown in blue in Fig. 5A,B, while seven residues of type 2 are colored in red. Nine residues, whose behavior could not be safely defined mainly because they fall in crowded spectral regions, are colored in gray.

As expected, we identified differences in backbone NH chemical shifts between free PTPN3-PDZ and when bound to HPV16E6 PBM. Residues G522, G527 and H572, are among the most affected and are located in the PBM binding site. I559 and L590 also experience a strong effect. Indeed, they make hydrophobic contacts of 5.4 Å and 4 Å with F523 and F521 of the "GLGF motif", respectively. E589 (in red), the direct neighbor of L590, displays perturbations both in chemical shift or intensity. Some residues of α 2-helix in the binding site are affected: D573 and F578 (in blue) and V575 (in red). We find also residues such as F521 and F523 of the "GLGF" motif and D516 in the "GLGF" loop affected in their NMR resonances. D530 in the β 2- β 3 loop forms a H-bond with H572, which interacts directly with the PBM peptide (position T₂).

Interestingly, G556 (in red) is part of a path of perturbation starting from G527 (in red) and G528 (in blue) at the end of β 2-strand in the binding groove, transmitted to the close V536 (in red) at the beginning of β 3-strand, affected

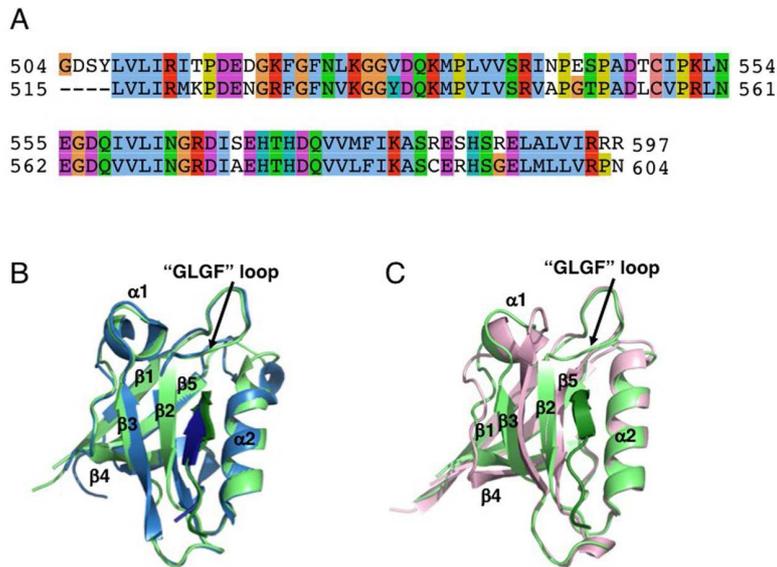


Figure 4. Structural comparison of the X-ray structure of PTPN3-PDZ_{Next} bound to HPV16E6 PBM with PTPN4-PDZ complexed to Cyto8retev, and with the CS-ROSETTA model of PTPN3-PDZ. (A) Clustalw2 alignment of PTPN3-PDZ (top) and PTPN4-PDZ (bottom). (B) Superposition of the structures of PTPN3-PDZ_{Next} bound to HPV16E6 PBM (light green/green) with PTPN4-PDZ bound to Cyto8retev (PDB 5EYZ) (light blue/blue). (C) Superposition of PTPN3-PDZ_{Next} bound to HPV16E6 PBM (light green/green) with the CS-ROSETTA model of PTPN3-PDZ (pink).

in signal intensity. The perturbation of V536 seems to be transferred to the close G556 on the β 4-strand. I559 (in red) on β 4-strand is also affected, possibly experiencing the perturbation of the F523 located 5 Å away of the “GLGF motif”. Finally, the path of perturbation finishes with R595 at the C-terminus of the PDZ domain (in blue). Thus, long-range perturbations seem to be induced upon PBM binding and transmitted through an interconnected network of residues throughout PTPN3-PDZ. This network involves the β 2, β 3 and β 4 strands and the C-terminus.

Discussion

PTPN3 is a PDZ domain-containing phosphatase that has been demonstrated to function as a tumor suppressor or, conversely, as an oncoprotein in a context-dependent manner. It integrates a number of signaling pathways via interactions mediated by its PDZ domain^{18–20}. The PDZ domain of PTPN3 is also a cellular target of oncoviruses. Until now, no structural data on this domain were available. We investigated the structural and functional properties of PTPN3-PDZ and characterized its interaction with the PBMs of one cellular partner, the MAP kinase p38 γ , and of its viral partners, the HPV E6 protein and the HBV core protein. We defined in particular the molecular basis of the PDZ-PBM interaction between HPV E6 and PTPN3.

Structural properties of the PDZ domain of PTPN3. The crystal structure of PTPN3-PDZ_{Next} complexed to HPV16E6 PBM reveals a typical PDZ fold with five β -strands and two α -helices (Fig. 3A,B). It was previously reported that some PDZ domains can be stabilized by extensions of either the N-, C-, or both termini of a PDZ²¹. These extended regions were required for correct folding and ligand binding. For example, the PDZ1 of the membrane-associated guanylate kinase inverted MAGI-1 is unstable without N- and C-terminal extensions of 14 and 26 residues, respectively, even though these sequences are unstructured in both free and ligand-bound states²². Our minimal PTPN3-PDZ construct, deduced from sequence alignments with PDZ domains of known structure (Fig. 1), started at residue 504, while the first N-terminal secondary structure element observed in the crystal structure of PTPN3-PDZ_{Next}, the β 1-strand, started at residue 507 and not at 510 as expected from the sequence alignments. We showed that this PTPN3-PDZ construct is folded but unstable. We were able to increase its stability and production yield, while decreasing its tendency to auto-associate, by the addition of an unfolded 15-residue N-terminal extension to the minimal delimitation of PTPN3-PDZ. 3D structural models of PDZ-PTPN3 in complex with HPV16E6 PBM determined by CS-Rosetta using backbone and ¹³C₃ NMR chemical shift assignments revealed a conformation in solution congruent with the X-ray structure of the PDZ domain.

Our AUC data showed that PTPN3-PDZ auto-association is prevented by PBM binding, as already reported for other PDZ domains²³. The molecular mechanism by which PBM binding interferes with PTPN3 PDZ auto-association remains unknown. Our NMR results on long-range perturbations upon PBM binding are consistent with a network involving the β 2, β 3 and β 4 strands and the C-terminus residues throughout PTPN3-PDZ. This connected pathway communicates the binding event to regions that are distal to the binding cleft. These findings are compatible with previous data on PDZ domains suggesting that energetic pathways within PDZ domains may support allostery^{24–27}. A regulation controlled by the equilibrium between PBM binding and dimerization is attractive since over 30% of PDZ domains are known to form dimers in solution²⁸.

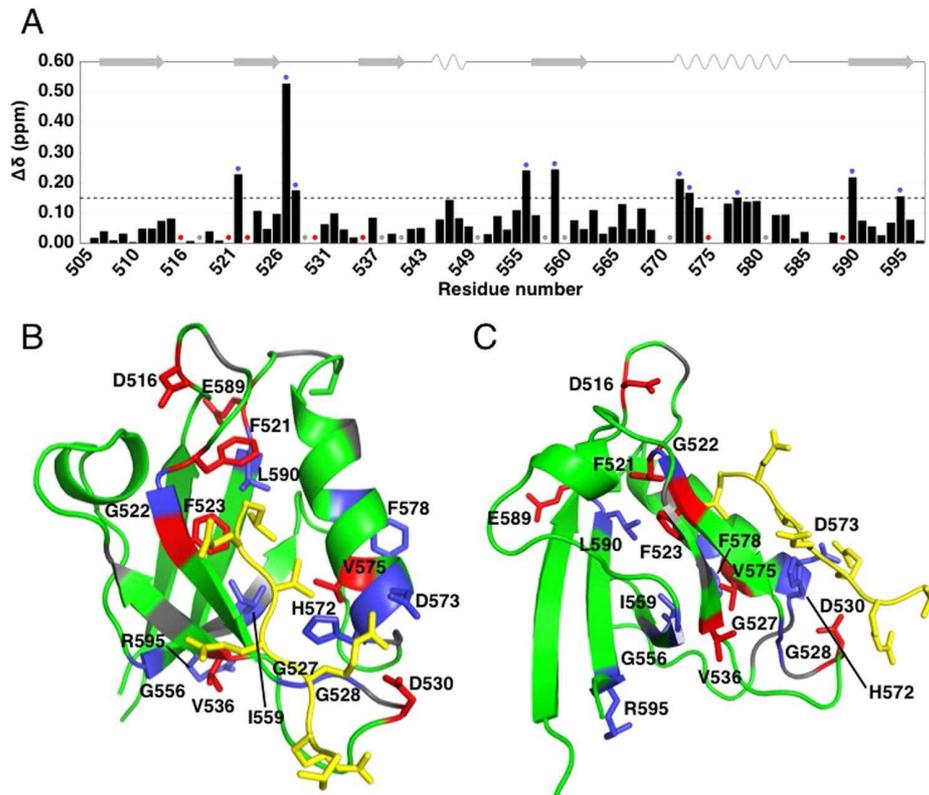


Figure 5. $\Delta\delta$ between free PTPN3-PDZ and PTPN3-PDZ bound to HPV16E6 PBM. **(A)** $\Delta\delta$ values (^1H , ^{15}N) computed as $\Delta\delta = [(\Delta\delta\text{H})^2 + (\Delta\delta\text{N} \cdot 0.15)^2]^{1/2}$. The dashed black line represents the threshold used to determine the perturbed region with $\Delta\delta > 0.15$ ppm; blue circles mark residues perturbed with $\Delta\delta > 0.15$ ppm; red circles mark residues with no resonance in free PTPN3-PDZ; grey circles mark residues whose behavior could not be safely defined, mainly because they fall in crowded spectral regions. **(B)** Two views of the structure of PTPN3-PDZ bound to HPV16E6 PBM (PDB ID 6HKS) with different colors depending on $\Delta\delta$ values. Blue: $\Delta\delta > 0.15$ ppm; red: resonances absent from the free PTPN3-PDZ spectrum; grey: residues whose resonances fall in crowded spectral regions. HPV16E6 PBM peptide is represented in yellow. Side chains of affected residues are represented as sticks.

The PDZ domain of PTPN3 as a target of viruses. We showed that the viral PBMs interact with PTPN3-PDZ with similar affinities to that of the endogenous PTPN3 ligand p38 γ . Despite variations in the PBM sequences, all the affinities fall in the 25–55 μM range (Table 1), in agreement with affinities previously reported for HPV E6 PBM PDZ binders¹². This result is consistent with the hypothesis of a competition between cellular PBM-containing partners and PBMs on viral proteins to bind PTPN3 and hijack signalling pathways in infected cells. E6 binding to PTPN3 should be favoured in HPV infected cells since E6 is highly expressed in HPV-derived cancer cells. Only the C-termini of E6 oncoproteins in high-risk HPV strains contain PBMs, and this is a marker of high oncogenic potential. The presence of a PBM on E6 confers the capacity to interact with various cellular PDZ domain-containing proteins including PTPN3²⁹. Furthermore, this interaction causes degradation of most of its targets through the ubiquitin–proteasome pathway³⁰. Therefore, PDZ–protein interactions with E6 PBM not only disrupt protein–protein interactions but also promote degradation of the PDZ-containing protein. Our results raise the question of how specificity is achieved in this interaction, and whether or not PTPN3 PDZ domain is able to discriminate its binding partners.

Molecular basis and specificities of the interaction with the PDZ domain of PTPN3. The crystal structure and the NMR chemical shift mapping of PDZ–PTPN3 in complex with the PBM of the E6 oncogenic protein of HPV16 highlight the main structural determinants of recognition of the C-terminal sequence of the E6 protein. HPV16E6 PBM binding to PTPN3-PDZ is consistent with the binding mode observed for canonical class I PDZ domains, where L_0 and T_{-2} are involved in key interactions with the PDZ domain. Indeed, L_0 interacts with the conserved GLGF motif, which is crucial for the hydrogen bond coordination of the terminal carboxylate group during the PDZ domain–PBM ligand interaction. Moreover, the T_{-2} forms a hydrogen bond with a conserved Histidine or Arginine in the PDZ domain, signature of class I PDZ domains^{14,31}. The histidine, H572 in PTPN3, is very well conserved in the PTPN3 and PTPN4 orthologs (Fig. 6), and was also identified as a preferred



Figure 6. Sequence alignment of PTPN3 and PTPN4 from different species.

residue for E6 binding by the alignment of 209 PDZ domains ranked according to their highest E6 binding intensity, as determined by a high-throughput assay¹².

At position -1 of HPV16E6 PBM, a glutamine forms a H-bond with a water molecule that is also H-bonded to the N524 of the PDZ domain of PTPN3, whereas the N531 of PTPN4 forms a H-bond with the D in position -1 in the PBM of the cellular partner GluN2A-16¹¹. This asparagine is very well conserved in the orthologs of PTPN3 and PTPN4, while a serine or threonine is more commonly found at this position in the PDZome.

N524 and S538 (S545 for PTPN4) form also H-bonds with the E in position -3 of the HPV16E6 PBM (Figs 3B and 4A). S538 and N524 are strictly conserved in PTPN3 and PTPN4 orthologues, with only 2 and 1 exceptions respectively (Fig. 6); although in the PDZome, S or T residues are often located at this position 524, and in position 538, we often find S or T but also K, H, Q and R. E₋₃ is also stabilized by K526 in PTPN3. This K is conserved in PTPN3 and PTPN4 orthologues but only a few K are present at this position in the human PDZome and I, V, A, S and M are more often found. We previously reported that one single mutation (Q to E) at position -3 of the PBM of the envelope G protein of rabies virus (RABV) switches the fate of the infected cell from survival to apoptosis³². Indeed, the Q to E change increases the number and change the pattern of cellular partners for the PBM in the infected cells, enabling the attenuated RABV strain G protein to interact with PTPN4 inducing cell death. PTPN4-PDZ is able to discriminate between these peptides, displaying a substantially higher affinity for E rather than Q in position -3. Altogether, these observations and the fact that a glutamic acid is often found in position -3 in all PBMs interacting with PTPN3 and PTPN4^{3,4,10,11,19,20,32,33} suggest that interactions with this E in position -3 are a critical determinant of specificity of PDZ/PBM interaction in the NT5 group comprising PTPN3 and PTPN4 phosphatases.

Finally, the arginine in position -4 is linked by ionic bonds to D573 in PTPN3-PDZ (D580 in PTPN4-PDZ) complexed to HPV16E6 PBM. This interaction is most likely maintained in the complexes with viral HPV18E6 PBM, HBVc PBMs, and even with the p38 γ PBM, which presents a lysine instead of an arginine. D573 is well conserved in the orthologous PTPN3 and PTPN4 with only 3 exceptions, yet it is not the most frequently found amino acid at this position in the human PDZome. It might also be a specificity determinant common to the PDZ of PTPN3 and PTPN4 phosphatases. These two closely-related non-receptor PTPs that compose the NT5 sub-family emerged in holozoa³⁴ and duplicated in vertebrates with 54% of sequence identity, which rises to 71% for their PDZ domains. However, the expression pattern, substrates and interacting partners of PTPN3 and PTPN4 have a limited overlap.

Methods

Production and purification of recombinant proteins and PBM ligands. PTPN3-PDZ and PTPN3-PDZ_{Next} (Fig. 1) are encoded as an N-terminal glutathione S-transferase (GST) tagged protein in a pDEST15 expression plasmid and a pGST//2 (derived from pGEX-4T-1; Amersham) expression plasmid respectively. A TEV cleavage site is introduced between the N-terminal tag and the protein sequence. The vectors were used to transform *E. coli* BL21 Star (DE3) star cells (Invitrogen, Carlsbad, CA, USA). Uniformly ¹⁵N-labeled, ¹³C, ¹⁵N-labeled and unlabeled PTPN3-PDZ and PTPN3-PDZ_{Next} constructs were expressed and purified as previously described⁹.

Briefly, harvested cells were resuspended in buffer A (50 mM Tris/HCl, pH 7.5, 150 mM NaCl), 2 mM β-mercaptoethanol and protease inhibitor cocktail (Roche), and then disrupted in a French press. The clarified supernatants were loaded onto a GST column (GSTrap HP, GE Healthcare) and washed with the same buffer. The GST tag was cleaved by overnight incubation at 4 °C by TEV protease (1% mol/mol) directly injected into the column. The eluted fractions containing the protein were pooled and loaded onto a size exclusion column (HiLoad Superdex 75 pg; GE) equilibrated with buffer A with 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP). For crystallization of PTPN3-PDZ_{Next}, the same protocol was followed, replacing the Tris/HCl in buffer A by 20 mM HEPES pH 8 on the size exclusion chromatography step. Purified proteins were concentrated using centrifugal filter devices (Vivaspin, Sartorius). Protein concentration was estimated from its absorbance at 280 nm.

The peptides, p38γ PBM, HBVc PBM, HPV16E6 PBM and HPV18E6 PBM, were synthesized in solid phase using Fmoc strategy (Proteogenix) and resuspended in H₂O.

CD experiments. All CD measurements were acquired with an Aviv 215 spectropolarimeter. Far-UV (195–240 nm) spectra were recorded at 25 °C on 8.4 μM PTPN3-PDZ samples in a cylindrical cell with a 0.2-mm path-length. Ellipticity was measured every 1 nm. The final spectrum of the protein sample was obtained by averaging three successive scans and subtracting the baseline spectrum of the buffer recorded under the same conditions. The CONTIN program was used for quantitative decomposition of the far-UV CD spectrum³⁵.

NMR experiments. The NMR samples for the PTPN3-PDZ and PTPN3-PDZ_{Next} constructs were prepared in buffer A with 0.5 mM TCEP and D₂O (5–10% vol:vol). All NMR experiments were performed on a 600-MHz Varian NMR System spectrometer equipped with a triple resonance ¹H{¹³C/¹⁵N} cryoprobe.

The NMR titration experiments to measure PTPN3-PDZ-PBM peptide affinities and the NMR experiments for backbone assignment of PTPN3-PDZ in complex with HPV16E6 PBM were performed with the PTPN3-PDZ construct at 15 °C. Briefly, the unlabeled peptides (stock solutions ranging from 2.8 to 5.7 mM) at pH 7.5 were added stepwise in a sample initially containing 240–260 μL of ¹⁵N-labeled PTPN3-PDZ at a concentration of 95 or 149 μM. A series of ¹H, ¹⁵N HSQC spectra was recorded for the different titration points. Chemical shift changes were calculated using the free PTPN3-PDZ signals as a reference. Chemical shift differences in the cross-peaks by titration were calculated using the relationship: $\Delta\delta = ((\Delta\delta_{\text{HN}})^2 + (0.15 \times \Delta\delta_{\text{N}})^2)^{0.5}$ where $\Delta\delta_{\text{HN}}$ and $\Delta\delta_{\text{N}}$ are changes in the ¹H and ¹⁵N chemical shifts in ppm, respectively. The weighting factor of 0.15 was used to adjust the relative magnitudes of the amide nitrogen chemical shift range and the amide proton chemical shift range. Signals broaden in the moderate fast-exchange regime observed with PTPN3-PDZ and the PBM peptides, increasing the experimental errors on the chemical shift measurements used for the fitting of the K_d. The following of chemical shift changes during titrations and the fitting of curves were performed with the CcpNmr Analysis software³⁶. A pool of 8 to 14 peaks with the best fit for each titration were kept to deduce the K_D, and the errors are the standard deviations of all the K_D values fitted from the curves.

The sequence specific ¹H^N, ¹⁵N, ¹³C^α, ¹³C^β and ¹³CO resonance assignments were performed using TROSY-based versions of the following experiments: 2D ¹H, ¹⁵N HSQC and 3D HNC(O), HN(CO)CA, CBCA(CO)NH and HNCACB. The spectra were processed with NMRPipe³⁷ and subsequently analysed in CcpNmr Analysis software³⁶. The chemical shifts have been deposited in the BioMagResBank (BMRB) under accession number 27645.

A series of ¹H, ¹⁵N HSQC spectra was recorded over a week at 20 °C for a sample of PTPN3-PDZ_{Next} at 170 μM to follow the stability of the construct.

DSC experiments. DSC experiments were carried out on a Microcal VP-capillary DSC (Malvern). 18, 21, 94 and 28 μM of PTPN4-PDZ, free PTPN3-PDZ, free PTPN3-PDZ_{Next} and PTPN3-PDZ complexed with peptides respectively were scanned from 10 °C to 90–100 °C with a scan rate of 200 °C/h. Large excess of peptides were used: about 600 μM for p38γ PBM, about 300 μM for HBVc PBM and HPV16E6 PBM and about 200 μM for HPV18E6 PBM. The buffer A was repeatedly scanned to ensure a stable buffer baseline. The buffer baseline was subtracted from the protein thermogram. Subsequent to normalization of the data by the protein concentration, a non-linear curve fitting algorithm was employed to obtain the T_m of the transition. All experiments were repeated three times.

AUC experiments. Sedimentation velocity experiments were carried out at 20 °C using a Beckman Coulter XL-I centrifuge equipped with a AN60-Ti rotor. Various protein and protein-peptide complex concentrations (protein concentrations ranging from 12.5 to 70 μM) samples were centrifuged for 17 h at 42000 rpm. Data were analyzed with SEDFIT 15.1 using a continuous size distribution c(S) model. The partial specific volume, the viscosity and the density of the samples were calculated with SEDNTERP. The processed data were used to obtain values of sedimentation coefficients at null concentration in our experimental conditions (S₀) and to get the standard sedimentation coefficients in water at 20 °C (S_{0,w,20}) (Table 2).

Crystallisation, data collection, and structure determination. The HPV16E6 PBM peptide used for co-crystallization was added in excess to form >95% of the complex with the protein. The PDZ domain-peptide complex for crystallization was generated by mixing PTPN3-PDZ_{Next} and the peptide at a ratio of 1:2. Initial screening of crystallization conditions was carried out by the vapor diffusion method using a Mosquito™ nanoliter-dispensing system (TTP Labtech). Sitting drops were set up using 400 nL of a 1:1 mixture of each sample protein and crystallization solutions (672 different commercially available conditions) equilibrated against a 150 µL reservoir in multiwell plates (Greiner Bio-One). The crystallization plates were stored at 4 °C in a RockImager1000® (Formulatrix) automated imaging system to monitor crystal growth. The best crystals were obtained by mixing 200 nL of PTPN3-PDZ_{Next}·HPV16E6 PBM complex solution (concentration of the PDZ domain at 4.8 mg/mL) in 20 mM HEPES pH 8, 150 mM NaCl, 0.5 mM TCEP mixed with 200 nL of reservoir solution containing 20% w/v PEG 3350, 0.2 mM KI at pH 7. Crystals were then flash-cooled in liquid nitrogen using Paratone-paraffin 50%(V/V)/50%V(V) oil as the cryoprotectant.

X-ray diffraction data were collected at a wavelength of 0.979 Å on the beamline PROXIMA-2A at Synchrotron SOLEIL (St. Aubin, France). The data were processed with XDS³⁸ and Xdsme³⁹, and other programs from the CCP4 suite⁴⁰. The structures were solved by molecular replacement with PHASER⁴¹ using the search atomic model of PTPN4-PDZ (PDB ID 5EZ0). The locations of the bound peptides were determined from a $F_o - F_c$ difference electron density maps. Models were rebuilt using COOT⁴², and refinement was done with phenix.refine of the PHENIX suite⁴³. The overall assessment of model quality was performed using MolProbity⁴⁴. The crystal parameters, data collection statistics, and final refinement statistics are shown in Table 3. All structural figures were generated with the PyMOL Molecular Graphics System, Version 1.7 (Schrödinger).

Sequence alignment. Sequence of PTPN3 (accession number NP_002820.3) was used as query on the InterEvoAlign server⁴⁵ to retrieve one single homolog per species assessed as probable ortholog through a reciprocal blast search procedure against the non-redundant database. Retrieved full-length sequences were re-aligned using MAFFT⁴⁶ and displayed using Jalview⁴⁷.

Data Availability

All data generated or analysed during this study are included in this published article.

References

- Gjörloff-Wingren, A. *et al.* Subcellular localization of intracellular protein tyrosine phosphatases in T cells. *Eur. J. Immunol.* **30**, 2412–2421 (2000).
- Hendriks, W. J. A. J. & Böhmer, F.-D. *Non-transmembrane PTPs in Cancer in Protein Tyrosine Phosphatases in Cancer* (ed. Neel B. G. & Tonks N. K.). (Springer, 2016).
- Töpffer, S., Müller-Schiffmann, A., Matentzoglou, K., Scheffner, M. & Steger, G. Protein tyrosine phosphatase H1 is a target of the E6 oncoprotein of high-risk genital human papillomaviruses. *J. Gen. Virol.* **88**, 2956–2965 (2007).
- Hsu, E.-C. *et al.* Suppression of hepatitis B viral gene expression by protein-tyrosine phosphatase PTPN3. *J. Biomed. Sci.* **14**, 731–744 (2007).
- Human papillomavirus (HPV) and cervical cancer. Available at: [http://www.who.int/en/news-room/fact-sheets/detail/human-papillomavirus-\(hpv\)-and-cervical-cancer](http://www.who.int/en/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer). (Accessed: 17th September 2018).
- Thomas, M., Myers, M. P., Massimi, P., Guarnaccia, C. & Banks, L. Analysis of Multiple HPV E6 PDZ Interactions Defines Type-Specific PDZ Fingerprints That Predict Oncogenic Potential. *PLoS Pathog.* **12**, e1005766 (2016).
- Chen, K.-E. *et al.* Reciprocal allosteric regulation of p38 γ and PTPN3 involves a PDZ domain-modulated complex formation. *Sci. Signal.* **7**, ra98 (2014).
- Chen, K.-E. *et al.* Substrate specificity and plasticity of FERM-containing protein tyrosine phosphatases. *Struct. Lond. Engl.* **19**(3), 653–664 (2015).
- Maisonneuve, P. *et al.* Regulation of the catalytic activity of the human phosphatase PTPN4 by its PDZ domain. *FEBS J.* **281**, 4852–4865 (2014).
- Maisonneuve, P. *et al.* Molecular Basis of The Interaction of the Human Protein Tyrosine Phosphatase Non-receptor Type 4 (PTPN4) with the Mitogen-Activated Protein Kinase p38 γ . *J. Biol. Chem.*, <https://doi.org/10.1074/jbc.M115.707208> (2016).
- Babault, N. *et al.* Peptides targeting the PDZ domain of PTPN4 are efficient inducers of glioblastoma cell death. *Struct. Lond. Engl.* **19**(19), 1518–1524 (2011).
- Vincentelli, R. *et al.* Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat. Methods* **12**, 787–793 (2015).
- Christopherson, K. S., Hillier, B. J., Lim, W. A. & Bredt, D. S. PSD-95 assembles a ternary complex with the N-methyl-D-aspartic acid receptor and a bivalent neuronal NO synthase PDZ domain. *J. Biol. Chem.* **274**, 27467–27473 (1999).
- Manjunath, G. P., Ramanujam, P. L. & Galande, S. Structure function relations in PDZ-domain-containing proteins: Implications for protein networks in cellular signalling. *J. Biosci.* **43**, 155–171 (2018).
- Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
- Ortega, A., Amorós, D. & García de la Torre, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **101**, 892–898 (2011).
- Lee, H.-J. & Zheng, J. J. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun. Signal.* **8**, 8 (2010).
- Li, M.-Y. *et al.* Protein tyrosine phosphatase PTPN3 inhibits lung cancer cell proliferation and migration by promoting EGFR endocytic degradation. *Oncogene* **34**, 3791–3803 (2015).
- Hou, S.-W. *et al.* PTPH1 dephosphorylates and cooperates with p38 γ MAPK to increase ras oncogenesis through PDZ-mediated interaction. *Cancer Res.* **70**, 2901–2910 (2010).
- Zheng, Y., Schlondorff, J. & Blobel, C. P. Evidence for regulation of the tumor necrosis factor α -convertase (TACE) by protein-tyrosine phosphatase PTPH1. *J. Biol. Chem.* **277**, 42463–42470 (2002).
- Wang, C. K., Pan, L., Chen, J. & Zhang, M. Extensions of PDZ domains as important structural and functional elements. *Protein Cell* **1**, 737–751 (2010).
- Charbonnier, S. *et al.* The structural and dynamic response of MAGI-1 PDZ1 with noncanonical domain boundaries to the binding of human papillomavirus E6. *J. Mol. Biol.* **406**, 745–763 (2011).

23. Terrien, E. *et al.* Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. *Sci. Signal.* **5**, ra58 (2012).
24. Delhommel, F. *et al.* Deciphering the unconventional peptide binding to the PDZ domain of MAST2. *Biochem. J.* **469**, 159–168 (2015).
25. Fuentes, E. J., Der, C. J. & Lee, A. L. Ligand-dependent Dynamics and Intramolecular Signaling in a PDZ Domain. *J. Mol. Biol.* **335**, 1105–1115 (2004).
26. Hultqvist, G. *et al.* Energetic Pathway Sampling in a Protein Interaction Domain. *Structure* **21**, 1193–1202 (2013).
27. van den Berk, L. C. J. *et al.* An Allosteric Intramolecular PDZ–PDZ Interaction Modulates PTP-BL PDZ2 Binding Specificity†. *Biochemistry* **46**, 13629–13637 (2007).
28. Chang, B. H. *et al.* A Systematic Family-wide Investigation Reveals that ~30% of Mammalian PDZ Domains Engage in PDZ–PDZ Interactions. *Chem. Biol.* **18**, 1143–1152 (2011).
29. Ganti, K. *et al.* The Human Papillomavirus E6 PDZ Binding Motif: From Life Cycle to Malignancy. *Viruses* **7**, 3530–3551 (2015).
30. Javier, R. T. & Rice, A. P. Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. *J. Virol.* **85**, 11544–11556 (2011).
31. Liu, X. & Fuentes, E. J. Emerging Themes in PDZ Domain Signaling: Structure, Function, and Inhibition. In *International Review of Cell and Molecular Biology*, <https://doi.org/10.1016/bs.ircmb.2018.05.013> (Elsevier, 2018).
32. Préhaud, C. *et al.* Attenuation of rabies virulence: takeover by the cytoplasmic domain of its envelope protein. *Sci. Signal.* **3**, ra5 (2010).
33. Hironaka, K., Umemori, H., Tezuka, T., Mishina, M. & Yamamoto, T. The protein-tyrosine phosphatase PTPMEG interacts with glutamate receptor delta 2 and epsilon subunits. *J. Biol. Chem.* **275**, 16167–16173 (2000).
34. Chen, M. J., Dixon, J. E. & Manning, G. Genomics and evolution of protein phosphatases. *Sci Signal* **10**, eaag1796 (2017).
35. Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260 (2000).
36. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
37. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
38. Kabsch, W. X. D. S. & Acta Crystallogr., D. Biol. Crystallogr. **66**, 125–132 (2010).
39. Legrandp, Aishima, J. & CV-GPhL. Legrandp/Xdsme: Version Working With Xds Version: June 1, 2017 And Previous., <https://doi.org/10.5281/zenodo.837885> (2017).
40. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
41. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 32–41 (2007).
42. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
43. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
44. Lovell, S. C. *et al.* Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **50**, 437–450 (2003).
45. Faure, G., Andreani, J. & Guerois, R. InterEvol database: exploring the structure and evolution of protein complex interfaces. *Nucleic Acids Res.* **40**, D847–856 (2012).
46. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
47. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinforma. Oxf. Engl.* **25**, 1189–1191 (2009).

Acknowledgements

Mariano Genera was supported by the Pasteur - Paris University (PPU) International PhD Program. The authors are grateful to the staff of the Crystallography Platform and the Molecular Biophysics Platform at Institut Pasteur for robot-driven crystallization screening and biophysics expertise respectively. We thank Sébastien Brûlé and Ariel Mechaly for helpful discussions. We acknowledge the Synchrotron SOLEIL (St Aubin, France) staff for assistance and advice during data collection on PROXIMA-2A beamline.

Author Contributions

M.G., N.W. and C.C.S. planned the experiments. M.G., D.S., B.R., A.H., B.B., C.S., R.G. and C.C.S. performed the experiments. M.G., D.S., B.R., B.B., R.G. and C.C.S. analysed the data. M.G., N.W. and C.C.S. wrote the paper.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-43932-x>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

5.3 Comments

In this article, we present the first structural data on the PDZ domain of PTPN3, as well as data on binding affinity for its PDZ partners. Although several screenings to obtain the crystallization conditions for the unbound PTPN3-PDZ were performed, I was unable to obtain protein crystals. I used protein concentrations ranging from 4 mg/mL to > 10 mg/mL, temperatures of 18 °C and 4 °C, and a large range of crystallization buffers. Some PDZ domains fail to crystallize in the free form because of higher flexibility/dynamics of the unbound form compared to the complex with a PBM. This could be also due to the tendency of PDZ domains to auto-associate¹⁷⁸. We found that ligand binding prevents auto-association of PTPN3-PDZ, something that had been previously observed in our lab for another PDZ domain¹⁷¹. Most likely, the ligand binding induces changes the dynamics of the PDZ domain and stabilizes the monomeric form at increasing concentrations, facilitating the crystallization process.

Another point worth discussing concerns the measurement of the binding affinities. NMR titration has the advantage of providing information on the state and folding of the labelled protein at every titration point. Moreover, as we showed in our article, the titration lead us to identify the residues involved in the binding of the ligand, the assignment of the backbone resonances being available. However, in our case, the errors on the affinities values introduced with this technique are quite large because of signal broadening in the moderate fast-exchange regime observed by NMR with PTPN3-PDZ and all the peptides, which increases the experimental errors on the chemical shift measurements used for the calculation of the K_D s. In addition, we fitted several titration curves simultaneously (i.e., following the chemical shift of various signals on the spectra) to estimate the K_D ³³⁸. We used 8 to 14 curves for each titration to deduce the K_D and calculated the standard deviation.

Before I joined the project, it was attempted to obtain these measurements by Microscale Thermophoresis and Isothermal Titration Calorimetry. NMR titration proved to be the most efficient way to obtain the affinities. The affinities of PTPN3-PDZ for its ligands fall in the expected tenth-of-micromolar range, and that they are comparable among each other. Moreover, the values obtained by NMR are consistent with those calculated from the binding intensities of the holdup assay (developed in section 7.2).

5.4 Additional results: kinetic studies of PTPN3

5.4.1 Context

PTPN3 is composed of three domains: a N-terminal FERM domain, a central PDZ domain and a C-terminal PTP domain (figure 5.1). When PTPN3 was first purified and biochemically characterized, it was observed that the proteolytic cleavage of the protein increased its catalytic activity⁵⁶. PTPN3 is 913 amino acids long, and the linker that connects the FERM and PDZ domains is about 200 amino acids long predicted mostly unstructured, and thus susceptible of being cleaved by proteases. Indeed, the FERM-PDZ linker was shown to be cleaved *in vitro* by trypsin, releasing a fragment of around 50 kDa that corresponds to the PDZ and PTP domains⁵⁶. Further cleavage resulted in the apparition of a fragment of 39 kDa, and N-terminal sequencing indicated that this fragment started at residue 601, thus encompassing the PTP domain. This separation of the PTP domain was concomitant with an increase of the phosphatase activity⁵⁶. At the time that these experiments were performed, PDZ domains had not been identified, and it was only known that some proteins presented a conserved GLGF motif with unknown functionality.

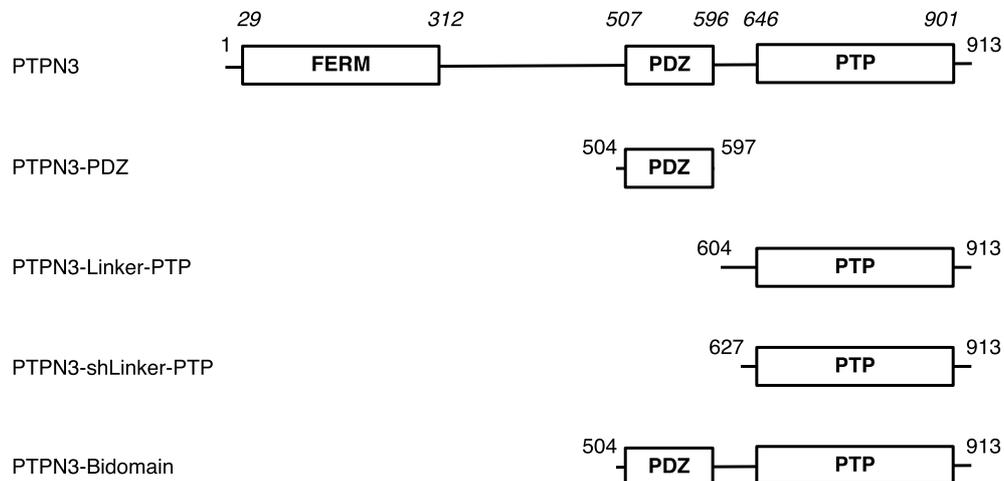


Figure 5.1: Schematic representation of the PTPN3 constructs. The numbers on both extremities indicate the boundary residues for each construct. The numbers in italic above the schematic construct of full-length PTPN3 correspond to the boundary residues of each protein domain.

More recently, it has been reported that the PDZ domains of PTPN3 and of its homolog PTPN4, exert an inhibitory effect on the catalytic activity of the adjacent PTP domain^{70,198}. It was shown in our lab for PTPN4 that binding of a PBM partially releases this

catalytic inhibition¹⁹⁸ and that the linker between the PDZ and the PTP is necessary for the regulation by both the PDZ domain and the PBM^{199,339}. PTPN3 has various reported PDZ ligands, and is targeted through its PDZ domain by oncogenic viruses that encode PBMs. However, it has not been studied whether these interactions have an effect on the catalytic regulation.

To further understand the regulatory role of PTPN3-PDZ, we characterised the kinetic parameters of PTPN3 catalysis, and studied the inhibition by the PDZ domain and the potential effects of PBM binding. Four constructs of PTPN3 were expressed in *E. coli* and purified: PTPN3-Bidomain, PTPN3-linker-PTP, PTPN3-shLinker-PTP (short linker, missing 23 residues) and PTPN3-PDZ (figure 5.1). In all experiments, the phosphatase activity was assessed using the synthetic nonspecific phosphatase substrate p-nitrophenyl phosphate (pNPP), whose hydrolysis into p-nitrophenol (pNP) by phosphatases can be followed spectrophotometrically at 410 nm.

5.4.2 Results

5.4.2.1 Kinetic parameters of PTPN3. Effect of the PDZ domain

First, the kinetic parameters of the dephosphorylation reaction catalysed by PTPN3-Bidomain and PTPN3-linker-PTP were analysed. The initial reaction rates were measured independently at pH 7.5 and at 25°C for a 1 to 10 mM range of pNPP concentrations. The dephosphorylation reaction followed Michaelis-Menten kinetics and exhibited a substrate inhibition effect at high concentrations of pNPP (figure 5.2.A). The experimental data was therefore fitted to a corrected Michaelis-Menten equation to take into consideration this inhibition. The Michaelis constant (K_M) and the turnover number (k_{cat}) derived from adjusting the experimental data to this model are listed in table 5.1. In agreement with previous reports, the catalytic activity of PTPN3-Bidomain was lower than the one of PTPN3-linker-PTP. We observe that the PDZ domain affects mainly the k_{cat} , reducing it by half ($3.0 \pm 0.1 \text{ s}^{-1}$ vs $1.5 \pm 0.1 \text{ s}^{-1}$) (table 5.1), whereas the K_M that reflects the affinity for the substrate is unaffected. The K_M is weak with a value around 1 mM, meaning that the enzyme is saturated at only high substrate concentrations.

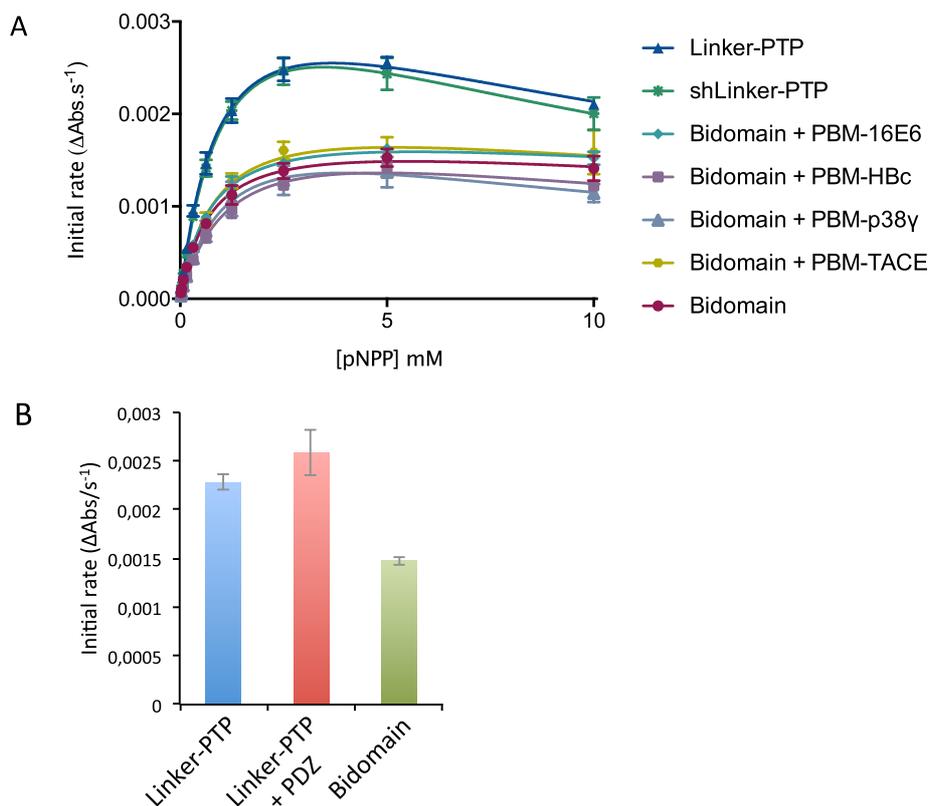


Figure 5.2: Regulation of the phosphatase activity of PTPN3 by its PDZ domain. (A) Michaelis-Menten plots of the initial rates of pNPP hydrolysis by PTPN3 constructs. The K_M and k_{cat} constants were deduced by fitting the data to a modified Michaelis-Menten equation, taking into account the substrate inhibition observed at high concentrations of pNPP. The data and error bars are representative of three independent experiments. The curves are nonlinear fits to a substrate-inhibition equation. (B) Initial rates of pNPP dephosphorylation at 25 mM pNPP by PTPN3-linker-PTP, PTPN3-linker-PTP pre-incubated with PTPN3-PDZ, and PTPN3-Bidomain. PTPN3-linker-PTP and PTPN3-Bidomain were at a concentration of 75 nM, while PTPN3-PDZ was added at a concentration of 6 μM . The data and error bars are representative of three independent experiments.

Table 5.1: kinetic parameters of hydrolysis of pNPP by PTPN3 constructs. The data are representative of three independent experiments.

Construct	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M
Linker-PTP	$1,0 \pm 0,1$	$3,0 \pm 0,1$	2941 ± 241
shLinker-PTP	$1,2 \pm 0,1$	$3,2 \pm 0,2$	2706 ± 340
Bidomain	$0,8 \pm 0,1$	$1,5 \pm 0,1$	1848 ± 234
Bidomain + PBM-HBc	$1,3 \pm 0,1$	$1,6 \pm 0,1$	1233 ± 151
Bidomain + PBM-E6	$0,8 \pm 0,1$	$1,5 \pm 0,1$	2000 ± 180
Bidomain + PBM-p38 γ	$1,2 \pm 0,3$	$1,7 \pm 0,2$	1397 ± 346
Bidomain + PBM-TACE	$0,9 \pm 0,2$	$1,7 \pm 0,1$	1874 ± 490

5.4.2.2 Effect of the binding of PBM ligands to PTPN3-Bidomain

To test whether the binding of a PBM ligand to the PDZ domain releases the catalytic inhibition, we added to PTPN3-Bidomain peptides comprising the PBMs of HPV 16 E6 (PBM-16E6), HPV 18 E6 (PBM-18E6), HBV core (PBM-HBc), and the cellular PTPN3 partners TACE (PBM-TACE) and p38 γ (PBM-p38 γ) (table 5.2). As the binding affinity of PTPN3 for PBM peptides we reported in our article was calculated using the PTPN3-PDZ construct³⁴⁰, I validated that the interaction is not affected in the context of the Bidomain, using PBM-38 γ to measure the interaction by isothermal titration calorimetry. Indeed, I confirmed that PTPN3-Bidomain binds PBM peptides with micromolar affinity. Then, a large excess of each peptide was incubated with PTPN3-Bidomain for 30' at 25°C, and the initial rates of the dephosphorylation reaction were measured in the same range of pNPP concentrations. For all peptides, as in the unbound Bidomain, there is a 0.5-fold decrease in k_{cat} compared to PTPN3-linker-PTP, while the K_M remains largely unaffected (table 5.1). Accordingly, the specificity constants (k_{cat}/K_M) of the complexed Bidomain decrease by 1.5- and 2.4-fold in comparison with the unbound Bidomain. These data confirm the existence of an auto-inhibited state of the enzyme in the Bidomain construct. The binding of PBM ligands of either cellular or viral origin did not affect the regulation of the PTP activity by the PDZ domain in the conditions assayed.

Table 5.2: Affinities of the synthetic PBM peptides. The sequences and nomenclature of the synthetic peptides comprising the PBMs of HPV 16 E6 (PBM-16E6), HPV 18 E6 (PBM-18E6), HBV core (PBM-HBc), TACE (PBM-TACE), and p38 γ (PBM-p38 γ) are indicated. The K_D values correspond to the complex between PTPN3-PDZ and each peptide determined by ¹⁵N-HSQC NMR titration.

Peptide	Sequence	K_D (μ M)
PBM-16E6	RSSRTRRETQL	53 \pm 31
PBM-18E6	RQERLQRRETQV	37 \pm 20
PBM-HBc	RRRRSRSRESQC	29 \pm 24
PBM-p38 γ	SWARVSKETPL	29 \pm 24
PBM-TACE	RQNRVDSKETEC	30 \pm 17

Additionally, we observed that the PTPN3-shLinker-PTP construct, in which the linker lacks the 23 N-terminal residues (figure 5.1.) has the same catalytic activity than

PTPN3-Linker-PTP, with the complete linker (figure 5.2 and table 5.1). This indicates that the linker alone is not sufficient for the inhibition to take place.

5.4.2.3 Role of the interdomain linker

Next, we assessed whether the linker that connects the PDZ and PTP domains is required for the catalytic regulation to take place, as observed for PTPN4. For this, we measured the hydrolysis of pNPP at a fixed concentration of 2.5 mM and at 25°C, where the Bidomain and linker-PTP constructs exhibited the highest initial rate of reaction (figure 5.2.A). We compared the initial rate of the reaction for PTPN3-Bidomain, PTPN3-linker-PTP, and for PTPN3-linker-PTP with PTPN3-PDZ added in trans in large excess (molar ratio 80:1) and incubated for 1h at 4°C. We observed similar initial reaction rates for PTPN3-Linker-PTP with and without PTPN3-PDZ added in trans (figure 5.2.B) whereas as expected, the Bidomain presents a lower initial rate. These results indicate that the regulatory mechanism is not trans-acting and requires the two domains to be covalently linked.

5.4.3 Discussion and perspectives

Protein-protein interaction modules like PDZ domains are often found in combination with catalytic domains, where they can impact the biological function of a protein by enhancing substrate specificity or targeting the protein to macromolecular complexes. PTPN3, for example, although lacking a ‘kinase interaction motif’ is still able to specifically dephosphorylate the MAPK p38 γ thanks to the recognition by its PDZ domain of the C-terminal PBM of p38 γ ⁶⁶. There is an increasing awareness that non-catalytic scaffold domains can, in some cases, perform direct regulatory functions on the catalytic domain to which they are linked, exceeding their traditional roles as inert binding domains. A well-documented case is the regulation of the catalytic activity of the tyrosine phosphatase SHP-2 by its two SH2 adjacent domains³⁴¹. The unbound N-terminal SH2 domain of SHP-2 interacts with the phosphatase domain, sterically blocking the active site in an open but inactive conformation, preventing the closure of the WPD loop resulting in competitive inhibition. The binding of a phosphoprotein ligand to the SH2 domain triggers allosteric conformational rearrangements that prevent binding of the complexed SH2 to the PTP domain, releasing the inhibition³⁴¹. The SH2 domain thus works as an allosteric molecular switch. Similarly, the PDZ domain of PTPN4 locks the phosphatase in an auto-inhibited conformation, and the catalytically active state is restored upon binding of a PBM¹⁹⁸. Both the inhibition and release of inhibition

processes occur through long-range intramolecular allosteric mechanisms that require the covalent binding of the two domains. In the case of PTPN4, the inhibition is non competitive (both k_{cat} and K_{M} affected, the specific activity decreases).

Here, I performed a kinetic characterization of the phosphatase activity of PTPN3 in the context of the isolated PTP domain and the PDZ-PTP bidomain construct. The PDZ domain inhibits the activity of the adjacent PTP domain by decreasing the turnover number, without affecting the affinity for the substrate. This indicates that the PDZ domain is not blocking the accessibility of the phosphatase substrate to the PTP active site. Therefore, the inhibition is non competitive, as found for PTPN4. Similar studies of PTPN3 activity were done using phosphorylated full-length p38 γ , which is both a substrate of PTPN3 and a ligand of its PDZ domain through a PBM found in its C terminus⁷⁰. It was observed that the K_{M} is lower with the Bidomain compared to the PTP (K_{M} of 3 μM and 16 μM respectively), which can be attributed to the presence of the PBM.

We were interested in exploring whether PTPN3 features a similar allosteric inhibitory mechanism as PTPN4. We have previously shown by sequence analysis of orthologues that a conserved hydrophobic FQYI sequence in the PDZ-PTP linker in PTPN4 is implicated in the regulation³³⁹. However, these positions are not strongly conserved between PTPN3 and PTPN4 (figure 5.3). The linker of PTPN4 is largely unstructured and flexible as monitored by NMR, yet resistant to proteolysis, which could indicate that it interacts with the PTP and PDZ domains with enough frequency to protect it against proteolysis³³⁹. On the contrary, secondary structure prediction indicates that the linker of PTPN3 is likely to feature a α -helical region (figure 5.4). We also found that in purified PTPN3-Bidomain samples, the linker is sensitive to proteolysis even in the presence of protease inhibitors (figure 5.5), indicating that it is predominantly exposed. This could be an indication that the linker does not actively participate in the regulation by interacting with the PTP or PDZ domains. This is supported by our observation that the PTPN3-shLinker-PTP construct, which features a truncated linker, displays the same catalytic activity as PTPN3-Linker-PTP, in which the linker is complete.



Figure 5.3: Sequence alignment of the PDZ-PTP linker of PTPN3 and PTPN4 homologs. The sequences of PTPN3 homologs are shown on the top, and those of PTPN4 (highlighted in grey) are below. The black box indicates the conserved FQYI sequence in PTPN4 homologs. The red triangles above the PTPN3 sequence indicate the conserved V599, F602, and F615 positions.

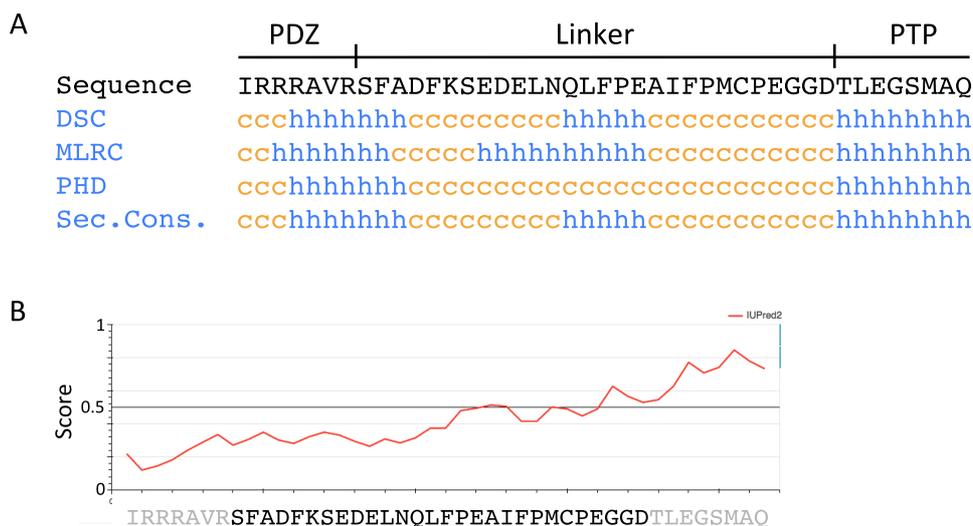


Figure 5.4. Secondary structure prediction for the PTPN3 PDZ-PTP linker. (A) Consensus secondary structure predictions for the PTPN3 linker. For the different softwares used: h is the prediction of α -helix, e of β -strand, and c of unstructured (<http://npsa-pbil.ibcp.fr>). (B) Disordered

region prediction of the PTPN3 PDZ-PTP linker using IUPred2A. The red curve indicates disorder tendency (<https://iupred2a.elte.hu>).

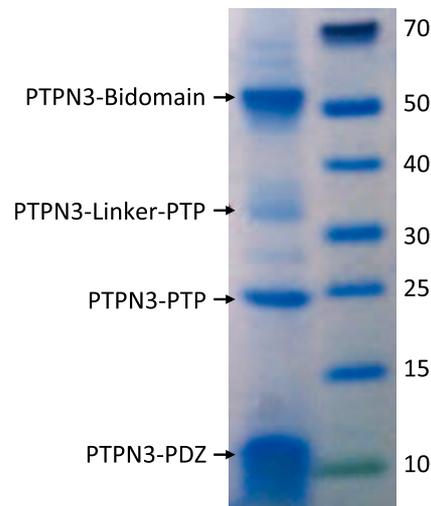


Figure 5.5: Sensitivity of PTPN3-Bidomain to proteolysis. PTPN3-Bidomain was purified in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM TCEP and cOmplete protease inhibitor cocktail (Roche), and kept at 25°C for 15h. The state of the sample was then checked by SDS-PAGE, showing that a large proportion of the protein had been cleaved at the interdomain linker.

A next step in this work could be to evaluate the specific role of the interdomain linker. One possibility is that specific conserved positions establish interactions with the domains, triggering allosteric mechanisms, as previously observed for PTPN4³³⁹. The sequence alignment of PTPN3 homologues shows that V599, F602, and F615 are conserved (figure 5.3), and are thus good candidates for a mutational analysis. However, given the small effects observed in PTPN4, and considering that PTPN3 is less active than PTPN4, it is possible that any effects would be difficult to detect.

Direct regulation of the adjacent catalytic domains is a relatively unknown and poorly documented function of PDZ domains. The proteases HtrA2 and DegS present an intramolecular inhibitory mechanism of the catalytic domain by the PDZ domain^{342,343}. In these two cases, as it happens for PTPN4, the binding of a PDZ ligand releases the inhibition. Our data showed that in the case of PTPN3, the binding of peptides to the PDZ domain does not affect the catalytic regulation. We cannot exclude that the full-length PBM-containing proteins could be necessary to have an effect on the regulation of the catalytic activity, instead of a PBM peptide. PTPN3 and PTPN4 share 51% of global sequence identity, but this rises to 71% for their PDZ domains and 61% for their catalytic domains. NMR mapping of the chemical shift changes that occur in PTPN3-PDZ upon ligand binding shows induced perturbations that extend from the peptide-binding groove until the opposite side of the domain³⁴⁰. This raises the question whether these perturbations could be part of an allosteric

network that could eventually trigger long-range conformational changes in PTPN3. Studying the internal dynamic changes of free and bound PTPN3-Bidomain by NMR would provide information about any structural rearrangements that might occur upon PDZ ligand binding. ^{15}N relaxation experiments detail the local and global dynamics of a protein. The ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates can be used to probe the relative motions of the domains in the free and liganded forms. Unfortunately, our attempts to obtain useful HSQC spectra of PTPN3-Bidomain were not successful. In agreement with our previous results, the dynamics of the construct and a concomitant proteolysis of the linker prevent us to record usable spectra (figure 5.6).

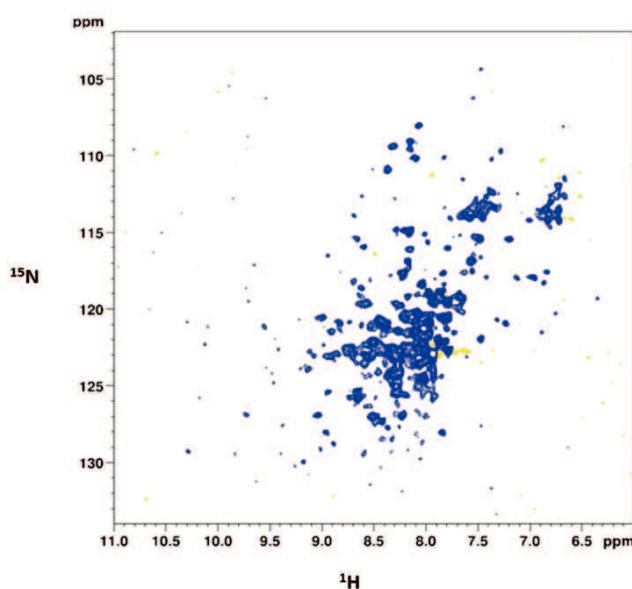


Figure 5.6: ^1H - ^{15}N TROSY spectra of PTPN3-Bidomain at $70\mu\text{M}$, 25°C , recorded at 600 MHz.

5.4.4 Materials and methods

Production and Purification of Recombinant Proteins

PTPN3-PDZ is encoded as an N-terminal glutathione S-transferase (GST) tagged protein in a pDEST15 expression plasmid. PTPN3-Bidomain and PTPN3-linker-PTP are encoded as N-terminal 6xHis tagged proteins in pET15b expression plasmids. In the three cases, a TEV cleavage site is inserted between the N-terminal tags and the protein sequences. The vectors were used to transform *E. coli* BL21 Star (DE3) star cells (Invitrogen, Carlsbad, CA, USA). PTPN3-PDZ, PTPN3-Bidomain and PTPN3-linker-PTP constructs were expressed and purified as previously described¹⁹⁸.

Briefly, harvested cells were resuspended in buffer A (50 mM Tris/HCl, pH 7.5, 150 mM NaCl), 2mM β -mercaptoethanol and protease inhibitor cocktail (Roche), and then disrupted in a French press. The clarified supernatants were loaded onto a GST column (GSTrap HP, GE Healthcare) or a nickel affinity chromatography column (HiTrap HP, GE healthcare) and washed with the same buffer. The GST/6xHis tag was cleaved by overnight incubation at 4°C by TEV protease (1% mol/mol) directly injected into the column. The eluted fractions containing the protein were pooled and loaded onto a size exclusion column (HiLoad Superdex 75 pg; GE) equilibrated with buffer A with 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP). Purified proteins were concentrated using centrifugal filter devices (Vivaspin, Sartorius). Protein concentration was estimated from its absorbance at 280 nm.

PBM Ligands

The peptides, p38 γ PBM, HBVc PBM, HPV16E6 PBM and HPV18E6 PBM, were synthesized in solid phase using Fmoc strategy (Proteogenix) and resuspended in H₂O.

Enzymatic assays

Phosphatase activity was measured by following the hydrolysis of p-nitrophenyl phosphate (pNPP) as previously described¹⁹⁸. Absorbances were measured continuously at 410 nm for pNP, using a Thermo Scientific UV spectrometer equilibrated at 25 °C. Initial linear reaction rates were calculated during a 60 seconds reaction. The k_{cat} and K_M constants were deduced from fitting the Michaelis-Menten equation with the Prism software. The equation takes into account excess-substrate inhibition.

Reactions were performed in 50mM Tris-HCl, pH 7.5, 1 mM MgCl₂, 150 mM NaCl, 0.5 mM TCEP. pNPP was assayed for concentrations ranging from 19 μ m to 10 mM. Specific activities towards pNPP were measured at an enzyme concentration of 75 nM.

NMR experiments

Uniformly labelled PTPN3-Bidomain samples were expressed as previously described¹⁹⁸ and purified as detailed above. The samples were prepared in buffer A with 0.5 mM TCEP and D₂O (5–10% vol:vol). NMR ¹H, ¹⁵N HSQC spectra were recorded on a 600-MHz Varian NMR System spectrometer equipped with a triple resonance 1H{13C/15N} cryoprobe.

6 Structural determinants of PDZ ligand selectivity for the NT5 subfamily of phosphatases

6.1 Introduction

One of the goals of the project was to gain insights into the way PTPN3 specifically interacts with its viral and cellular PBM-containing partners. I presented in the first chapter of the results the structural analysis of the interaction between PTPN3-PDZ and the PBM of HPV 16 E6 (PBM-16E6). To document the binding mode of PTPN3 towards different PBMs, I solved the crystal structure of the complexes formed by PTPN3-PDZ with peptides comprising the PBMs of HPV 18 E6 (PBM-18E6), HBV Core (PBM-HBc), and the cellular PTPN3 partner TACE (PBM-TACE). In this section, I compare these structures to highlight the interactions that determine the recognition of these ligands by PTPN3-PDZ. Moreover, I am interested in finding which of these specificity determinants are shared with its close homolog PTPN4. Indeed, PTPN3 and PTPN4 compose the NT5 subfamily of non-receptor PTPs and come from a gene duplication during holozoan³⁴⁴. They share the same modular organization and 73% of sequence identity in their PDZ domains, which were inserted during evolution into *Monosiga* PTPN3 (after yeast and before sponge)³⁴⁴. *Monosiga* are the closest unicellular organism to metazoans from an evolutionary point of view. The insertion of a PDZ domain in *Monosiga* as in eukaryotes suggests that PDZ domain appeared concomitantly with multi-cellular organisms.

We have performed a sequence alignment for the PDZ domains of PTPN3 and PTPN4 in different species to identify conserved positions that could be involved in the specificity of recognition of PBMs by these phosphatases. In the following description, the numbering of the residues in PTPN3-PDZ corresponds to the numbering of PTPN3 from Uniprot P26045. The equivalent positions in PTPN4 can be found in figure 6.1.

504 **G**DSYLVLI**R**IT**P**DE**D**G**K**FGFN**L**KGGVD**Q**K**M**PLV**S**RIN**P**ES**P**AD**T**C**I**PK**L**N 554
 515 ----L**V**LI**R**M**K**P**D**EN**G**R**F**GF**N**V**K**GG**Y**D**Q**K**M**PV**I**V**S**R**V**A**P**GT**P**AD**L**C**V**PR**L**N 561

 555 **E**GD**Q**I**V**LI**N**GR**D**ISE**H**TH**D**Q**V**VM**F**IK**A**S**R**ESH**S**REL**A**L**V**I**R**RR 597
 562 **E**GD**Q**V**V**LI**N**GR**D**IA**E**H**T**H**D**Q**V**VL**F**IK**A**S**C**ER**H**SG**E**L**M**LL**V**RP**N** 604

Figure 6.1: Sequence alignment of the PDZ domains of PTPN3 (top) PTPN4 (bottom).

6.2 Results and discussion

6.2.1 Crystal structures of PTPN3-PDZ/peptide ligand complexes

The sequences of the peptides derived from viral and cellular PTPN3 ligands used to form the complexes with PTPN3-PDZ and parameters of the crystal structures (space group, resolution, and $R_{\text{work}}/R_{\text{free}}$ of the structures) are listed in table 6.1. The data collection and refinement statistics are detailed in section 9.4.

Table 6.1: Parameters of the crystal structures of PTPN3-PDZ in complex with PBM peptides.

Peptide	Sequence	Space group	Resolution (Å)	$R_{\text{work}}/R_{\text{free}}$
PBM-16E6	RSS <u>R</u> TR <u>R</u> ET <u>Q</u> L	P 1 21 1	2.19	0.21/0.26
PBM-18E6	RQ <u>E</u> R <u>L</u> O <u>R</u> RE <u>T</u> O <u>V</u>	P 65 2 2	2.10	0.19/0.23
PBM-HBc	RRRR <u>S</u> <u>Q</u> S <u>R</u> ES <u>Q</u> C	P 32 2 1	1.86	0.20/0.23
PBM-TACE	RQ <u>N</u> R <u>V</u> D <u>S</u> K <u>E</u> T <u>E</u> C	P 32 2 1	1.70	0.17/0.20

In all cases, the global PDZ fold is well conserved, and the peptides bind in the conventional mode as an anti-parallel extension to the β 2-strand. Compared to our PTPN3-PDZ structure in complex with PBM-16E6 (PDB ID 6HKS), all structures present a very low root medium square deviation (rmsd) ranging from 0.19 Å to 0.30 Å for the backbone atoms, indicating that none of these peptides produce significant backbone conformational changes when binding to PTPN3-PDZ. All the complexes possess the interaction network of a class I PDZ/PBM interaction (S/T-X- Φ_{COOH}). The C-terminal residues that present a well-defined electron density map are underlined in the peptide sequence in table 6.1.

6.2.2 Specificity of recognition of PTPN3-PDZ

The interactions at positions 0 and -2 are essential in PDZ/PBM recognition (section 2.4.2). Position -2 in particular can be considered as the class determinant for PBMs¹⁶¹. The binding modes of each PBM peptide to PTPN3 are shown in figure 6.2. As expected, the C-terminal carboxylate in each peptide forms three H-bonds with the amide nitrogens of F521, G522 and F523 of the “GLGF motif” on PTPN3-PDZ. The PBM-18E6 V₀ is additionally bonded to the carbonyl of G519 on the α 1- β 1 loop and the N ζ of K580 of the α 2-helix through a molecule of water (figure 6.2.B). In position -2, the S or T side chains form H-bonds with the N ϵ 2 of the H572 at the base of the α 2-helix of PTPN3-PDZ, which is conserved in class I PDZ domains. These interactions found in these two key positions are in agreement with the expected bonding pattern of a class I PBM for all the PBMs tested even when the expected C-terminal hydrophobic amino acid is a cysteine.

In position -1, the Q side chain in the PBM-16E6 and PBM-HBc and the E side chain in the PBM-TACE are forming a H-bond with a water molecule that is in turn bonded to the N δ 2 of N524 in the β 2-strand of PTPN3-PDZ. In PBM-18E6, on the contrary, the side chain is not making any contacts with the PDZ domain (figure 6.2). This position is not considered to be a significant determinant for PDZ/PBM interaction. In fact, position -1 is not specified in any of the three main classes of PDZ domains, although some studies have shown that certain PDZ domains display a high specificity in this position³⁴⁵. For example, in a large-scale study of PDZ domain binding specificity to a random peptide library, the class I PDZ domain of PDLIM-2 binds exclusively PBMs with a W₋₁³⁴⁵.

In the case of PTPN3, sequence analysis of its known partners suggests a bias towards residues with long and polar side chains in position -1 (table 6.1). In line with this, N524 is strictly conserved in PTPN3 and PTPN4 orthologs, while in the PDZome the short polar residues S or T are more often found in this position (figures 6.3 and 6.4). The short side chains of S and T make them unsuited to establish bonds with Q-1 or E-1. Moreover, I describe below that N524 is also H-bonded to position -3 of the PBM that requires also a long side chain. Thus, the conservation of N524 could originate from these two interactions with the -1 and the -3 residues of the PBM.

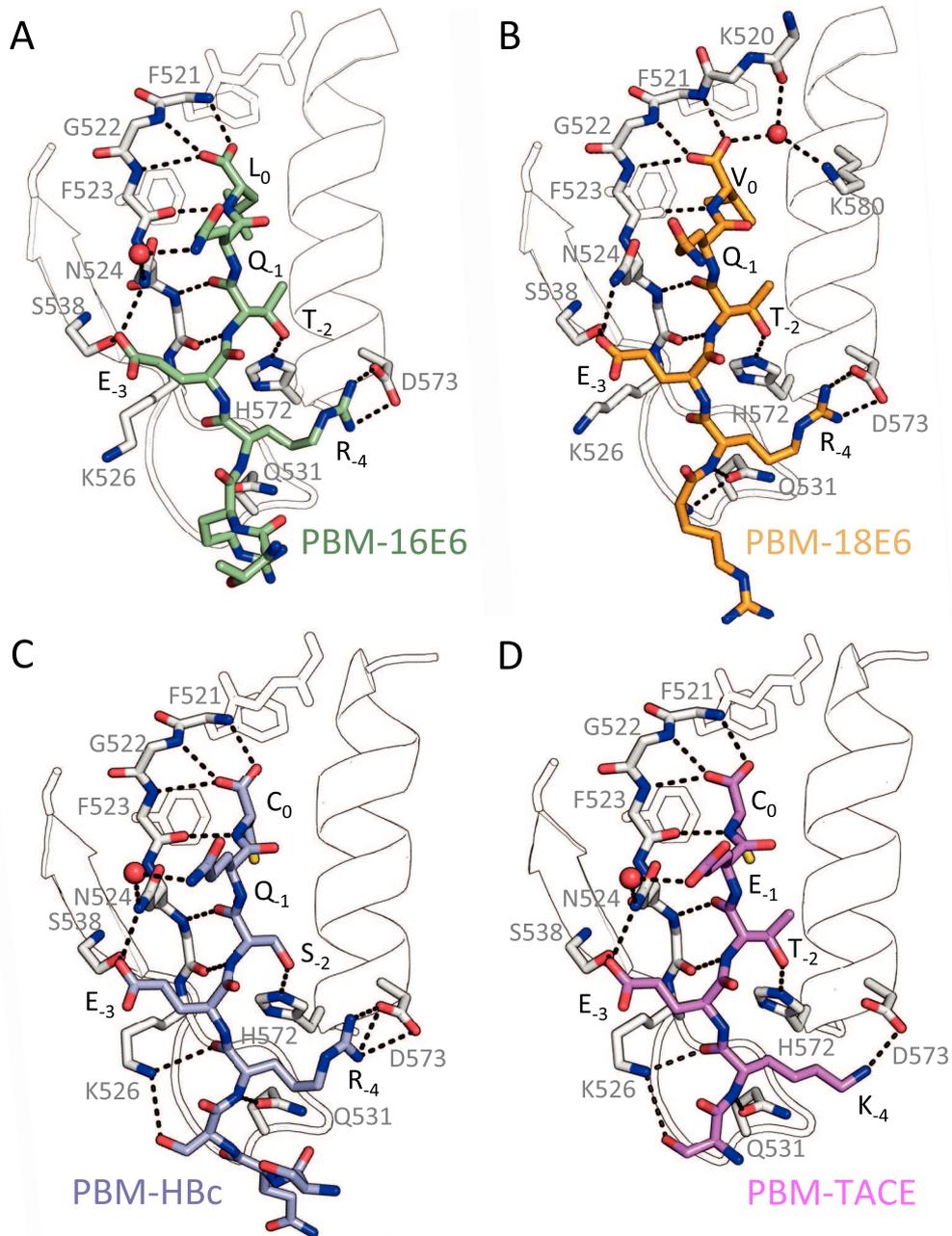


Figure 6.2: Detail of the bonding network of PTPN3-PDZ and the PBM peptides of HPV16 E6, HPV 18 E6, HbC and TACE. Secondary structure elements of PTPN3-PDZ are shown in black trace, and relevant residues are shown as sticks. Peptides are shown as coloured sticks, with the corresponding name indicated below the structure. Polar interactions are shown as black dashes.

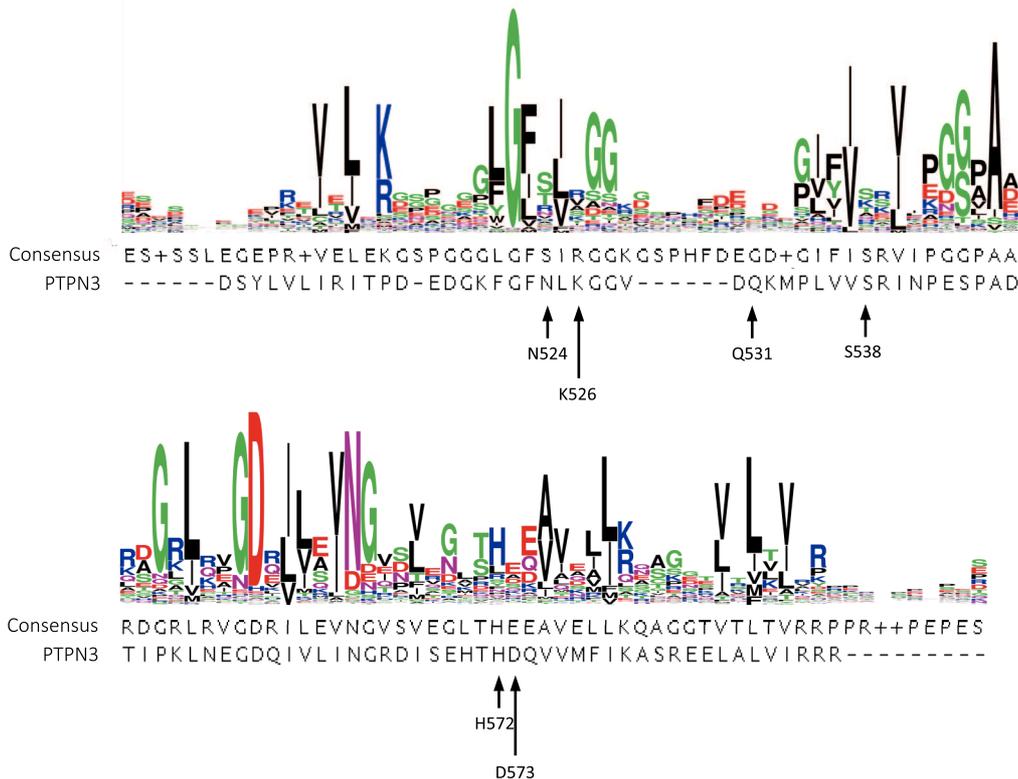


Figure 6.3: Sequence conservation of the PDZome. Sequence logo showing the PDZome conservation. Below, the PDZome consensus sequence and the aligned PTPN3 sequence are shown. Relevant positions in PTPN3 are indicated.

We propose that positions -3 and -4 have a significant influence on PTPN3-PDZ ligand selection. E is found in position -3 in all reported ligands of PTPN3-PDZ. Similarly, most reported ligands of PTPN4-PDZ feature as well an E₋₃, with the exception of the glutamate receptor δ -2 subunit, which has a G₋₃. In this case, the glycine cannot engage in any interactions and a lower affinity of the peptide for PTPN4-PDZ is observed. In all of our PTPN3-PDZ/peptide complexes, the E₋₃ side chain forms a bifurcated H-bond with the amide nitrogen of N524 and with the side chain hydroxyl of S538 (figure 6.2). S538 is also strictly conserved in PTPN3 and PTPN4 orthologs, with only one exception, while in the PDZome this position is conserved only in 25% of cases: although S is most frequently found, also K, T, A, and Q are common (figure 6.3). The aliphatic carbon chain of E₋₃ is also well positioned to establish hydrophobic contacts with the carbon side chain of K526, which contributes to the stabilization of E₋₃. In PTPN3 and PTPN4 orthologs, a K residue is conserved at position 526, while in the PDZome R and V are more commonly found, followed by A, S, and K. In agreement with this, similar hydrophobic contacts have been observed in the complexes of PTPN4-PDZ with the PBMs of p38 γ , the ionotropic glutamate receptor GluN2A, and the

attenuated rabies virus G protein^{199,253} (figure 6.5), suggesting that this is also a specific interaction of the NT5 subfamily.

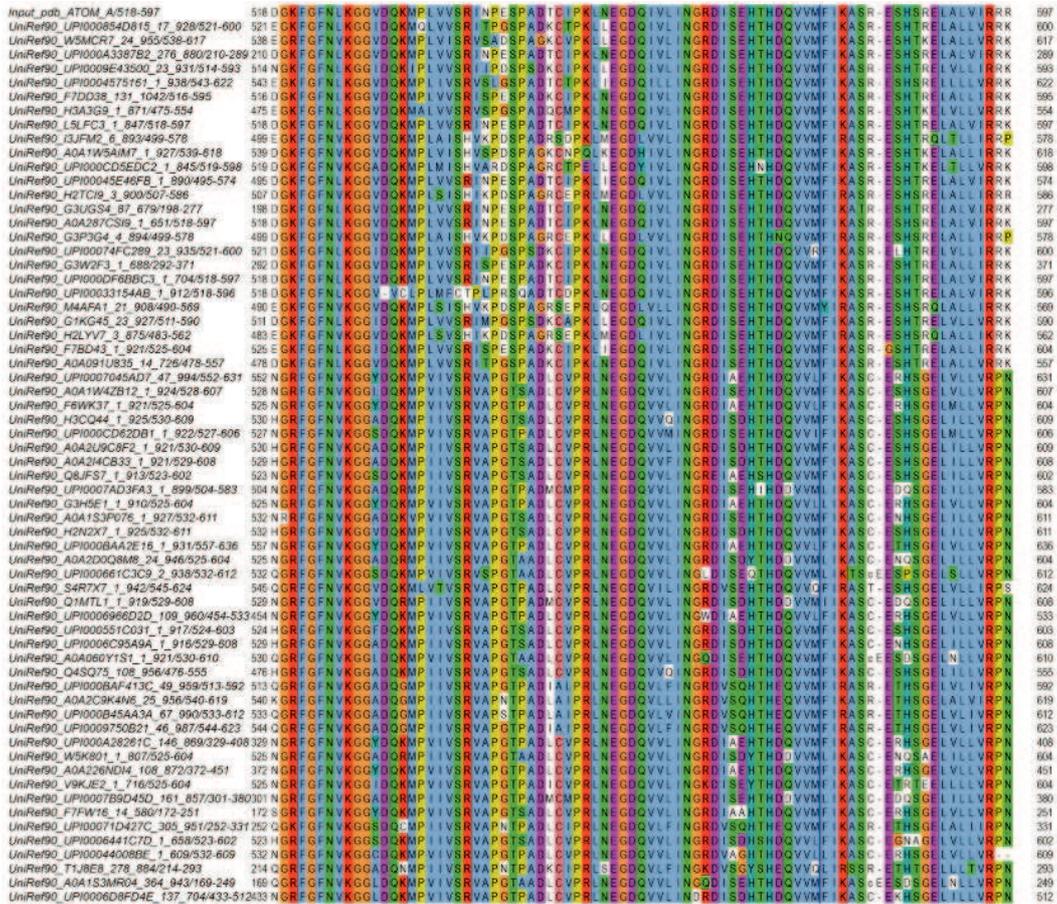


Figure 6.4: Sequence alignment of PTPN3 and PTPN4 homologs.

In PBM-16E6, PBM-18E6, and PBM-HBc, the strictly conserved R₄ forms ionic bonds with the carboxylate oxygens of D573 at the base of the α 2-helix. PBM-TACE features a K₄ instead of R. Its long and positively charge side chain is also able to form an ionic bond with the D573 side chain (figure 6.2). Most PTPN3 and PTPN4 orthologs present D in position 573. In the PDZome, E is most frequently found in this position, although the consensus is less than 30%, followed by A, D, S, Q, and K (figures 6.3 and 6.4). Both E and D should be able to establish ionic bonds with a positively charged residue in position -4 of the PBM, and indeed we find that 9% (6 out of 64) of PTPN3 and 4 orthologs have E instead of D. However, in the two PTPN4-PDZ ligands, attenuated rabies virus G protein and glutamate receptor GluN2A, G and I are found in position -4 respectively (figure 6.5). The side chains of these residues are not able to bind to the D573, and the affinity of the interaction decreases accordingly. Using optimization of sequence targeting PTPN4-PDZ,

demonstration has been done that that D573 is also a significant determinant of the NT5 family binding specificity.

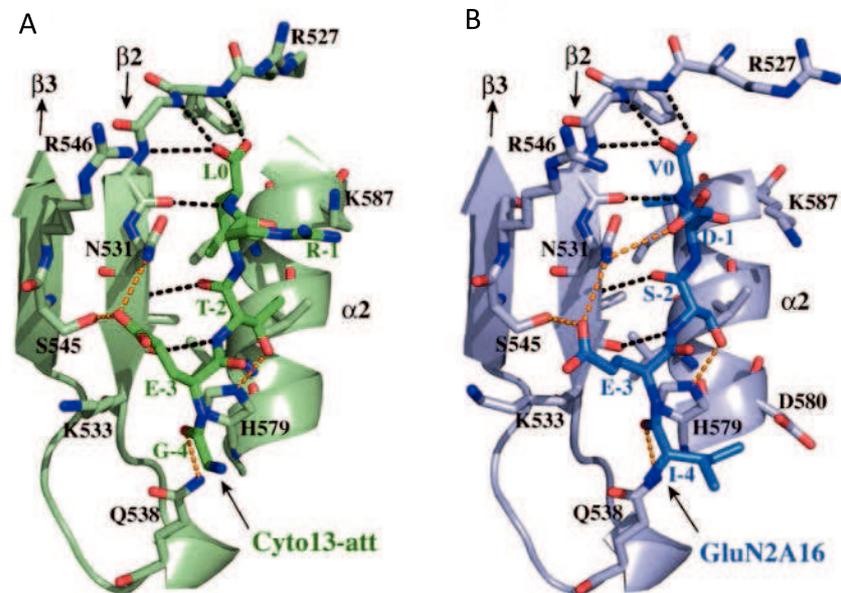


Figure 6.5: Structures of PTPN4-PDZ Bound to Cyto13-att and GluN2A-16. Close-up views of the PTPN4-PDZ/peptide-binding sites. Important residues are shown as sticks in CPK colors. Peptides form an antiparallel β -sheet with the β 2-strand via a complete set of intermolecular canonical backbone hydrogen bonds (black dashed lines). Other intermolecular H-bonds are shown in orange dashed lines. Adapted from Babault *et al.* 2011²⁵³.

Then, in all the PTPN3-PDZ/peptide complexes, the amide N of R₋₅ or K₋₅ forms a H-bond with the N ϵ 2 of the Q531 side chain. Q531 is strictly conserved in 63 out of 64 PTPN3 and 4 orthologs analyzed, but in the PDZome the conservation in this position is very low; G, D, E, S, F, P, and H are all found with higher frequency than Q (figures 6.3 and 6.4). However, although these interactions with Q531 likely contribute to the affinity of the complex, they are not implicated in NT5 subfamily ligand selectivity, as they involve backbone of the PBM and thus any residue (except proline) could fill the position. This can be observed in the complexes of PTPN4-PDZ with the attenuated rabies virus G protein and the glutamate receptor GluN2A, which have G₋₄ and L₄ respectively (figure 6.5). These residues do not establish ionic contacts with the conserved D (D573 in PTPN3) as do R₋₄ and K₋₄ in PTPN3, but both still form H-bonds with Q531.

One final observation is that the side chain of K526 from the β 2-strand is pointed towards the PBM-HBc and PBM-TACE, allowing its N ζ to form H-bonds with the carbonyl oxygen of the K or R in position -4 and with the side chain hydroxyl of S₋₅, contributing to the

stability of the complex. In the other two PTPN3 partners, there is R instead of S in position -5. It appears that the possibility of forming a H-bond with the short, polar side chain of S₋₅ favours the K526 side chain to orient towards the peptide, while in the other cases it adopts an extended conformation to maximize the hydrophobic contacts with the C β -C γ carbon chain of E₋₃.

In conclusion, the interactions with positions -3 and -4 of the PBM appear to be significant determinants of ligand recognition by the PDZ domains of the NT5 subfamily phosphatases. PBM-containing partners with E in position -3 are favoured because of their capacity to form H-bonds with the conserved N524 and S538, as well as hydrophobic contacts with the aliphatic carbon side chain of K526. Additionally, partners with R or K in position -4 are able to form ionic bonds with the conserved D573. Although these interactions are not determinant for the binding to occur, they likely contribute to the affinity of the complex, which will favour their binding over other potential partners.

6.2.3 Recognition of atypical PBMs with a C-terminal cysteine

An interesting feature in the complexes of PTPN3-PDZ with PBM-TACE and PBM-HBc is the presence of the C-terminal cysteine in each PBM. The carboxylate-binding pocket at the top of the peptide-binding groove PDZ domains is lined with hydrophobic side chains (F521, F523, L525, I579 for PTPN3-PDZ), which determines the preference for peptides with hydrophobic C-terminal residues¹⁶¹ (figure 6.6). Although sometimes classified among hydrophobic amino acids since they are frequently found in the interior of proteins, the sulfhydryl group in cysteine is chemically similar to a hydroxyl group, making it a polar residue. We observe that in the complexes of PTPN3-PDZ with PBM-TACE and PBM-HBc, the cysteine side chain is oriented towards the interior of the peptide-binding groove, without making any contacts with the hydrophobic side chains that line this pocket (figure 6.6.B). The cysteine side chain of the PBM C-terminal residue is short enough to fit within the binding pocket. Other polar amino acids with longer side chains in this position might cause steric clashes that could prevent binding of the peptide.

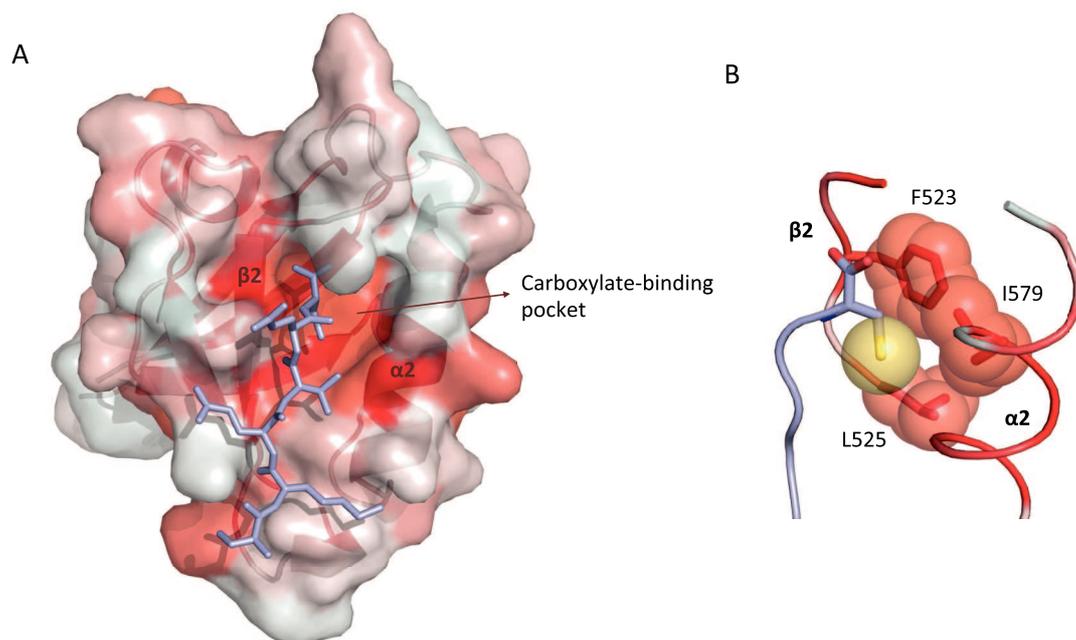


Figure 6.6: Surface hydrophobicity of PTPN3-PDZ. (A) The surface of PTPN3-PDZ is color-coded according to the normalized consensus hydrophobicity scale³⁴⁶. Darker red indicates a more hydrophobic environment, while white indicates polar regions. The peptide shown is PBM-TACE. (B) Detail of the position of the terminal cysteine of PBM-TACE in the carboxylate-binding pocket of PTPN3-PDZ. The spheres represent the Van der Waals radii of the sulphur (yellow) and carbon (red) atoms.

Binding of PBMs with a terminal cysteine (hereafter, C_0 -PBM) has only been reported in a handful of cases, many of which involve the PDZ domain of GIPC1. The PBM of Hbc is an case of C_0 -PBM of bound by GIPC1²⁹⁴. Other examples are the lutropin-choriogonadotropic hormone receptor (LHCGR, $-RYTECCOOH$)³⁴⁷, the complement component C1q receptor (CD93, $-PGTDCOOH$)³⁴⁸, the insulin-like growth factor 1 receptor (IGF1R, $-PQSSTCCOOH$)³⁴⁹, and the dopamine 2 and 3 receptors (DRD2, DRD3), which share the atypical C-terminal sequence $-KILHCCOOH$ ³⁵⁰.

Then, the PDZ domains in other proteins can also bind C_0 -PBMs. Both the MAGI-1 PDZ5 and the Scrib and PDLIM-4 PDZ domains bind the C_0 -PBM of the thyroid receptor interacting protein 6 (TRIP6, $-VTTDCOOH$)^{351–353}. However, there seems to be a lack of structural data on complexes involving this type of PBM. Indeed, the only crystal structure that we found of a PDZ domain bound to a C_0 -PBM is the one of the GRIP1 PDZ6 domain complexed to the liprin- α PBM ($-RTYSCOOH$, figure 6.7)¹⁸⁷. We observe that in this case, just as in the PTPN3-Hbc complex, the cysteine side chain has the same orientation and can be accommodated towards the interior of the hydrophobic carboxylate-binding pocket.

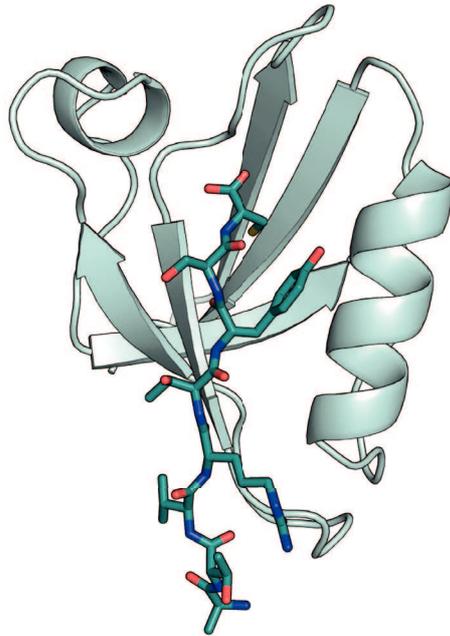


Figure 6.7: Crystal structure of the PDZ6 domain of GRIP bound to the C-terminal peptide of liprin- α , shown as dark green sticks and CPK colors (PDB ID 1N7F).

It is widely considered that the hydrophobic character of the side chains that compose the carboxylate binding-pocket imposes a specificity requirement for PBMs containing hydrophobic residues in this position¹⁵². The preference of some hydrophobic residues over others in P₀ has been attributed to variations in the size and the geometry of the hydrophobic carboxylate-binding pocket¹⁶¹, although it is evident that selectivity at P₀ is not highly stringent. As observed for PTPN3, PDZ domains can accommodate different residues in position 0.

It is interesting that only a few PDZ domains have been reported to bind C₀-PBMs. It is possible that some of these interactions could have been overlooked in some large-scale screenings due to restraints imposed on the sequence analysis (for example, if peptides with C-terminal cysteines are filtered out). Nonetheless, it is remarkable that both GIPC and PTPN3 have multiple C₀-PBM partners, while this residue not frequently found in position 0. This opens the possibility that the C₀ in the HBc PBM is an evolutionary adaptation to specifically target a subset of PDZ-containing proteins to favour particular steps of its life cycle. On the next section of this thesis, I will show our results on the specificity of the HBc C₀-PBM towards the human PDZome. We report a number of PDZ-containing proteins that bind the HBc PBM with higher affinity and are potentially significant targets of HBV.

6.2.4 Binding specificity of PTPN3-PDZ for cellular proteins

We performed pull-down experiments to find new cellular PBM-containing partners of PTPN3-PDZ (as described in section 7.2). The data obtained provides information on the cellular partners of PTPN3-PDZ and on the consensus sequence of the PBMs that it binds. However, since the data come from cellular proteins and not from a library of random peptides, there might be a bias towards finding certain residues in certain positions, and so it would be incorrect to use this data to define a preferred motif of PTPN3-PDZ. We used purified GST-PTPN3-PDZ as bait to fish partners in HeLa S3 cell lysates, which were then identified by LC-MS/MS. 326 proteins bind exclusively GST-PTPN3-PDZ samples but are absent from GST controls. These are potential PTPN3 interactants through a PDZ/PBM interaction. We then narrowed our list of candidate proteins by selecting those that feature a PBM at their C terminus. This way, we shortlisted 128 potential PBM-containing proteins: 40 of class I (S/T-X- ϕ_{COOH}), 62 of class II (ϕ -X- ϕ_{COOH}), and 26 of classes III (H/R/K-X- ϕ_{COOH}) and IV (D/E-X- ϕ_{COOH}) (section 9.5).

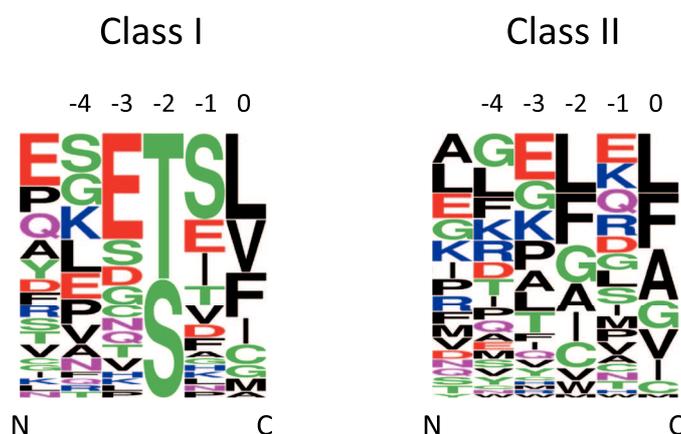


Figure 6.8: Frequency plot of the C-terminal residues of the class I and class II PTPN3-PDZ partners identified by pull-down. Amino acids are coloured according to their chemical properties: polar amino acids (G,S,T,Y,C,Q,N) are green, basic (K,R,H) blue, acidic (D,E) red and hydrophobic (A,V,L,I,P,W,F,M) amino acids are black. The positions of the PBM are indicated above the sequences.

As PTPN3-PDZ is a class I PDZ domain, we focused on the 40 class I PBM-containing partners (table 6.2). Sequence conservation analysis shows the most frequently found residues in each position of the PBM (figure 6.8). We can see that the preferred residues of PTPN3-PDZ at P₀ are L, V, F, and I. Together, these account for 32 over the 40 binders. Then, we found 3 proteins with C₀-PBMs: the microtubule-stabilizing protein

esconsin (MAP7), the mitochondrial NADH-ubiquinone oxidoreductase 75 kDa subunit (NDUFS1), and the PCI domain-containing protein 2 (PCID2). In P₋₁, PTPN3-PDZ shows a slight tendency to bind proteins with S₋₁. In our crystal structures, the Q or E found in this position is bonded to N524 through a molecule of water (figure 6.2). A S₋₁ would be able to bond directly to N524 thanks to its shorter side chain, which could explain why this residue is favoured in this position. In P₋₃ we find a preference for E, although S, D and G are also allowed. E and D have side chains with similar chemical properties, so a D₋₃ should be able to form a H-bond with S538 in a similar way to E₋₃ in our structures. S and G cannot form this bond and thus do not contribute to the interaction, but are accepted because their side chains do not cause any steric clashes. Finally, the conservation in P₋₄ is very low. This could be explained because the frequency of occurrence of a certain residue in this position does not reflect the affinity of the binding, but are mostly affected by the frequency in which said residue is found in cellular PBMs, and their level of expression.

Table 6.2: PTPN3-PDZ Class I PBM-containing partners.

PDB ID	Name	PBM
P48960	CD97 antigen	SESGI
P12814	Alpha-actinin-1	GESDL
P11169	Solute carrier family 2, facilitated glucose transporter member 3	TTTNV
O96008	Mitochondrial import receptor subunit TOM40 homolog	GLTIG
Q8TBC3	SH3KBP1-binding protein 1	NETSF
Q96TA1	Niban-like protein 1	VQTEF
P81605	Dermeidin	LDSVL
P53985	Monocarboxylate transporter 1	EESPV
O00592	Podocalyxin	EDTHL
Q13151	Heterogeneous nuclear ribonucleoprotein A0	GGSSF
P62701	40S ribosomal protein S4, X isoform	KQSSG
P26599	Polypyrimidine tract-binding protein 1	SKSTI
P27816	Microtubule-associated protein 4	QETSI
P02545	Prelamin-A/C	NCSIM
P55060	Exportin-2	SVTLL
P23219	Prostaglandin G/H synthase 1	PSTEL

Q9HB71	Calcyclin-binding protein	GDTEF
Q9BSE5	Agmatinase, mitochondrial	KVTTV
O15427	Monocarboxylate transporter 4	PETSV
P35354	Prostaglandin G/H synthase 2	RSTEL
O95336	6-phosphogluconolactonase	KHSTL
P49189	4-trimethylaminobutyraldehyde dehydrogenase	VESAF
Q9Y6M5	Zinc transporter 1	PESSL
P20020	Plasma membrane calcium-transporting ATPase 1	LETSL
Q9Y6M7	Sodium bicarbonate cotransporter 3	AETSL
P52569	Cationic amino acid transporter 2	KTSEF
Q02543	60S ribosomal protein L18a	PNTFF
P62829	60S ribosomal protein L23	AGSIA
O43707	Alpha-actinin-4	GESDL
Q96T17	MAP7 domain-containing protein 2	LNTFC
Q8TAA9	Vang-like protein 1	SETSV
Q15758	Neutral amino acid transporter B(0)	KESVM
Q15046	Lysine--tRNA ligase	VGTSV
Q96BN8	Ubiquitin thioesterase otulin	EETSL
Q5JVF3	PCI domain-containing protein 2	LSTVC
P09543	2',3'-cyclic-nucleotide 3'-phosphodiesterase	SCTII
O14734	Acyl-coenzyme A thioesterase 8	SESKL
P28331	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	EPSIC
P23634	Plasma membrane calcium-transporting ATPase 4	LETSV
Q9Y2T3	Guanine deaminase	FSSSV

It is interesting that the number of class II partners is bigger than the one of class I. In these partners, L, F, and A, are most often found in P₀, and some C₀-PBMs are also present (2 out of 61 partners). No tendencies seem to emerge from P₋₁. The preferred residues at P₋₂ for the class II partners are L, F, and G. Although a G₋₂ should not cause any steric clashes if inserted into the peptide-binding cleft, L and F would be more difficult to accommodate, since they have longer and hydrophobic side chains. As with the pull-down methodology we are fishing with full-length proteins, we cannot exclude that the interactions occur through a different interaction motif, for example via internal PBMs. Additionally, it is possible that the

class II PBMs interact with PTPN3-PDZ through a non-canonical binding mode. For example, in a growing number of PDZ domain structures^{354–356}, the peptide ligands are inserted perpendicular to the PDZ domain, with only P₀ and sometimes P₋₁ interacting with the PDZ domain. Although it is possible that perpendicular binding is solely an artifact of the crystal packing, the observation of this binding mode in the NMR solution structure of the autoinhibited X11 α PDZ1 domain²¹⁴ suggests that this type of noncanonical binding could be more relevant than what is currently thought.

In sum, the data derived from the cellular partners of PTPN3-PDZ identified by pull down experiments confirm the observations derived from the crystal structures of PTPN3-PDZ and PBM peptides. First, we identified other C₀-PBMs that can be bound by PTPN3-PDZ. Although these partners might not be physiologically relevant, they show that the characteristics of the carboxylate-binding pocket of PTPN3-PDZ allow it to accommodate PBMs with a terminal cysteine. Then, by analysing the preferred residues in positions -1 and -3 we confirmed that N524 and S538 are determinant for the selectivity of the NT5 subfamily. In position -4 we did not find a strong conservation to support the relevance of D573. However, it is important to keep in mind that with the methodology we used, the result is strongly dependent on the level of expression of the cellular proteins. Thus, as this data does not reflect the affinities of the interactions, the low conservation in position -4 does not mean that *in vivo* the D573 does not affect partner selection by PTPN3-PDZ. Indeed, as shown for PTPN4, the interactions established at this position have a significant impact on the binding affinity.

6.3 Materials and methods

Crystallisation, data collection, and structure determination

The peptides used for co-crystallization was added in excess to form >95% of the complex with the protein. The PDZ domain-peptide complex for crystallization was generated by mixing PTPN3-PDZ and the peptide at a ratio of 1:2. Initial screening of crystallization conditions was carried out by the vapor diffusion method using a MosquitoTM nanoliter-dispensing system (TTP Labtech). Sitting drops were set up using 400 nL of a 1:1 mixture of each sample protein and %(v/v)/50%(v/v) oil as the cryoprotectant. crystallization solutions (672 different commercially available conditions) equilibrated against a 150 μ L reservoir in multiwell plates (Greiner Bio-One). The crystallization plates were stored at 4 °C in a RockImager1000[®] (Formulatrix) automated imaging system to monitor crystal growth. The

best crystals were obtained by mixing 200 nL of PTPN3-PDZ_{Next}-PBM-peptide complex solution in 20 mM HEPES pH 8, 150 mM NaCl, 0.5 mM TCEP mixed with 200 nL of reservoir solution, as follows. For the complex with PBM-16E6, the PDZ domain was at a concentration of 4.6 mg/mL, and the reservoir solution contained 20% w/v PEG 3350, 0.2 M KI at pH 7; for PBM-18E6, the PDZ domain was at a concentration of 5 mg/mL, and the reservoir solution contained 20% w/v PEG 3350, 0.2 M NaI at pH 7; for PBM-HBc, the PDZ domain was at a concentration of 4.9 mg/mL, and the reservoir solution contained 20% w/v PEG 3350, 0.2 M NaBr at pH 7; for PBM-TACE, the PDZ domain was at a concentration of 4.8 mg/mL, and the reservoir solution contained 20% w/v PEG 3350, 0.2 M NaSCN at pH 7. Crystals were flash-cooled in liquid nitrogen using Paratone-paraffin 50%(v/v)/50%V/V) oil as cryoprotectant.

X-ray diffraction data were collected at a wavelength of 0.979 Å on the beamline PROXIMA-1 at Synchrotron SOLEIL (St. Aubin, France). The data were processed with XDS and Xdsme, and other programs from the CCP4 suite. The structures were solved by molecular replacement with PHASER using the search atomic model of PTPN3-PDZ (PDB ID 6HKS). The locations of the bound peptides were determined from a Fo–Fc difference electron density maps. Models were rebuilt using COOT, and refinement was done with phenix.refine of the PHENIX suite. The overall assessment of model quality was performed using MolProbity. The crystal parameters, data collection statistics, and final refinement statistics are shown in Table X. All structural figures were generated with the PyMOL Molecular Graphics System, Version 1.7 (Schrödinger).

Sequence logos

The sequence and frequency logos were created using the online WebLogo service³⁵⁷ at <https://weblogo.berkeley.edu/>. The amino acids are coloured according to their chemical properties: polar amino acids (G,S,T,Y,C,Q,N) are green, basic (K,R,H) blue, acidic (D,E) red and hydrophobic (A,V,L,I,P,W,F,M) amino acids are black. The sequence conservation is shown as a frequency plot.

Pull down experiments

The pull down experiments to fish PTPN3-PDZ partners were performed as described in section 7.2.

7 Study of the interaction of PTPN3 with the hepatitis B virus core protein

7.1 Summary

The hepatitis B virus has been the object of extensive research, much of which has focused on its capsid protein HBc. This small and versatile protein has been shown to have functions that extend well beyond its structural role. Zlotnick and collaborators adequately referred to it ‘a pleiotropic keystone in the HBV lifecycle’¹. Most of the HBc functions are achieved thanks to its CTD tail, which contains nucleic acid-binding sites, transport signals to regulate trafficking, post-translational modification sites, and a C-terminal PBM that mediates interactions with cellular PDZ proteins^{2,3}. However, despite the well-established relevance of viral targeting of PDZ proteins, and in comparison to other viral PBM-containing proteins, little is known about the functions of the HBc PBM. Until now, only two cellular partners had been reported: PTPN3 and GIPC1. Previous studies showed that PTPN3 reduced the levels of HBV gene expression under transfection conditions, and this effect required the catalytic activity of PTPN3 but was independent of the PDZ/PBM interaction⁴.

Our working hypothesis was that, given the very high concentrations of viral proteins in infected cells, HBc could displace natural PDZ ligands of PTPN3 during infection, impairing their dephosphorylation and producing a benefit for the virus. Our previous data supports this hypothesis showing that the PBM of HBc bound the PDZ domain of PTPN3 with a comparable affinity to the one of PTPN3 cellular partners.

We aimed to gain insights about fundamental aspects of PTPN3 signalling, and to understand in particular the interaction of PTPN3 and HBc and its disruptive effects in the context of HBV infection.

To this end, we investigated the recognition specificity of the PDZ binding motif of HBc for human PDZ domains, and we identified new host proteins potentially targeted by HBc during infection. Then, we showed that PTPN3 is able to bind the HBc PBM in the context of the full-length protein forming capsids or dimers, confirming that HBc is a partner of PTPN3. We structurally characterized this interaction by combining the crystal structure to study the bonding pattern of PTPN3-PDZ complexed to the HBc PBM with NMR chemical shift mapping to highlight short- and long-range perturbations induced in PTPN3-PDZ upon

binding. Then, to begin to shed light on the role of PTPN3 in the viral life cycle, we performed PTPN3 overexpression in HBV-infected cells and we observed that multiple markers of the viral life cycle were affected indicating that PTPN3 may have several effects. Finally, we investigated the PDZ-mediated interactions of PTPN3 with cellular proteins to uncover potential disruptive effects produced during HBV infection.

7.2 Article: Molecular basis of the interaction of the hepatitis B virus core protein with the human phosphatase PTPN3

Molecular basis of the interaction of the hepatitis B virus core protein with the human tyrosine phosphatase PTPN3

Mariano Genera^{1,2}, Barbara Quioc-Salomon^{3,4,5}, Antonin Nourisson¹, Baptiste Colcombet-Cazenave^{1,2}, Ahmed Haouz⁶, Ariel Mechaly⁶, Mariette Matondo⁷, Magalie Duchateau⁷, Alexander König⁸, Marc P. Windisch⁸, Christine Neuveut^{3,4}, Nicolas Wolff¹ and Célia Caillet-Saguy^{1,*}

¹ Récepteurs-Canaux, Institut Pasteur, UMR 3571, CNRS, F-75724 Paris, France

² Sorbonne Université, Complexité du Vivant, F-75005 Paris, France

³ CNRS, UMR 3569, 75015 Paris, France

⁴ Département de Virologie, Institut Pasteur Paris, France.

⁵ Université Paris Diderot, Sorbonne Paris Cité, Paris, France

⁶ Institut Pasteur, C2RT-Plate-forme de Cristallographie, Institut Pasteur UMR 3528, CNRS, F-75724 Paris, France

⁷ Proteomics Platform, Mass Spectrometry for Biology Utechs (MSBio), USR 2000 CNRS, Institut Pasteur, F-75724 Paris, France

⁸ Applied Molecular Virology Laboratory, Institut Pasteur Korea, 696 Sampyung-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, South Korea.

* Correspondence should be addressed to Célia Caillet-Saguy (email: celia.caillet-saguy@pasteur.fr)

7.2.1 Highlights

- Hepatitis B virus core protein (HBc) interactions with cell proteins are poorly known
- 28 potential human HBc binders were identified using a human PDZ domain library
- PTPN3 binds HBc in capsid-like particles and has multiple effects on HBV infection
- Potential PBM-containing cellular partners of PTPN3 were identified
- Our study provides a new resource for the HBV-host interactions research

7.2.2 Abstract

Interactions between the hepatitis B virus core protein (HBc) and host cell proteins are poorly understood, although they may be essential for the propagation of the virus and its pathogenicity. HBc has a C-terminal PDZ-binding motif (PBM) that is responsible for interactions with host PDZ domain-containing proteins. In this work, potential cellular partners of HBc through PDZ-PBM interactions were identified by screening a human PDZ domain library. Most of the HBc-interacting partners detected herein are also targeted by others viruses through PBM-PDZ interactions. Among these partners, we focused on the human protein tyrosine phosphatase non-receptor type 3 (PTPN3) and its interaction with HBc. We solved the crystal structure of the PDZ domain of PTPN3 in complex with the PBM of HBc, revealing a network of interactions specific to class I PDZ domains despite the presence of a C-terminal cysteine in this atypical PBM. We further showed that PTPN3 binds the HBc protein within capsids. The effects of PTPN3 overexpression on HBV infection in HepG2 NTCP cells indicate a significant and potentially pleiotropic role of PTPN3 in HBV replication. Finally, by performing pull-down assays with PTPN3-PDZ as bait, we identified a pool of human PBM-containing proteins that might interact with PTPN3 in cells and that could be in competition by the HBc PBM during infection.

Keywords

Protein tyrosine phosphatase PTPN3; PDZ domain; Hepatitis B virus core protein; host-pathogen interactions; protein-protein interaction

Abbreviations

BI, binding intensity; GST, glutathione *S*-transferase; HBc, Hepatitis B virus core protein; HSQC, heteronuclear single quantum coherence; LC, liquid chromatography; MS, Mass Spectrometry; PBM, PDZ-binding motif; PDZ, PSD-95/Dlg/ZO-1; pNPP, para-nitrophenylphosphate; PTP, protein tyrosine phosphatase; PTPN3, protein tyrosine phosphatase non-receptor type 3; TCEP, Tris(2-carboxyethyl)phosphine; WT, wild type; root mean square deviation (rmsd).

7.2.3 Introduction

Hepatitis B virus (HBV) remains a major global health problem, with 257 million chronic HBV carriers and almost 1 million deaths estimated in 2015 (WHO). Despite available prophylactic vaccines, chronic hepatitis B is still incurable; existing treatments focus on stopping the spread of the virus to prevent the progression of the disease. The acute infection can be asymptomatic or symptomatic, in some cases, resulting in fulminant hepatitis. Chronic infection can lead to the development of cirrhosis and hepatocellular carcinoma (HCC).

HBV is a small, enveloped DNA virus. The infectious particle contains a 3.2kb partial double-stranded relaxed circular DNA (rcDNA) surrounded by the capsids composed by the core protein (HBc). The different steps of the viral life cycle are: the virus entry in the cell, the translocation of the nucleocapsid to the nuclear pore allowing the release of the rcDNA into the nucleoplasm and its 'repair' into covalently closed circular DNA (cccDNA), the transcription of the viral RNAs from the cccDNA, the export of the pgRNA to the cytoplasm and viral protein translation, the encapsidation of the pgRNA and its reverse transcription into rcDNA inside the capsids. Nucleocapsids are then enveloped at the endoplasmic reticulum and released, or alternatively they return to the nucleus to form additional cccDNA molecules. The HBc capsid protein has functions that surpass its structural role. Indeed, HBc has been implicated in most stages of the viral life cycle, including subcellular trafficking and release of the viral genome, pgRNA encapsidation, and reverse transcription [1]. Although HBV is relatively well studied, little is known about its HBc-mediated protein interactions with host proteins. To clarify the role of the virus structural protein in the pathogenesis of HBV-infected hepatocytes and to identify human proteins involved in important stages of the virus life cycle, the interactions with host-cell proteins have to be investigated.

HBc protein (183 residues) presents an N-terminal capsid assembly domain (residues 1-149) [2] and an arginine-rich RNA-binding C-terminal domain (CTD, residues 150-183), which is

dispensable for capsid assembly *in vitro* but is required for viral genome replication in cells [2]. The CTD also contains nuclear localization signals and cytoplasmic retention signals that regulate nuclear import and export of the HBV capsid [3], and multiple serine and threonine residues that are susceptible of being phosphorylated [4]. Finally, the C-terminal residues of the CTD form a non-canonical class I PDZ-binding motif (PBM) with a C-terminal cysteine (-ESQC_{COOH}); Class I PBMs are defined as -[X-S/T-X-Φ], where Φ is any hydrophobic amino acid and X is any amino acid). This PBM has been shown to interact with the PDZ (PSD-95, Dlg1, ZO-1) domains of protein tyrosine phosphatase non-receptor type 3 (PTPN3) [5] and GIPC1 [6].

PTPN3 possesses three domains: a N-terminal FERM (band 4.1, ezrin, radixin, moesin) domain involved in localization to the cytoskeleton–membrane interface or interaction with transmembrane proteins, a central PDZ domain that mediates interaction with other proteins, and a C-terminal protein tyrosine phosphatase (PTP) domain bearing the catalytic activity of PTPN3. PTPN3 is a signalling protein and its FERM and PDZ domains account for specificity in its functions. The FERM domain targets the protein to the plasma membrane providing spatial regulation, while the PDZ domain mediates interactions with specific partners and substrates, or anchors the phosphatase in multi-protein signalling complexes. A few ligands and substrates and known functions of PTPN3 have been reported [7–16]. However, its precise role in cell signalling has not yet been clearly established. We performed pull-down assays with the PDZ domain of PTPN3 (PTPN3-PDZ) to detect endogenous PBM-containing proteins that potentially interact with PTPN3 in cells and that could be displaced by the PBM-HBc during viral infection. Several of these potential PTPN3 partners have been implicated in HBV infection and/or in HCC highlighting potential signalling pathways disrupted by HBV and/or involved in HCC. We also report the effects of PTPN3 overexpression on HBV infection in HepG2 NTCP cells to further support the role of PTPN3 in the HBV life cycle. Moreover, we report a high-throughput analysis of the HBc-interacting partners through PDZ-PBM interactions. As expected, PTPN3 binds the PBM of HBc (PBM-HBc) via its PDZ domain, although the role of this interaction in HBV pathogenesis is unknown. It has been shown that PTPN3 has a suppressive role on HBV gene expression that is independent of its PBM-mediated interaction with HBc [17]. PTPN3 was also reported to be targeted by the high-risk human papillomavirus types 16 and 18 (HPV 16 and 18) through the PBM of the viral E6 oncoprotein [18]. Interestingly, HBV and HPV 16 and 18 are all oncoviruses and accumulating evidence suggests that PTPN3 plays a critical role in the

progression of various human cancers (breast, lung, colorectal cancer, intrahepatic cholangiocarcinoma, hepatocellular carcinoma) [7,19–23].

We investigated the molecular mechanism of interaction between PTPN3 and HBc. We previously reported the similar binding affinities of the PBMs from E6 and HBc for PTPN3-PDZ, and we solved the structure of PTPN3-PDZ in complex with the PBM of the HPV16 E6 protein (PBM-16E6) [24]. To gain structural insights into the recognition of HBc by PTPN3, here we solved the X-ray structure of PTPN3-PDZ in complex with the PBM-HBc, revealing a network of interactions specific to class I PDZ domains despite the atypical PBM of HBc containing a cysteine at its C-terminus. We showed that PTPN3 binds HBc within capsids *in vitro*, supporting a model in which the CTD comprising the PBM is at least transiently exposed at the surface of the capsids.

7.2.4 Results

7.2.4.1 HBV targets PDZ-containing proteins frequently targeted by other viruses

To investigate the recognition specificity of the PBM-HBc for human PDZ domains (the PDZome), we used the automated high-throughput holdup assay, which allows measuring binding intensities (BIs) for a large number of PDZ domain-PBM pairs [25]. We used an updated PDZome library that contains all the 266 known human PDZ domains [26]. The PDZome-binding profile (Fig.1a) represents the individual BIs of each PDZ domain for the PBM-HBc peptide (peptide sequence with the PBM underlined: RRRRSQSRESQC). The binding intensities are directly linked to the PBM/PDZ affinities. Using BI=0.2 as the minimum threshold for a significant PDZ-peptide interaction [25], 28 potential HBc binders containing PDZ domains were identified (Fig.1b, Table 1). This dataset represents about 10% of the human PDZome targeted by PBM-HBc with high-to-medium affinities. The best binders, GIPC1 and PTPN3, have already been reported as targets of the HBV core protein through PDZ/PBM interactions [6,17]. PBM-HBc also bound the PDZ domain of the non-receptor protein tyrosine phosphatase 4 (PTPN4), the closely related isozyme of PTPN3, and the PDZ domain of GIPC2, closely related to GIPC1. The interacting proteins include various PDZ proteins targeted by other oncoviruses. The membrane-associated guanylate kinase, WW and PDZ domain-containing proteins 1, 2 and 3 (MAGI1, MAGI2 and MAGI3), the disc-large homolog 1 and 4 (DLG1 and DLG4), the protein scribble homolog (SCRIB), the multi-PDZ domain scaffolding protein (MPDZ), the INADL protein (InaDI) and the Sorting Nexin

27 (SNX27) [27] are all targeted by HPV types 16 and 18 through their E6 proteins [28]. PTPN3 and GIPC1 have also been reported as E6 binding partners. Thus, the same large pool of human PDZ-containing proteins is targeted by the oncoviruses HBV and the high-risk HPVs. Additionally, DLG1, MAGI1 and 3, and SCRIB are also targeted by another oncovirus, the Human T Lymphotropic Virus Type I (HTLV-I), through the PBM of its Tax protein. Likewise, non-oncogenic viruses target some of the PDZ-containing proteins that bind the PBM-HBc. Indeed, Adenovirus binds DLG1, MAGI-1, MPDZ, and the tight junction protein ZO-2 (TJP2) through the PBM of the early 4 ORF1 protein [28]. Then, the Microtubule-associated serine/threonine-protein kinase 2 (MAST2), the discs-large homolog 2 (DLG2), and PTPN4 interact with the PBM of the protein G of Rabies virus, and the beta-1-syntrophin (SNTB1) binds the HTLV-I Tax PBM [29]. All these HBc binders identified through the holdup assay and already reported as targets of other viral PBMs are good candidates for validation in cells.

The 6 remaining binders have not been reported to be targeted by viruses. These are the discs-large homolog 3 (DLG3), the alpha-1 and beta-2-syntrophin (SNTA1 and SNTB2), the Signal-induced proliferation-associated 1-like protein 2 (SIPA1L2), the Syntaxin-binding protein 4 (STXB4) and the Synaptojanin-2-binding protein (SYNJ2BP). They might be new relevant binders of PBM-HBc, or they interact because of PDZ domain homology, as in DLG3 with the DLG family and SNTA1 and SNTB2 with the syntrophin family. Alternatively, they might emerge from the degenerate specificity of PBMs or the promiscuity of PDZ/PBM interactions. Thus, the holdup assay enabled us to identify new host proteins potentially targeted by HBc during infection in addition to PTPN3 and GIPC1. To follow, we further investigated the interaction between HBc and PTPN3.

7.2.4.2 The atypical PDZ binding motif of HBc interacts with PTPN3-PDZ through a network specific to class I PDZ domains

PDZ domains can be grouped into in three main specificity classes: class I recognizes a serine or threonine at position -2 (P_{-2}) within a C-terminal peptide sequence-motif defined as $-[X-S/T-X-\Phi_{COOH}]$, where Φ is a hydrophobic residue and X is any residue; class II recognizes any hydrophobic residue at P_{-2} within a motif defined as $-[X-\Phi-X-\Phi_{COOH}]$ [30,31]; and class III, which recognizes the motif $-[X-D/E-X-\Phi_{COOH}]$ [32]. The PBM-HBc is an atypical PBM because it features a cysteine at its C-terminus. All the sequence consensus classes commonly used to describe PBMs present a hydrophobic residue in the last C-terminal position.

Although cysteine is often considered to be hydrophobic its classification is ambiguous, as some consider it to be polar since it is often found close to or at the surface of proteins. However, as we previously reported, its affinity for PTPN3-PDZ is similar to the affinity of canonical class I PBMs such as the PBM-16E6, which presents a leucine at the C-terminus (HPV16 E6 C-terminal sequence ETQL) [24].

To decipher the molecular basis of recognition of the C-terminal sequence of HBc by PTPN3, we solved the crystal structure of the complex formed by PTPN3-PDZ and the PBM-HBc peptide ($\text{NH}_2\text{RRRRSQSRESQC}_{\text{COOH}}$) by molecular replacement at 1.86 Å resolution. The structure of the complex between PTPN3-PDZ and the class I PBM-16E6 (PDB ID: 6HKS) that we previously solved was used as a template (Fig. 2, Table 2). The structure factors and coordinates have been deposited in the Protein Data Bank under accession code 6T36. We compared the set of intermolecular bonds in both complexes to obtain structural insights on the binding mode of the unconventional HBc PBM.

As expected, the overall structures of PTPN3-PDZ in complex with PBM-HBc and PBM-16E6 are similar with a very low root mean square deviation (rmsd) of 0.18 Å for all the backbone atoms (Fig. 2a). PTPN3-PDZ adopts the typical PDZ fold comprising five β -strands and two α -helices. A clear electron density map was seen for the last eight C-terminal residues of PBM-HBc (underlined in the peptide sequence -RRRRSQSRESQC COOH). The last five residues are inserted into the binding groove (Fig. 2b). PBM-HBc binds PTPN3-PDZ in a conventional way, inserting into a hydrophobic cleft formed by the β 2-strand, the α 2-helix and the “GLGF” loop (Fig. 2a) and pairing as an anti-parallel extension of the β 2-strand, as observed for the PBM-16E6. Indeed, the backbone amide proton of the cysteine (C_0) and the serine (S_{-2}) and the backbone carbonyl oxygen of S_{-2} of the PBM-HBc are H-bonded to the b2 sheet of PTPN3-PDZ through the backbone carbonyl oxygen of F523, L525 and the backbone amide proton of L525 respectively (black dashed lines in Fig. 2c).

PTPN3-PDZ possesses the interaction network specific to class I PDZ domains that differ from other classes by the nature of the residue at P_{-2} . The key interactions of the PBM residues at positions 0 and -2 with PTPN3-PDZ are identical to those we reported for the complex formed between PTPN3-PDZ and the PBM-16E6, in agreement with the same range of affinity of PTPN3-PDZ for PBM-HBc and classical PBMs [24]. The carboxylate of the C-terminal cysteine (C_0) of PBM-HBc forms three H bonds with the amide protons of F521, G522 and F523 of the ‘GLGF’ motif on PTPN3-PDZ (Fig. 2c). The side-chain of the C_0 occupies the same position as the leucine side-chain in the complex with HPV16E6 PBM (Fig. 2b, 2c). The space between the isoleucine 579 side-chain and the C_0 permits to

accommodate the cysteine side-chain in this hydrophobic environment without additional polar contacts to maintain its position. In position -2, the hydroxyl group of the serine (S₂) forms a H-bond with the Ne2 of the conserved H572 from the α 2-helix, as those we observed for T₂ in the PTPN3-PDZ/PBM-16E6 complex.

Interestingly, the same residues are found in positions -1, -3 and -4 (Q, E and R, respectively) in both viral PBMs, PBM-16E6 and PBM-HBc. Accordingly, we observed the same network of polar contacts in both complexes (Fig. 2c). In position -1, the ϵ NH₂ of the exposed glutamine (Q₁) side-chain is H-bonded to the side-chain carbonyl and the terminal amine of N524 of the b2-strand through a water molecule. The hydroxyl of S538 and the amine proton of N524 form H-bonds with the side-chain carboxylate of the glutamate in position -3 (E₃), and the carboxylate oxygens of D573 at the beginning of the α 2-helix form ionic bonds with the guanidinium nitrogens of arginine at position -4 (R₄) (Fig. 2c). These bonds involving residues at positions -3 and -4 are known to modulate the affinity of PDZ domain for PBM sequence while they do not strictly form part of the PBM [24,33–35]. A few water molecules conserved in the two complexes participate as well in the PBM binding network, such as the one making a bridge between R₄ and G527 and Q531 of the b2-b3 loop. Moreover, the backbone carboxyl oxygen of R₄ forms H-bonds with the side-chain amine of K526, which is in turn H-bonded to the hydroxyl of the serine of the PBM-HBc at position -5 (S₅) (Fig. 2c). Finally, the backbone amide of R₄ forms a H-bond with the side-chain carboxyl of Q531, whose amine group is also bonded to the backbone carboxyl of S₈, whereas a water molecule makes the bridge with the threonine at position -7 (T₇) within the complex with HPV16E6 PBM.

Thus, we determined that a C-terminal cysteine does not affect the PBM binding, considering the similar binding affinities of the PBM-HBc and the PBM-16E6 for PTPN3-PDZ, and the similar bonding patterns typical of class I PDZ complexes.

7.2.4.3 The PDZ-binding motif of HBc induces direct perturbations on residues at the PBM binding site of PTPN3-PDZ and in its vicinity

To provide insights at the atomic level on the residues of PTPN3 affected upon interaction with HBc, and on potential distal effects, we probed the binding of PTPN3-PDZ to the PBM-HBc by solution NMR experiments. ¹H, ¹⁵N HSQC spectra were recorded on ¹⁵N-labeled PTPN3-PDZ in its free form and in the presence of the unlabeled PBM-HBc peptide. We assigned the signals using the previous assignment of *the* backbone of the PTPN3-PDZ

complexed with the HPV16 E6 PBM (BMRB entry 27645) and we analyzed the chemical shift perturbations (Dd) of the ^1H , ^{15}N signals of PTPN3-PDZ induced by the binding to PBM-HBc. Two types of signals were used for the analysis (Fig. 3a): (1) signals experiencing significant chemical shift changes ($\text{Dd} > 0.19$ ppm; standard deviation 0.14 ppm); (2) signals experiencing large chemical shift changes or severe line broadening effects caused by exchange and disappear upon complex formation. Six residues belonging to signals of type 1, G522, L525, G527, G528, I559, E589, are shown in red in Figures 3a and 3b, while 9 residues of signals of type 2, D516, F521, F523, K526, V536, V575, K580, R583, S587, are colored in magenta in Figure 3a and 3b. Seven residues, D505, D518, I550, K552, H570, T571, R588, whose behaviour could not be safely defined mainly because they fall in crowded spectral regions, are colored in gray (Fig. 3b).

As expected, signals of the residues located in the PBM binding site such as L525, K526 and G527 of the β 2-strand; K580, R583 and V575 of the helix α 2; or F521, G522, F523 of the “GLGF motif” are among the most affected when the complex with PBM-HBc is formed (Fig. 3b). G528 in the β 2- β 3 loop, V536 on the β 3-strand, and the S587 and E589 in the α 2- β 5 loop are in close proximity to the PBM binding site and are also affected. In addition, L525, in direct contact with the ligand, is located 3.9 Å away from I559 on the β 4-strand, likely transmitting the perturbation through hydrophobic contacts between the two. D516 also experiences a strong effect upon PBM binding due to a network of water molecules that links it to the carboxylate of C₀. Thus, direct and distal perturbations are induced upon PBM binding, as we previously observed in PTPN3 and PTPN4 PDZ domains [24,36].

7.2.4.4 PTPN3-PDZ binds to full-length HBc assembled into CLPs

We then tested the complex formation between PTPN3-PDZ and the full-length HBc (Fig. 3c). ^1H , ^{15}N HSQC-TROSY spectra were recorded on ^{15}N -labeled PTPN3-PDZ in its free form (red signals in Fig. 3c) and in the presence of the unlabeled HBc full-length protein at different ratios. After stepwise additions of HBc to a final ratio 1:19.4 (black signals in Fig. 3c), nearly all of the TROSY resonances observed with free PTPN3-PDZ had disappeared, almost certainly because transverse relaxation in the complexes was too fast to observe NMR signals even with the TROSY effect. Indeed, WT HBc proteins form homodimers and self-assemble into capsids [37] and the full-length HBc expressed in *Escherichia coli* mainly assembles into capsid-like particles (CLPs) of 3 to 5 million Da, as measured by analytical

ultracentrifugation (AUC)(Fig. 3d). Indeed, we observed by AUC a large peak corresponding to a main species in our sample with a sedimentation coefficient of 85 S, in agreement with the values previously reported for HBV CLPs at 82.5 S [38]. The molecular tumbling of PTPN3-PDZ is drastically slowed down when forming high molecular weight complexes with the CLPs, which induced the broadening of the linewidths of the PTPN3-PDZ signals beyond detection. Yet, the TROSY spectra of the complexes did contain a few peaks, which we attributed to regions of PTPN3-PDZ that remained mobile, probably the N-terminal and C-terminal extremities. These results indicate that PTPN3-PDZ can bind the CTD of full-length HBc within CLPs *in vitro*.

7.2.4.5 PTPN3 binds to the HBc protein within capsids as well as to the assembly-deficient mutant Y132A

To investigate if the PBM at the end of CTD of HBc is accessible for PTPN3 binding when capsids are assembled in cells, a GST pull-down *in vitro* assay was performed with lysates of HeLa S3 cells overexpressing either the HA-tagged wild-type (WT) full-length HBc or its Y132A mutant that forms dimers but is unable to assemble into capsids [39] (Fig. 4a). GST protein was used as a negative control. The expression of HA-HBc WT and HA-HBc Y132A in the cell lysates was verified by Western-blot using the anti-HA antibody (Fig. 4a lane 5). We also confirmed, using native western blot, the presence of capsid in the HeLaS3 HBc WT cell lysate and its absence in the HeLaS3 HBc Y132A cell lysate (Fig. 4b).

The GST alone and its fusion with PTPN3-PDZ (GST-PTPN3-PDZ), used as baits, were immobilized on glutathione beads and tested for their ability to pull down HA-HBc WT by Western blot using the anti-HA antibody. After washing, identical amounts of beads were analysed for the presence of the tagged proteins, and HBc WT was detected only in interaction with GST-PTPN3-PDZ (Fig. 4a upper part, lane 2). Thus, the interaction between HBc and GST-PTPN3-PDZ is specific of PTPN3-PDZ.

Similarly, GST-PTPN3-PDZ, but not GST alone, was able to bind to HA-tagged HBc Y132A mutant protein as efficiently as WT HBc (Fig. 4a lower part, lane 2). These results demonstrate that PTPN3-PDZ interacts directly with the HBc protein both in the context of the full-length protein within capsids and in the dimeric form, and suggest that the interaction with PTPN3-PDZ could occur at different steps of viral infection.

7.2.4.6 PTPN3 overexpression has multiple effects on HBV infection in hepatoma cells

Human hepatoma HepG2 NTCP cells stably expressing full-length PTPN3 were established to monitor the effect of PTPN3 on HBV infection. Overexpression of PTPN3 was assessed by the quantification of PTPN3 RNA in HepG2 NTCP-PTPN3 compared to HepG2 NTCP and by western blot analysis (Fig. 5a). Neither cell death nor morphological differences were observed in HepG2 NTCP-PTPN3 cells compared to HepG2 NTCP cells. This indicates that PTPN3 overexpression is well tolerated by HepG2 NTCP cells.

Because HBc may be involved in different steps of the HBV life cycle, from the first steps of infection to viral RNA transcription, encapsidation and reverse transcription, we studied the impact of PTPN3 on HBV replication by following multiple virological parameters after HepG2 NTCP-PTPN3 and HepG2 NTCP infection. We measured the level of cccDNA using qPCR, HBV transcription using RTqPCR, and calculated the ratio RNA/cccDNA to evaluate cccDNA transcriptional activity. We also assessed HBV replication by measuring cytoplasmic RC-DNA by qPCR, and finally we monitored HBV secretion by calculating the ratio secreted RC-DNA/cytoplasmic RC-DNA.

We observed a 2-fold decrease of cccDNA levels in HepG2 NTCP-PTPN3 compared to HepG2 NTCP (Fig. 5b). This might be due to a default of DNA import through the capsids on the nucleus or an effect of PTPN3 on capsid stability. In parallel, a 2-fold increase in viral genome transcription and a 2.5-fold increase of secreted particles were also observed in PTPN3 overexpression conditions (Fig 5b). These results indicate a significant role of PTPN3 in HBV replication.

7.2.4.7 PTPN3 targets multiple PBM-containing binding partners in cells

We aimed at identifying new endogenous partner proteins of PTPN3 through a PDZ-PBM interaction to highlight potential biological pathways involving the phosphatase. We performed pull-down analyses using PTPN3-PDZ as a bait and HeLa S3 cell lysate. Purified GST-PTPN3-PDZ and GST alone were incubated with the soluble fraction of HeLa S3 cell lysates. The interaction partners recovered in each sample were then identified by LC-MS/MS. In total, 348 and 54 different proteins, common between triplicates, were identified with GST-PTPN3-PDZ and GST alone, respectively. 326 proteins were specifically detected in GST-PTPN3-PDZ samples and were absent from GST controls (Supplementary Table S1). These proteins are potential PTPN3 interactants through a PDZ/PBM interaction. Then, we

narrowed our list of candidate proteins by selecting those that feature a PBM at their C-terminus. We shortlisted 104 potential PBM-containing proteins: 40 of class I (S/T-X- ϕ COOH), 62 of class II (ϕ -X- ϕ COOH), and 2 of class III (D/E-X- ϕ COOH) with ϕ : V, I, L, A, G, W, C, M, F (Supplementary Table S2). PTPN3-PDZ is known to bind preferentially type I PBMs, so we focused on the 40 class I PBM-containing partners. Chronic infection with HBV can lead to the the development of hepatocellular carcinoma (HCC) [40]. Therefore, we searched for potential links between the 40 class I PBM-containing partners and HBV and/or HCC. Only Vangl1 (planar cell polarity protein 1), component of the Wnt/PCP pathway, was known to bind PTPN3, although it was not determined if the interaction occurred via its PBM [41]. Vangl1 is dysregulated in human cancers [42]. Interestingly, Vangl1 has been reported to be upregulated in the HBV-related cancer HCC, and associated to the invasion capacity of HCC cells [43]. Several other PTPN3-PDZ partners identified in our analysis have been linked to HCC. The nuclear transport receptor Exportin 2 is overexpressed in HCC and its upregulation correlates with de-differentiation, proliferation, and poor prognosis [44]. The PTPN3-PDZ domain also binds Podocalyxin, whose expression is increased in HCC, and participates in migration and invasion processes [45]. Acot8 is often upregulated in HCC patients [46], whereas its silencing reduces *in vitro* tumorigenesis in HCC cells.

Moreover, among the PTPN3 partners identified, some are directly linked to HBV infection. COX-2 is upregulated by HBV [47] and its expression is also increased in the liver of cancer patients with chronic HBV infection. COX-2 might represent an important cellular effector of the HBV protein HBx, which contributes to hepatitis B virus-associated hepatocarcinogenesis [48]. Interestingly, hnRNP A0, another PTPN3 partner, is implicated in the LPS-induced post-transcriptional regulation of COX-2 mRNA and specifically bind COX-2 mRNA [49]. The PTPN3 binder 2',3'-Cyclic Nucleotide 3'-Phosphodiesterase (CNP) is a gene stimulated by interferon. It also strongly inhibits HBV production by blocking viral proteins synthesis and reducing viral RNAs, respectively [50]. In patients with chronic hepatitis B, CNP is expressed in most hepatocytes of HBV-infected liver specimens. Inactivation of CNP expression moderately enhances viral production in HepG2.2.15 cells treated with IFN- α . Thus, CNP could be an intermediary of the response induced by interferon against HBV. PTB, the polypyrimidine tract-binding protein, is another interesting PBM-containing partner of PTPN3. Indeed, PTB can bind to the HBV PRE (posttranscriptional regulatory element), an RNA element that facilitates the export of unspliced mRNAs of the nucleus. Thus, PTB could be important for the PRE activity and may function as an export factor for mRNAs containing PRE [51].

In sum, we have identified potential PTPN3 PBM-containing cellular partners. Some of them have been reported in HCC and/or in HBV infection and could be displaced by HBc during HBV infection by competition with PBM-HBc.

7.2.5 Discussion

Thus far, only a few interactions of HBc with host proteins have been described. There are the ones mediated by the NLS present in the CTD of HBc, which can bind to importin $\alpha\beta$ complexes and target the capsid to the nuclear pores [3]. Several host kinases have also been proposed as candidates for CTD phosphorylation [52] and two human PDZ-containing proteins GIPC1 and PTPN3 [5,6] are known to bind the PBM of HBc. GIPC1 couples other proteins to myosin VI movement and participates in the recycling of membrane receptors [53], and has been proposed to be involved in short intracellular transport of HBc proteins or capsids from the cytoplasm to the nucleus [54]. Its PDZ domain is targeted by at least two other viral proteins: Tax protein of HTLV-1 [55] and E6 protein of HPV18 [56]. Concerning PTPN3, the precise role of the interaction between HBc and PTPN3 in HBV pathogenesis is still unknown.

In this study, we found that the effects of PTPN3 overexpression on HBV infection in hepatoma cells are multiple, suggesting that PTPN3 could have a significant and pleiotropic role in the HBV replication. In particular, we observed a decrease of cccDNA levels during HBV infection in HepG2 NTCP-PTPN3 compared to HepG2 NTCP. This might be due to a default of DNA import on the nucleus or an effect of PTPN3 on capsid stability or phosphorylation. We showed that PTPN3 binds the HBc CTD within CLPs and capsids *in vitro*. We hypothesized that PTPN3 might interact *in vivo* with the HBc CTD of HBV capsids through a PDZ/PBM interaction.

Indeed, several data indicate that at least a fraction of the HBc CTD are exposed on the external surface of the particle [57–59]. The CTD of cytosolic HBc is responsible for binding and encapsidating viral pgRNA through its CTD. It is generally accepted that the arginine-rich C-terminal region of the HBc CTD interacts with the viral genome inside the particle. Once the viral genome is matured to rcDNA within the capsids, a fraction of the rcDNA-filled capsids is shuttled back to the nucleus to maintain the levels of cccDNA. At this step, the nuclear localization signals (NLSs) of CTD are required. Therefore, to perform these functions, the CTD should be able to shuttle between the interior and the exterior of the

capsids. Previous observations support this model [60,61]. In addition, the linker at the boundary between the assembly and arginine-rich domains of HBc seems to be mobile and may allow a large mobility of the C-terminal domain [62].

Interestingly, PTPN3 contains 3 functional NLSs (1 in the FERM domain, 1 in the linker between the FERM and PDZ domains and 1 in a loop of the PTP domain), and PTPN3 lacking its FERM domain is found in the nucleus [17]. Three isoforms of PTPN3 are produced by alternative splicing, two of which have truncated FERM domains. PTPN3 might play a role in the DNA import into the nucleus through the PBM-HBc/PTPN3-PDZ interaction and the NLSs. Another attractive possibility is a modulation of capsid stability involving the tyrosine dephosphorylation activity of PTPN3. The PBM-HBc/PTPN3-PDZ interaction would allow PTPN3 to dephosphorylate the tyrosine of HBc. Several tyrosines in HBc are susceptible of being phosphorylated, including the tyrosine 132 (Y132), which is exposed in the HBc capsid and involved in capsid assembly [39]. To our best knowledge, there is currently no data reported in the literature on tyrosine phosphorylation in HBc. We can hypothesize that capsid assembly is regulated by phosphorylation of Y132, and that PTPN3 could dephosphorylate HBc and modulate the stability of the capsid. Thus, PBM-HBc might be responsible for PTPN3 recruitment to dephosphorylate HBc and regulate the capsid assembly.

In this study, we reported a pool of human PDZ-containing proteins potentially targeted by PBM-HBc. Despite the non-canonical composition of the PBM sequence (i.e. a cysteine at the last C-terminal position), a large panel of interactors with high-to-medium affinities for PBM-HBc was detected *in vitro*. As expected, we identified GIPC1 and PTPN3, but also new cellular partners potentially involved in the HBV life cycle. Notably, a large majority of these binders are targeted by other viruses. For example, the three proteins DLG1, MAGI-1 and MPDZ are all targeted by the oncoviruses HBV, HPV 16 and 18, and HTLV-1, as well as by adenovirus. Thus, viruses with quite distinct replication cycles encode PBM-containing proteins that target a common subset of cellular PDZ proteins during infection [29]. A majority of these PDZ-containing proteins are involved in cell-cell junction or polarity and cellular trafficking, suggesting that interfering with these functions is a strategy shared by viruses. The interaction of viral PBMs with PDZ-containing proteins frequently leads to a loss of function either by abnormal sequestration in cell structures or by proteasomal degradation, while in some cases an apparent gain of function of the cellular protein is observed. In all cases, the targeting of PDZ-containing proteins and the alteration of cellular processes

regulated by these proteins are likely to improve viral replication, dissemination in the host, or transmission to new hosts.

As the structural data on PDZ domains in complex with a C-terminal cysteine PBM are rather sparse, we have solved the high-resolution structure of the PTPN3-PDZ/PBM-HBc complex. The atypical PBM-HBc binds PTPN3-PDZ establishing a network of interactions typical of class I PDZ domains despite the cysteine in the last C-terminal position, which explains the binding affinity for PTPN3-PDZ being comparable to the one of canonical PBMs. Interestingly, PTPN3 is also able to bind to the cytoplasmic domain of TACE (tumor necrosis factor alpha-convertase), which also presents a C-terminal cysteine in its PBM [13], and GIPC1, the other reported PBM-HBc target, interacts as well with the human lutropin receptor, which also presents a C-terminal cysteine in its PBM (-YTEC_{COOH}) sequence. Structural insights about the binding of a cysteine at the C-terminal part of a PBM to a PDZ domain have been reported for a class II PBM (liprin-a peptide) with the GRIP1 PDZ6 [63]. We compared the binding mode of this complex with our crystal structure of the complex between the class I PBM of HBc and PTPN3-PDZ, and we found a similar orientation of the side-chains of the two cysteines in the hydrophobic pocket. The position of these side-chains is equivalent to the one of the classical C-terminal hydrophobic leucine in the PBM-16E6 bound to PTPN3-PDZ. Thus, it appears that a C-terminal cysteine in PBMs is not discriminant, since PDZ domains such as PTPN3-PDZ can interact with these PBMs and with canonical ones displaying a similar network of interaction and affinity.

HBV, through the PBM-HBc, may enter in competition with endogenous partners to hijack signalling pathways in infected cells as already reported for the rabies glycoprotein [34]. We propose a list of PBM-containing candidates that are possible partners of PTPN3 through PDZ/PBM interactions using a pull-down assay. Among the candidates, CNP, for example, appears to be involved in cellular response against HBV infection, as it has been found to block viral protein synthesis and reduce viral RNAs. PTPN3 might participate in a signalling pathway of interferon-stimulated genes. HBc could block this process by targeting the PDZ domain of PTPN3 partner and displacing CNP. Similarly, HBc could displace PTB, another PBM partner of PTPN3. PTB binds the regulatory protein PRE and both proteins appear to enhance the nuclear export of unspliced RNAs (such as the HBV RNAs). By targeting and disrupting the PTPN3-PTB interaction, HBV could potentially favour the export of its RNAs, and thus enhance virus production.

In conclusion, during infection the virus may target endogenous proteins for a direct and specific action as PTPN3 on encapsidation, but also for a more general perturbation of

cellular homeostasis in its favor. Further research on the specific set of PDZ-containing proteins targeted by the PBM of HBc and the PBM-containing proteins interacting with PDZ-PTPN3 will provide insights into the mechanisms whereby the interaction of HBc with these cellular proteins is advantageous for the HBV life cycle. Our integrative study identifies perturbations at the levels of proteins and cellular pathways and provides a new resource for the HBV-host interactions research.

7.2.6 Materials and Methods

Holdup assay

The holdup assay was carried out against the biotinylated peptide PBM-HBc (peptide sequence -RRRRS QSRESQC) in duplicates as previously described [25,26] with minor modifications. We measured PBM-HBc interactions against 255 human PDZ domains. The minimal BI threshold value is 0.2 to define a significant interaction as previously reported [25].

Production and purification of recombinant proteins and PBM peptides

PTPN3-PDZ is encoded as an N-terminal glutathione S-transferase (GST) tagged protein. Construct delimitations of PTPN3-PDZ are from residues 489 to 597 except for NMR titration with PBM-HBc peptide and Holdup assays where the construct originated from another plasmid and its delimitation was from residues 504 to 597. Uniformly ¹⁵N-labeled and unlabeled PTPN3-PDZ constructs were expressed and purified as previously described [24]. For pull-down assays with GST-PTPN3-PDZ, the cleavage step by TEV protease in the protocol was skipped. The samples for crystallogenesis of PTPN3-PDZ were prepared as previously described [24].

The HBc expression vector pRSF-T7-HBc183opt plasmid previously described [64] carrying the HBc gene, optimized for bacterial expression, was used to express the full-length HBc. Transformed *E. coli* BL21 CodonPlus (DE3) cells were grown in LB medium with 100 µg/mL of kanamycin and 25 µg/mL of chloramphenicol. Protein expression was induced at $D_{600\text{ nm}} 2$ with 1 mM isopropyl thio-β-d-galactoside at 18 °C for 16 h. Harvested cells were resuspended in buffer B (50 mM Tris/HCl, pH 7.5, 300 mM NaCl, 5mM DTT), 1 mg/mL lysozyme, 50 U/mL benzonase and protease inhibitor cocktail (Roche), and then incubated 45 min with 0.5% Triton X-100. The lysis is achieved by sonication. The cell lysate was cleared

by centrifugation at 20,000 *g*. HBV core particles were further purified by 10-60% sucrose gradient (prepared with buffer B) ultracentrifugation at 140,000 *g* for 3 hours (Optima L-80, Beckman ultracentrifuge). The eluted fractions containing the protein were pooled and ammonium sulfate was then added to 40% saturation [65]. The mixture was left for 1 hour to salt out the HBV core and then centrifuged at 20,000 *g* for 15 minutes to pellet the HBV core. Minimal volume of buffer B was added to resuspend the pellet and then the mixture was dialyzed against buffer B overnight. The purified HBV core sample was concentrated by ultra-filtration with a MW cut-off of 100 kD (Millipore, USA). Protein concentration was estimated by the Bradford method.

The peptides PBM-HBc and biotinylated PBM-HBc were synthesized in solid phase using Fmoc strategy (Proteogenix) and resuspended in H₂O.

Crystallisation, data collection, and structure determination

The PBM-HBc peptide used for co-crystallization was added in excess to form >95% of the complex with the protein. The PDZ domain-peptide complex for crystallization was generated as previously reported [24]. The best crystals were obtained by mixing 200 nL of PTPN3-PDZ·PBM-HBc complex solution (concentration of the PDZ domain at 4.5 mg/mL) in 20 mM HEPES pH 8, 150 mM NaCl, 0.5 mM TCEP mixed with 200 nL of reservoir solution containing 20% w/v PEG 3350, 0.2 M NaBr at pH 7. Crystals were then flash-cooled in liquid nitrogen using Paratone-paraffin 50%(v/v)/50%(v/v) oil as the cryoprotectant.

X-ray diffraction data were collected at a wavelength of 0.979 Å on the beamline PROXIMA-1 at Synchrotron SOLEIL (St. Aubin, France). The data were reprocessed and the structures were solved as previously reported [24] using the search atomic model of PTPN3-PDZ (PDB ID: 6HKS). The positions of the bound peptides were determined from a $F_o - F_c$ difference electron density maps. Models, refinement and the overall assessment of model quality were done as previously described [24]. The crystal parameters, data collection statistics, and final refinement statistics are shown in Table 2. All structural figures were generated with the PyMOL Molecular Graphics System, Version 1.7 (Schrödinger).

Analytic ultracentrifugation (AUC) experiments

HBc CLPs formation was verified by AUC. Sedimentation velocity experiments were carried out at 20 °C using an analytical ultracentrifuge (Beckman Coulter Optima AUC) equipped with a AN50-Ti rotor. The HBc protein sample at approximately 150 mg/ml was

centrifuged for 17 h at 12000 rpm. Interference profiles were analyzed with SEDFIT 16.1 [66] using a continuous size distribution $c(S)$ model with a constant diffusion coefficient D .

NMR experiments

The NMR samples of the PTPN3-PDZ constructs were prepared in buffer A with 0.5 mM TCEP and D₂O (5–10% v/v). The NMR titration experiments to map the PTPN3-PDZ·PBM interaction were performed on a 600-MHz Varian NMR System spectrometer equipped with a triple resonance $^1\text{H}\{^{13}\text{C}/^{15}\text{N}\}$ cryoprobe at 15 °C as previously reported [24] with a stock solution of the unlabeled peptide PBM-HBc of 3.6 mM at pH 7.5 and a sample of ^{15}N -labeled PTPN3-PDZ initially containing 260 μL at a concentration of 95 μM . The chemical shift changes were followed with the CcpNmr Analysis software [67].

The NMR binding experiments between PTPN3-PDZ and full-length HBc as CLPs were performed at 20 °C on a 600-MHz Bruker Avance III HD spectrometer equipped with a cryoprobe. Briefly, the unlabeled HBc (stock solutions at 7.1 mM) were added in a sample initially containing 270 μL of ^{15}N -labeled PTPN3-PDZ at a concentration of 95 μM . A series of ^1H , ^{15}N HSQC-TROSY spectra was recorded for different titration points with a ratio PDZ:HBc (mol:mol) 1:0, 1:1.4, 1:5.6 and 1:19.4.

Cells

HelaS3-HBc and HelaS3-HBcY132A were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% foetal calf serum (FCS). They are derived from HeLaS3 and express the HBc wild type protein or HBc Y132A mutant respectively, as a fusion with Flag and HA Tags. They have been established as previously described [68].

HepG2-NTCP cells (A3 clone) derive from HepG2 cells and express the human sodium taurocholate cotransporting polypeptide (NTCP). HepG2-NTCP cells are grown in DMEM with 10% fetal calf serum (FCS). HepG2-NTCP-PTPN3 cells are derived from HepG2-NTCP cells and stably express PTPN3. pCMV6-PTPN3 plasmid was from Origene (reference RC15851).

Pull-down assays

The construct PTPN3-PDZ tagged with a N-terminal GST (GST-PTPN3-PDZ) was expressed and purified as previously described [24] without TEV cleavage. The additional TEV cleavage step of GST-PTPN3-PDZ provides the GST alone. Pull-down assays were

performed as previously described [35] except that the lysate is from human cervical carcinoma *cell* line HeLa S3 cells overexpressing the HA-HBc.

Briefly, HeLa S3 cells were prepared as followed. Cells overexpressing the HA-HBc protein were lysed in 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 0.5% Nonidet P-40, complete protease inhibitor cocktail (Roche), 2 mM DTT by incubating cells on ice for 15 minutes followed by five sonication cycles and a $13,000 \times g$ spin for 20 min.

After lysate binding, the glutathione-agarose beads were washed four times with binding buffer, pelleted and resuspended in a SDS-PAGE sample buffer for Western blotting. After boiling, the bound proteins were analyzed using 12% SDS-PAGE followed by Western blotting using HRP-conjugated HA Epitope Tag monoclonal antibody (26183-HRP, ThermoFisher) diluted to 1:2000. The proteins were visualized by enhanced chemiluminescence according to the manufacturer's instructions (Immobilon Forte Western HRP substrate, Millipore).

For pull-down analyses by LC-MS/MS, the proteins bound to the glutathione-agarose resin were eluted by incubation with binding buffer containing 10 mM glutathione for 1h at 4°C. The beads were separated by centrifugation, and the proteins in the supernatant were precipitated using trichloroacetic acid/acetone [69].

Samples were re-suspended in 300 μ L 8 M urea/ 100 mM Tris HCl pH 8.5. Briefly, samples were reduced with 5 mM TCEP for 30 minutes at room temperature and alkylated with 10 mM iodoacetamide for 30 minutes at room temperature in the dark. Then, proteins were digested for 5 hours at 30°C with 500 ng rLys-C Mass Spec Grade (Promega, Madison, WI, USA). Samples were then treated as previously described [70] as well as tryptic peptides analysis [70] with minor changes.

All data were searched using Andromeda [71] with MaxQuant software [72,73] version 1.5.3.8 against Homo sapiens reference proteome from Uniprot (20 399 entries) concatenated with GST-PTPN3-PDZ protein sequence, usual known mass spectrometry contaminants and reversed sequences of all entries as previously described [70]. The “match between runs” feature was applied between replicates with a maximal retention time window of 0.7 minutes. One unique peptide to the protein group was required for the protein identification. A false discovery rate (FDR) cut-off of 1 % was applied at the peptide and protein levels. Reverse proteins and usual MS contaminants were removed before the analysis of the data. Quantification of each identified protein was performed by summing the intensities of its associated peptides.

HBV production and infection

For virus production, HepAD38 cells were grown in Williams E medium supplemented with 5% FCS, 7.10^{-5} M hydrocortisone hemisuccinate, 5 mg/mL insulin, and 2% dimethylsulfoxide (DMSO). HBV particles were concentrated through ultracentrifugation with a SW-32 rotor at 32,000rpm and a 20% sucrose cushion. Titers of enveloped DNA-containing viral particles were determined by immunoprecipitation with an anti-preS1 antibody (gift of C. Sureau), followed by qPCR quantification of viral RC-DNA using RC primers RC 5' (5'-CACTCTATGGAAGGCGGGTA-3') and RC 3' (5'-TGCTCCAGCTCCTACCTTGT-3'). Enveloped DNA-containing viral particles (vp) quantification was used to normalize for virus infection, and multiplicities of infection (MOI) were expressed as vp per cell.

HepG2-NTCP or HepG2-NTCP-PTPN3 cells were infected with normalized amounts of virus at a MOI of 500 vp/cell in presence of 4% PEG 8000 and 3% DMSO. Infected cells were maintained in medium supplemented by 3% DMSO and collected 7 days after infection for virus replication analysis.

Quantitative RT-PCR (RT-qPCR)

Total RNA was prepared using TRIzol reagent (Invitrogen) and DNA contamination was removed by TURBO DNase treatment (Ambion). RNA (500 ng) was retrotranscribed using Oligo dT primers and RevertAid H Minus M-MuLV reverse transcriptase (Fermentas). RT-qPCR experiments were carried out as described [74]. Relative quantifications were performed as described previously [75]. The primers HBV RNAall-F and HBV RNAall-R amplify all HBV transcripts (pgRNA as well as the 2.4 and 2.1 kb mRNA) except the 0.8 Kb transcript encoding HBx.

Quantification of cccDNA

For nuclei isolation, cells were lysed in fractionation buffer containing 100 mM tris HCl pH 7 and 0,375% NP40 and shook using a vortex for 30 sec. Nuclei were pelleted by centrifugation (10 min at 15,000 rpm) and washed in fractionation buffer. Supernatant corresponding to cytoplasmic fraction were kept for RC-DNA quantification. Following centrifugation DNA was extracted using QIAamp DNA blood mini kit (Qiagen). For HBV cccDNA quantification, DNA was pre-treated with 10 U of plasmid-safe DNase (Epicenter) for 1 h at 37 °C and cccDNA was amplified using cccDNA primers: cccDNA 5' (5'-GTGCACTTCGCTTCACCTCT-3') and cccDNA 3' (5'-AGCTTGGAGGCTTGAACAGT-3'). Samples were normalized using CyclinA2 quantification by qPCR with the following

primers: CCNA2 5' (5'-CCTGCTCAGTTTCCTTTGGT-3') and CCNA2 3' (5'-AGACGCCCAGAGATGCAG-3').

Quantification of RC-DNA in the cytoplasm

Supernatant from nucleocytoplasmic fractionation described above is digested with DNase and RNase for 1 h at 37 °C in presence of 6.25 mM MgOAc to remove nucleic acid contaminants. Capsids were then digested with 7.5 µl Proteinase K (eurobio) and 12.5mM EDTA, 1% SDS, 125 mM NaCl. DNA was then purified using phenol chloroform technique followed by isopropanol precipitation and quantified by qPCR using RC primers: RC 5' (5'-CACTCTATGGAAGGCGGGTA-3') and RC 3' (5'-TGCTCCAGCTCCTACCTTGT-3').

Virus secretion

Supernatant from infected cells were recovered and centrifuge 1min at 13000g to remove cellular debris and PEG 8000 was added to a 5% final concentration. After precipitation at 4°C, supernatants were centrifuged 15min at 13000g. Pellets containing virions were resuspended in lysis buffer containing 1% SDS, 0,1M NaHCO₃, 0,1M TrisHCl pH 6,5 and 0,8mg/ml proteinase K and incubated 1h 56°C. DNA was then purified using phenol chloroform technique followed by isopropanol precipitation and quantified by qPCR using RC primers: RC 5' (5'-CACTCTATGGAAGGCGGGTA-3') and RC 3' (5'-TGCTCCAGCTCCTACCTTGT-3').

Accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 6T36.

Acknowledgements

Mariano Genera is part of the Pasteur-Paris University (PPU) International PhD Program. We thank the staff of the crystallography platform at Institut Pasteur for carrying out robot-driven crystallization screening. We thank Sébastien Brûlé and Bertrand Raynal from the Institut Pasteur molecular biophysics platform for the design and analysis of the AUC experiment. We thank Bruno Vitorge from the Institut Pasteur biological NMR technological platform for help with NMR experiments. We acknowledge synchrotron SOLEIL (Saint-Aubin, France) for granting access to their facility and the staff of Proxima1 for helpful assistance during the

data collection. We thank Lauriane Lecoq and Anja Böckmann for the pRSF-T7-HBc183opt plasmid and the advices to produce HBc.

Author Contributions Statement

MG and CCS planned the experiments. MG, BQS, AN, AH, MD, AK and CCS performed the experiments. MG, BQS, AN, BCC, AM, MM, MD, AK, MW, CN and CCS analysed the data. MG, NW and CCS wrote the paper with input from all authors.

Competing interests

The authors declare no competing financial and non-financial interests.

References

- [1] A. Diab, A. Foca, F. Zoulim, D. Durantel, O. Andrisani, The diverse functions of the hepatitis B core/capsid protein (HBc) in the viral life cycle: Implications for the development of HBc-targeting antivirals, *Antiviral Research*. 149 (2018) 211–220. <https://doi.org/10.1016/j.antiviral.2017.11.015>.
- [2] M. Nassal, The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly., *Journal of Virology*. 66 (1992) 4107–16.
- [3] M. Blondot, V. Bruss, M. Kann, Intracellular transport and egress of hepatitis B virus, *Journal of Hepatology*. 64 (2016) S49–S59. <https://doi.org/10.1016/j.jhep.2016.02.008>.
- [4] J. Jung, S.G. Hwang, Y.-J. Chwae, S. Park, H.-J. Shin, K. Kim, Phosphoacceptors threonine 162 and serines 170 and 178 within the carboxyl-terminal RRRS/T motif of the hepatitis B virus core protein make multiple contributions to hepatitis B virus replication., *Journal of Virology*. 88 (2014) 8754–67. <https://doi.org/10.1128/JVI.01343-14>.
- [5] E.-C. Hsu, Y.-C. Lin, C.-S. Hung, C.-J. Huang, M.-Y. Lee, S.-C. Yang, L.-P. Ting, Suppression of hepatitis B viral gene expression by protein-tyrosine phosphatase PTPN3., *Journal of Biomedical Science*. 14 (2007) 731–44. <https://doi.org/10.1007/s11373-007-9187-x>.
- [6] R. Razanskas, K. Sasnauskas, Interaction of hepatitis B virus core protein with human GIPC1, *Archives of Virology*. 155 (2010) 247–250. <https://doi.org/10.1007/s00705-009-0561-z>.
- [7] M.-Y. Li, P.-L. Lai, Y.-T. Chou, A.-P. Chi, Y.-Z. Mi, K.-H. Khoo, G.-D. Chang, C.-W. Wu, T.-C. Meng, G.-C. Chen, Protein tyrosine phosphatase PTPN3 inhibits lung cancer cell proliferation and migration by promoting EGFR endocytic degradation, *Oncogene*. 34 (2015) 3791–3803. <https://doi.org/10.1038/onc.2014.312>.
- [8] I. Pilecka, C. Patrignani, R. Pescini, M.-L. Curchod, D. Perrin, Y. Xue, J. Yasenchak, A. Clark, M.C. Magnone, P. Zaratini, D. Valenzuela, C. Rommel, R. Hooft van Huijsduijnen, Protein-tyrosine phosphatase H1 controls growth hormone receptor signaling and systemic growth, *J. Biol. Chem*. 282 (2007) 35405–35415. <https://doi.org/10.1074/jbc.M705814200>.
- [9] S. Han, S. Williams, T. Mustelin, Cytoskeletal protein tyrosine phosphatase PTPH1 reduces T cell antigen receptor signaling, *Eur. J. Immunol*. 30 (2000) 1318–1325. [https://doi.org/10.1002/\(SICI\)1521-4141\(200005\)30:5<1318::AID-IMMU1318>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1521-4141(200005)30:5<1318::AID-IMMU1318>3.0.CO;2-G).
- [10] M.S. Sozio, M.A. Mathis, J.A. Young, S. Wälchli, L.A. Pitcher, P.C. Wraage, B. Bartók, A. Campbell, J.D. Watts, R. Aebersold, R. Hooft van Huijsduijnen, N.S.C. van Oers, PTPH1 is a predominant protein-tyrosine phosphatase capable of interacting with and dephosphorylating the T cell receptor zeta subunit, *J. Biol. Chem*. 279 (2004) 7760–7769. <https://doi.org/10.1074/jbc.M309994200>.
- [11] T.J. Bauler, E.D. Hughes, Y. Arimura, T. Mustelin, T.L. Saunders, P.D. King, Normal TCR signal transduction in mice that lack catalytically active PTPN3 protein tyrosine phosphatase, *J. Immunol*. 178 (2007) 3680–3687. <https://doi.org/10.4049/jimmunol.178.6.3680>.
- [12] S.H. Zhang, J. Liu, R. Kobayashi, N.K. Tonks, Identification of the cell cycle regulator VCP (p97/CDC48) as a substrate of the band 4.1-related protein-tyrosine phosphatase PTPH1, *J. Biol. Chem*. 274 (1999) 17806–17812. <https://doi.org/10.1074/jbc.274.25.17806>.
- [13] Y. Zheng, J. Schlondorff, C.P. Blobel, Evidence for regulation of the tumor necrosis factor alpha-converter (TACE) by protein-tyrosine phosphatase PTPH1, *J. Biol. Chem*. 277

- (2002) 42463–42470. <https://doi.org/10.1074/jbc.M207459200>.
- [14] S.H. Zhang, R. Kobayashi, P.R. Graves, H. Piwnicka-Worms, N.K. Tonks, Serine phosphorylation-dependent association of the band 4.1-related protein-tyrosine phosphatase PTPH1 with 14-3-3beta protein, *J. Biol. Chem.* 272 (1997) 27281–27287. <https://doi.org/10.1074/jbc.272.43.27281>.
- [15] T. Jespersen, B. Gavillet, M.X. van Bemmelen, S. Cordonier, M.A. Thomas, O. Staub, H. Abriel, Cardiac sodium channel Na(v)1.5 interacts with and is regulated by the protein tyrosine phosphatase PTPH1, *Biochem. Biophys. Res. Commun.* 348 (2006) 1455–1462. <https://doi.org/10.1016/j.bbrc.2006.08.014>.
- [16] S.-W. Hou, H.-Y. Zhi, N. Pohl, M. Loesch, X.-M. Qi, R.-S. Li, Z. Basir, G. Chen, PTPH1 dephosphorylates and cooperates with p38gamma MAPK to increase ras oncogenesis through PDZ-mediated interaction, *Cancer Res.* 70 (2010) 2901–2910. <https://doi.org/10.1158/0008-5472.CAN-09-3229>.
- [17] E.-C. Hsu, Y.-C. Lin, C.-S. Hung, C.-J. Huang, M.-Y. Lee, S.-C. Yang, L.-P. Ting, Suppression of hepatitis B viral gene expression by protein-tyrosine phosphatase PTPN3, *J. Biomed. Sci.* 14 (2007) 731–744. <https://doi.org/10.1007/s11373-007-9187-x>.
- [18] M. Jing, J. Bohl, N. Brimer, M. Kinter, S.B. Vande Pol, Degradation of tyrosine phosphatase PTPN3 (PTPH1) by association with oncogenic human papillomavirus E6 proteins, *J. Virol.* 81 (2007) 2231–2239. <https://doi.org/10.1128/JVI.01979-06>.
- [19] H.-Y. Zhi, S.-W. Hou, R.-S. Li, Z. Basir, Q. Xiang, A. Szabo, G. Chen, PTPH1 cooperates with vitamin D receptor to stimulate breast cancer growth through their mutual stabilization, *Oncogene.* 30 (2011) 1706–1715. <https://doi.org/10.1038/onc.2010.543>.
- [20] Z. Wang, D. Shen, D.W. Parsons, A. Bardelli, J. Sager, S. Szabo, J. Ptak, N. Silliman, B.A. Peters, M.S. van der Heijden, G. Parmigiani, H. Yan, T.-L. Wang, G. Riggins, S.M. Powell, J.K.V. Willson, S. Markowitz, K.W. Kinzler, B. Vogelstein, V.E. Velculescu, Mutational analysis of the tyrosine phosphatome in colorectal cancers, *Science.* 304 (2004) 1164–1166. <https://doi.org/10.1126/science.1096096>.
- [21] Q. Gao, Y.-J. Zhao, X.-Y. Wang, W.-J. Guo, S. Gao, L. Wei, J.-Y. Shi, G.-M. Shi, Z.-C. Wang, Y.-N. Zhang, Y.-H. Shi, J. Ding, Z.-B. Ding, A.-W. Ke, Z. Dai, F.-Z. Wu, H. Wang, Z.-P. Qiu, Z.-A. Chen, Z.-F. Zhang, S.-J. Qiu, J. Zhou, X.-H. He, J. Fan, Activating mutations in PTPN3 promote cholangiocarcinoma cell proliferation and migration and are associated with tumor recurrence in patients, *Gastroenterology.* 146 (2014) 1397–1407. <https://doi.org/10.1053/j.gastro.2014.01.062>.
- [22] Z. Wang, Protein tyrosine phosphatase PTPN3 inhibits hepatocellular carcinoma growth and metastasis by dephosphorylation of EGFR | *OncologyPRO*, (n.d.). <http://oncologypro.esmo.org/Meeting-Resources/ESMO-Asia-2015-Congress/Protein-tyrosine-phosphatase-PTPN3-inhibits-hepatocellular-carcinoma-growth-and-metastasis-by-dephosphorylation-of-EGFR> (accessed March 11, 2018).
- [23] S. Ikuta, F. Itoh, Y. Hinoda, M. Toyota, Y. Makiguchi, K. Imai, A. Yachi, Expression of cytoskeletal-associated protein tyrosine phosphatase PTPH1 mRNA in human hepatocellular carcinoma, *J. Gastroenterol.* 29 (1994) 727–732.
- [24] M. Genera, D. Samson, B. Raynal, A. Haouz, B. Baron, C. Simenel, R. Guerois, N. Wolff, C. Caillet-Saguy, Structural and functional characterization of the PDZ domain of the human phosphatase PTPN3 and its interaction with the human papillomavirus E6 oncoprotein, *Scientific Reports.* 9 (2019) 7438. <https://doi.org/10.1038/s41598-019-43932-x>.
- [25] 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>.

- [26] 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 New Library of Human Single and Tandem PDZ Domains Allows Quantitative PDZ-Peptide Interaction Screening Through High-Throughput Holdup Assay, *Methods Mol. Biol.* 2025 (2019) 439–476. https://doi.org/10.1007/978-1-4939-9624-7_21.
- [27] K. Ganti, P. Massimi, J. Manzo-Merino, V. Tomaić, D. Pim, M.P. Playford, M. Lizano, S. Roberts, C. Kranjec, J. Doorbar, L. Banks, Interaction of the Human Papillomavirus E6 Oncoprotein with Sorting Nexin 27 Modulates Endocytic Cargo Transport Pathways, *PLoS Pathog.* 12 (2016) e1005854. <https://doi.org/10.1371/journal.ppat.1005854>.
- [28] C.D. James, S. Roberts, Viral Interactions with PDZ Domain-Containing Proteins-An Oncogenic Trait?, *Pathogens.* 5 (2016). <https://doi.org/10.3390/pathogens5010008>.
- [29] R.T. Javier, A.P. Rice, Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses, *J. Virol.* 85 (2011) 11544–11556. <https://doi.org/10.1128/JVI.05410-11>.
- [30] Z. Songyang, A.S. Fanning, C. Fu, J. Xu, S.M. Marfatia, A.H. Chishti, A. Crompton, A.C. Chan, J.M. Anderson, L.C. Cantley, Recognition of unique carboxyl-terminal motifs by distinct PDZ domains, *Science.* 275 (1997) 73–77. <https://doi.org/10.1126/science.275.5296.73>.
- [31] H. Tochio, Q. Zhang, P. Mandal, M. Li, M. Zhang, Solution structure of the extended neuronal nitric oxide synthase PDZ domain complexed with an associated peptide, *Nat. Struct. Biol.* 6 (1999) 417–421. <https://doi.org/10.1038/8216>.
- [32] S. Kalyoncu, O. Keskin, A. Gursoy, Interaction prediction and classification of PDZ domains, *BMC Bioinformatics.* 11 (2010) 357. <https://doi.org/10.1186/1471-2105-11-357>.
- [33] N. Babault, F. Cordier, M. Lafage, J. Cockburn, A. Haouz, C. Prehaud, F.A. Rey, M. Delepierre, H. Buc, M. Lafon, N. Wolff, Peptides targeting the PDZ domain of PTPN4 are efficient inducers of glioblastoma cell death, *Structure.* 19 (2011) 1518–1524. <https://doi.org/10.1016/j.str.2011.07.007>.
- [34] E. Terrien, A. Chaffotte, M. Lafage, Z. Khan, C. Préhaud, F. Cordier, C. Simenel, M. Delepierre, H. Buc, M. Lafon, N. Wolff, Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN, *Sci Signal.* 5 (2012) ra58. <https://doi.org/10.1126/scisignal.2002941>.
- [35] P. Maisonneuve, C. Caillet-Saguy, M.-C. Vaney, B.-Z. Edo, K. Sawyer, B. Raynal, M. Delepierre, M. Lafon, F. Cordier, N. Wolff, Molecular Basis of The Interaction of the Human Protein Tyrosine Phosphatase Non-receptor Type 4 (PTPN4) with the Mitogen-Activated Protein Kinase p38 γ , *J. Biol. Chem.* (2016). <https://doi.org/10.1074/jbc.M115.707208>.
- [36] C. Préhaud, N. Wolff, E. Terrien, M. Lafage, F. Mégret, N. Babault, F. Cordier, G.S. Tan, E. Maitrepierre, P. Ménager, D. Choppy, S. Hoos, P. England, M. Delepierre, M.J. Schnell, H. Buc, M. Lafon, Attenuation of rabies virulence: takeover by the cytoplasmic domain of its envelope protein, *Sci Signal.* 3 (2010) ra5. <https://doi.org/10.1126/scisignal.2000510>.
- [37] L.M. Stannard, M. Hodgkiss, Morphological irregularities in Dane particle cores, *J. Gen. Virol.* 45 (1979) 509–514. <https://doi.org/10.1099/0022-1317-45-2-509>.
- [38] C.M. Hilditch, L.J. Rogers, D.H. Bishop, Physicochemical analysis of the hepatitis B virus core antigen produced by a baculovirus expression vector, *J. Gen. Virol.* 71 (Pt 11) (1990) 2755–2759. <https://doi.org/10.1099/0022-1317-71-11-2755>.
- [39] C.R. Bourne, S.P. Katen, M.R. Fulz, C. Packianathan, A. Zlotnick, A mutant hepatitis B virus core protein mimics inhibitors of icosahedral capsid self-assembly, *Biochemistry.* 48 (2009) 1736–1742. <https://doi.org/10.1021/bi801814y>.
- [40] A.M. Di Bisceglie, Hepatitis B And Hepatocellular Carcinoma, *Hepatology.* 49 (2009)

S56–S60. <https://doi.org/10.1002/hep.22962>.

- [41] L. Yadav, F. Tamene, H. Göös, A. van Drogen, R. Katainen, R. Aebersold, M. Gstaiger, M. Varjosalo, Systematic Analysis of Human Protein Phosphatase Interactions and Dynamics, *Cell Systems*. 4 (2017) 430–444.e5. <https://doi.org/10.1016/j.cels.2017.02.011>.
- [42] J. Hatakeyama, J.H. Wald, I. Printsev, H.-Y.H. Ho, K.L. Carraway, Vangl1 and Vangl2: planar cell polarity components with a developing role in cancer, *Endocr. Relat. Cancer*. 21 (2014) R345–356. <https://doi.org/10.1530/ERC-14-0141>.
- [43] G.O. Cetin, A. Toyly, N. Atabey, Z. Sercan, M. Sakizli, Downregulation of VANGL1 inhibits cellular invasion rather than cell motility in hepatocellular carcinoma cells without stimulation, *Genet Test Mol Biomarkers*. 19 (2015) 283–287. <https://doi.org/10.1089/gtmb.2015.0014>.
- [44] M. Beck, P. Schirmacher, S. Singer, Alterations of the nuclear transport system in hepatocellular carcinoma – New basis for therapeutic strategies, *Journal of Hepatology*. 67 (2017) 1051–1061. <https://doi.org/10.1016/j.jhep.2017.06.021>.
- [45] T.N.J. Flores-Téllez, T.V. Lopez, V.R. Vásquez Garzón, S. Villa-Treviño, Co-Expression of Ezrin-CLIC5-Podocalyxin Is Associated with Migration and Invasiveness in Hepatocellular Carcinoma, *PLoS ONE*. 10 (2015) e0131605. <https://doi.org/10.1371/journal.pone.0131605>.
- [46] Y.-H. Hung, Y.-S. Chan, Y.-S. Chang, K.-T. Lee, H.-P. Hsu, M.-C. Yen, W.-C. Chen, C.-Y. Wang, M.-D. Lai, Fatty acid metabolic enzyme acyl-CoA thioesterase 8 promotes the development of hepatocellular carcinoma, *Oncology Reports*. 31 (2014) 2797–2803. <https://doi.org/10.3892/or.2014.3155>.
- [47] X. Yue, F. Yang, Y. Yang, Y. Mu, W. Sun, W. Li, D. Xu, J. Wu, Y. Zhu, Induction of Cyclooxygenase-2 Expression by Hepatitis B Virus Depends on Demethylation-associated Recruitment of Transcription Factors to the Promoter, *Virol J*. 8 (2011) 118. <https://doi.org/10.1186/1743-422X-8-118>.
- [48] A.S.-L. Cheng, H.L.-Y. Chan, W.K. Leung, K.F. To, M.Y.-Y. Go, J.Y.-H. Chan, C.T. Liew, J.J.-Y. Sung, Expression of HBx and COX-2 in chronic hepatitis B, cirrhosis and hepatocellular carcinoma: implication of HBx in upregulation of COX-2, *Mod. Pathol*. 17 (2004) 1169–1179. <https://doi.org/10.1038/modpathol.3800196>.
- [49] S. Rousseau, N. Morrice, M. Pegg, D.G. Campbell, M. Gaestel, P. Cohen, Inhibition of SAPK2a/p38 prevents hnRNP A0 phosphorylation by MAPKAP-K2 and its interaction with cytokine mRNAs, *EMBO J*. 21 (2002) 6505–6514. <https://doi.org/10.1093/emboj/cdf639>.
- [50] H. Ma, X.-L. Zhao, X.-Y. Wang, X.-W. Xie, J.-C. Han, W.-L. Guan, Q. Wang, L. Zhu, X.-B. Pan, L. Wei, 2',3'-Cyclic Nucleotide 3'-Phosphodiesterases Inhibit Hepatitis B Virus Replication, *PLOS ONE*. 8 (2013) e80769. <https://doi.org/10.1371/journal.pone.0080769>.
- [51] W.-Q. Zang, B. Li, P.-Y. Huang, M.M.C. Lai, T.S.B. Yen, Role of Polypyrimidine Tract Binding Protein in the Function of the Hepatitis B Virus Posttranscriptional Regulatory Element, *J Virol*. 75 (2001) 10779–10786. <https://doi.org/10.1128/JVI.75.22.10779-10786.2001>.
- [52] A. Zlotnick, B. Venkatakrishnan, Z. Tan, E. Lewellyn, W. Turner, S. Francis, Core protein: A pleiotropic keystone in the HBV lifecycle, *Antiviral Research*. 121 (2015) 82–93. <https://doi.org/10.1016/j.antiviral.2015.06.020>.
- [53] S.N. Naccache, T. Hasson, A. Horowitz, Binding of internalized receptors to the PDZ domain of GIPC/synectin recruits myosin VI to endocytic vesicles, *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006) 12735–12740. <https://doi.org/10.1073/pnas.0605317103>.
- [54] R. Razanskas, K. Sasnauskas, Interaction of hepatitis B virus core protein with human GIPC1, *Archives of Virology*. 155 (2010) 247–250. <https://doi.org/10.1007/s00705-009-0561-z>.

- [55] R. Rousset, S. Fabre, C. Desbois, F. Bantignies, P. Jalinot, The C-terminus of the HTLV-1 Tax oncoprotein mediates interaction with the PDZ domain of cellular proteins, *Oncogene*. 16 (1998) 643–654. <https://doi.org/10.1038/sj.onc.1201567>.
- [56] A. Favre-Bonvin, C. Reynaud, C. Kretz-Remy, P. Jalinot, Human papillomavirus type 18 E6 protein binds the cellular PDZ protein TIP-2/GIPC, which is involved in transforming growth factor beta signaling and triggers its degradation by the proteasome., *Journal of Virology*. 79 (2005) 4229–37. <https://doi.org/10.1128/JVI.79.7.4229-4237.2005>.
- [57] P. Vanlandschoot, F. Van Houtte, B. Serruys, G. Leroux-Roels, The arginine-rich carboxy-terminal domain of the hepatitis B virus core protein mediates attachment of nucleocapsids to cell-surface-expressed heparan sulfate, *J. Gen. Virol.* 86 (2005) 75–84. <https://doi.org/10.1099/vir.0.80580-0>.
- [58] V. Bruss, Hepatitis B virus morphogenesis, *World J. Gastroenterol.* 13 (2007) 65–73. <https://doi.org/10.3748/wjg.v13.i1.65>.
- [59] D. Meng, R.P. Hjelm, J. Hu, J. Wu, A theoretical model for the dynamic structure of hepatitis B nucleocapsid, *Biophys. J.* 101 (2011) 2476–2484. <https://doi.org/10.1016/j.bpj.2011.10.002>.
- [60] C. Chen, J.C.-Y. Wang, A. Zlotnick, A kinase chaperones hepatitis B virus capsid assembly and captures capsid dynamics in vitro, *PLoS Pathog.* 7 (2011) e1002388. <https://doi.org/10.1371/journal.ppat.1002388>.
- [61] X. Yu, L. Jin, J. Jih, C. Shih, Z.H. Zhou, 3.5Å cryoEM structure of hepatitis B virus core assembled from full-length core protein, *PLoS ONE*. 8 (2013) e69729. <https://doi.org/10.1371/journal.pone.0069729>.
- [62] N.R. Watts, J.F. Conway, N. Cheng, S.J. Stahl, D.M. Belnap, A.C. Steven, P.T. Wingfield, The morphogenic linker peptide of HBV capsid protein forms a mobile array on the interior surface, *EMBO J.* 21 (2002) 876–884. <https://doi.org/10.1093/emboj/21.5.876>.
- [63] Y.J. Im, S.H. Park, S.-H. Rho, J.H. Lee, G.B. Kang, M. Sheng, E. Kim, S.H. Eom, Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class II PDZ target recognition and PDZ domain-mediated multimerization, *J. Biol. Chem.* 278 (2003) 8501–8507. <https://doi.org/10.1074/jbc.M212263200>.
- [64] J. Heger-Stevic, P. Zimmermann, L. Lecoq, B. Böttcher, M. Nassal, Hepatitis B virus core protein phosphorylation: Identification of the SRPK1 target sites and impact of their occupancy on RNA binding and capsid structure, *PLoS Pathog.* 14 (2018). <https://doi.org/10.1371/journal.ppat.1007488>.
- [65] F. Birnbaum, M. Nassal, Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein, *J. Virol.* 64 (1990) 3319–3330.
- [66] P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling, *Biophys. J.* 78 (2000) 1606–1619. [https://doi.org/10.1016/S0006-3495\(00\)76713-0](https://doi.org/10.1016/S0006-3495(00)76713-0).
- [67] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins*. 59 (2005) 687–696. <https://doi.org/10.1002/prot.20449>.
- [68] Y. Nakatani, V. Ogryzko, Immunoaffinity purification of mammalian protein complexes, *Meth. Enzymol.* 370 (2003) 430–444. [https://doi.org/10.1016/S0076-6879\(03\)70037-8](https://doi.org/10.1016/S0076-6879(03)70037-8).
- [69] TCA Precipitation, *Methods in Enzymology*. 541 (2014) 3–10. <https://doi.org/10.1016/B978-0-12-420119-4.00001-X>.
- [70] B. Hommel, A. Sturny-Leclère, S. Volant, N. Veluppillai, M. Duchateau, C.-H. Yu, V. Hourdel, H. Varet, M. Matondo, J.R. Perfect, A. Casadevall, F. Dromer, A. Alanio, *Cryptococcus neoformans* resists to drastic conditions by switching to viable but non-

culturable cell phenotype, *PLoS Pathog.* 15 (2019) e1007945.

<https://doi.org/10.1371/journal.ppat.1007945>.

[71] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, *J. Proteome Res.* 10 (2011) 1794–1805. <https://doi.org/10.1021/pr101065j>.

[72] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372. <https://doi.org/10.1038/nbt.1511>.

[73] S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics, *Nat Protoc.* 11 (2016) 2301–2319.

<https://doi.org/10.1038/nprot.2016.136>.

[74] S. Benhenda, A. Ducroux, L. Rivière, B. Sobhian, M.D. Ward, S. Dion, O. Hantz, U. Protzer, M.-L. Michel, M. Benkirane, O.J. Semmes, M.-A. Buendia, C. Neuveut, Methyltransferase PRMT1 is a binding partner of HBx and a negative regulator of hepatitis B virus transcription, *J. Virol.* 87 (2013) 4360–4371. <https://doi.org/10.1128/JVI.02574-12>.

[75] L. Rivière, B. Quioc-Salomon, G. Fallot, B. Halgand, C. Féray, M.-A. Buendia, C. Neuveut, Hepatitis B virus replicating in hepatocellular carcinoma encodes HBx variants with preserved ability to antagonize restriction by Smc5/6, *Antiviral Res.* 172 (2019) 104618. <https://doi.org/10.1016/j.antiviral.2019.104618>.

Figure 1. PDZome binding of HBc investigated by *in vitro* holdup protein–peptide binding assays. (a) PDZome-binding profile of PBM-HBc. PDZ domains are ranked on the basis of their BIs for the PBM-HBc peptide. A dashed line indicates the cut-off for a significant PDZ–PBM interaction (BI > 0.2). (b) Zoom-in on the 28 HBc binders with a BI > 0.2. The data and error bars are representative of two independent experiments.

Figure 2. X-ray structure of PTPN3-PDZ bound to PBM-HBc. (a) Superposition of the structures of PTPN3-PDZ bound to PBM-HBc (PDB ID: 6T36, light grey) and PBM-HPV16E6 (PDB ID: 6HKS, light green). (b) Representation of the density of the peptide in the binding groove of the PDZ domain. (c) Detailed views of the PTPN3-PDZ bound to PBM-HBc. Important residues are shown as sticks in CPK colors, molecules of water are represented as red spheres, and intermolecular H-bonds and polar contacts are shown as yellow dashed lines. Black dashed lines show the H-bond pairing with the β 2-strand.

Figure 3. $\Delta\delta$ between free PTPN3-PDZ and PTPN3-PDZ bound to PBM-HBc. (a) $\Delta\delta$ values (^1H , ^{15}N) computed as $\Delta\delta = [(\Delta\delta\text{H})^2 + (\Delta\delta\text{N} \times 0.15)^2]^{1/2}$. Red bars mark residues perturbed with $\Delta\delta > 0.19$ ppm; magenta circles mark residues of PTPN3-PDZ severely affected when binding to the PBM with signals that disappear from their original well-resolved position in the spectrum upon complex formation (b) Views of the structure of PTPN3-PDZ bound to PBM-HBc (PDB ID: 6T36) with the color depending on $\Delta\delta$ values. Red: $\Delta\delta > 0.19$ ppm; magenta: resonances absent in the complexed PTPN3-PDZ spectrum; grey: residues whose behaviour could not be safely defined, mainly because they fall in crowded spectral regions. Residues affected are represented as spheres. PBM-HBc peptide is represented as yellow sticks with its H-bonds as dashed yellow lines. (c) HSQC NMR spectra of ^{15}N -labelled PTPN3-PDZ free (in red) or in presence of an excess of full-length HBc (in black). (d) AUC diagram of the full length HBc sample used for NMR.

Figure 4. GST pull-down assay of GST-tagged PTPN3 PDZ domain against HeLa S3 cell lysate overexpressing HBc. (a) GST (negative control) and GST fusion proteins were bound to the same amount of glutathione-agarose beads and allowed to interact with HA-tagged HBc proteins from lysates of HeLa S3 cells overexpressing either the HA-tagged full-length WT HBc or the Y132A mutant. Samples were washed and analyzed by western blot with anti-HA antibodies. (upper part) *Lanes 1 and 2*: GST-PTPN3-PDZ in absence or presence of lysate containing HA-HBc WT respectively; *Lanes 3 and 4*: GST in absence or presence of lysate containing HA-HBc WT respectively; Lane 5 shows the signal of HA-HBc WT in lysates of HeLa S3 cell lysate. (lower part) *Lanes 1 and 2*: GST-PTPN3-PDZ in absence or presence of lysate containing HA-HBc Y132A respectively; *Lanes 3 and 4*: GST in absence or presence of lysate containing HA-HBc Y132A respectively; Lane 5 shows the signal of HA-HBc Y132A in lysates of HeLa S3 cell lysate. (left) Ticks and labels indicate molecular weight markers. (b) Native Western blot on the HeLa S3 HA-HBc WT and HA-HBc Y132A using the anti-HA antibody.

Figure 5. Effects of PTPN3 overexpression on HBV life cycle (a) PTPN3 expression in HepG2 NTCP or HepG2 NTCP-PTPN3 cells was tested by RT-qPCR. *P* values were determined by Mann-Whitney test. Results are expressed as the average of at least three independent experiments. Standard error of the mean (SEM) are indicated. Error bars represent SD of at least three independent experiments. (b) Effect of PTPN3 expression: (left panel) on cccDNA levels in HepG2 NTCP or HepG2 NTCP-PTPN3 cells (cccDNA levels was analyzed by qPCR); (second panel) on HBV total RNA (HBV RNA was analyzed and

quantified by RT-qPCR); (third panel) on cytoplasmic RC DNA (RC DNA is quantified by qPCR); (right panel) on secreted RC DNA (RC DNA is quantified by qPCR).

	BI	HPV16&18	HTLV-1	Adenovirus	Rabies
GIPC1	0,85	x	x		
MAGI3	0,78	x	x		
PTPN4	0,58				x
SNTB1	0,54		x		
DLG4_3	0,53	x			
MAGI2	0,51	x			
MAGI1	0,49	x	x	x	
STXB4	0,47				
MPDZ	0,46	x		x	
PTPN3	0,45	x			
SYNJ2BP	0,43				
DLG4_1	0,43	x			
DLG1_2	0,40	x	x	x	
SNX27	0,39	x			
DLG2_2	0,38				x
SNTB2	0,38				
DLG3_2	0,36				
DLG4_2_	0,35	x			
DLG2_1	0,35				x
DLG1_3	0,34	x	x	x	
DLG1_1	0,34	x	x	x	
GIPC2	0,31				
InaD1	0,29	x			
TJP2	0,27			x	
SNTA1	0,25				
SIPA1L2	0,24				
SCRIB	0,23	x	x		
MAST2	0,21				x

Table 1. Best binders of HBc PBM in the holdup assay. Binding intensities (BIs) and occurrences between HBV and others viruses are reported.

PTPN3-PDZ _{Next} / PBM-HBc	
Data collection	
Space group	P 31 2 1
Unit cell (a, b, c) (Å)	75.58, 75.58, 46.59
a, b, g (°)	90.00, 90.00, 120.00
Resolution (Å)	46.59-1.86 (1.96-1.86)
No. of reflections (total/unique)	67510/13198
Redundancy	5.1 (5.0)
Completeness (%)	99.4 (97.8)
<i>I</i> / <i>s</i> (<i>I</i>)	17.38 (2.70)
<i>R</i> -meas [¶] (%)	5.6 (57.5)
CC(1/2)	99.9 (97.7)
Refinement	
Resolution (Å)	29.35-1.86
No. reflections	13066
<i>R</i> _{work} [†] / <i>R</i> _{free} [‡]	0.197 / 0.236
No. of protein atoms/ligand atoms	749/63
No. of solvent/hetero-atoms	68/1
Rmsd bond lengths (Å)	0.007
Rmsd bond angles (°)	0.836
<i>Wilson B</i> -factors	28.6
Ramachandran plot (favored/disallowed)*	96.9/0.0
PDB code	6T36

Values in parenthesis correspond to the highest resolution shell

[¶] $R_{meas} = \frac{\sum h(n/n-1)^{1/2} \sum_i |I_i(h) - \langle I(h) \rangle|}{\sum h \sum_i I_i(h)}$, where $I_i(h)$ and $\langle I(h) \rangle$ are the *i*th and the mean measurement, respectively, of the intensity of reflection *h*.

[†] $R_{work} = \frac{\sum h ||F_{obs}(h) - F_{calc}(h)||}{\sum h |F_{obs}(h)|}$, where $F_{obs}(h)$ and $F_{calc}(h)$ are the observed and calculated structure factors, respectively. No *I*/ σ cut-off was applied.

[‡] *R*_{free} is the *R* value obtained for a test set of reflections consisting of a randomly selected 5% subset of the data set excluded from refinement.

*Categories were defined by MolProbity.

Table 2. Data collection and refinement statistics

Figure 1

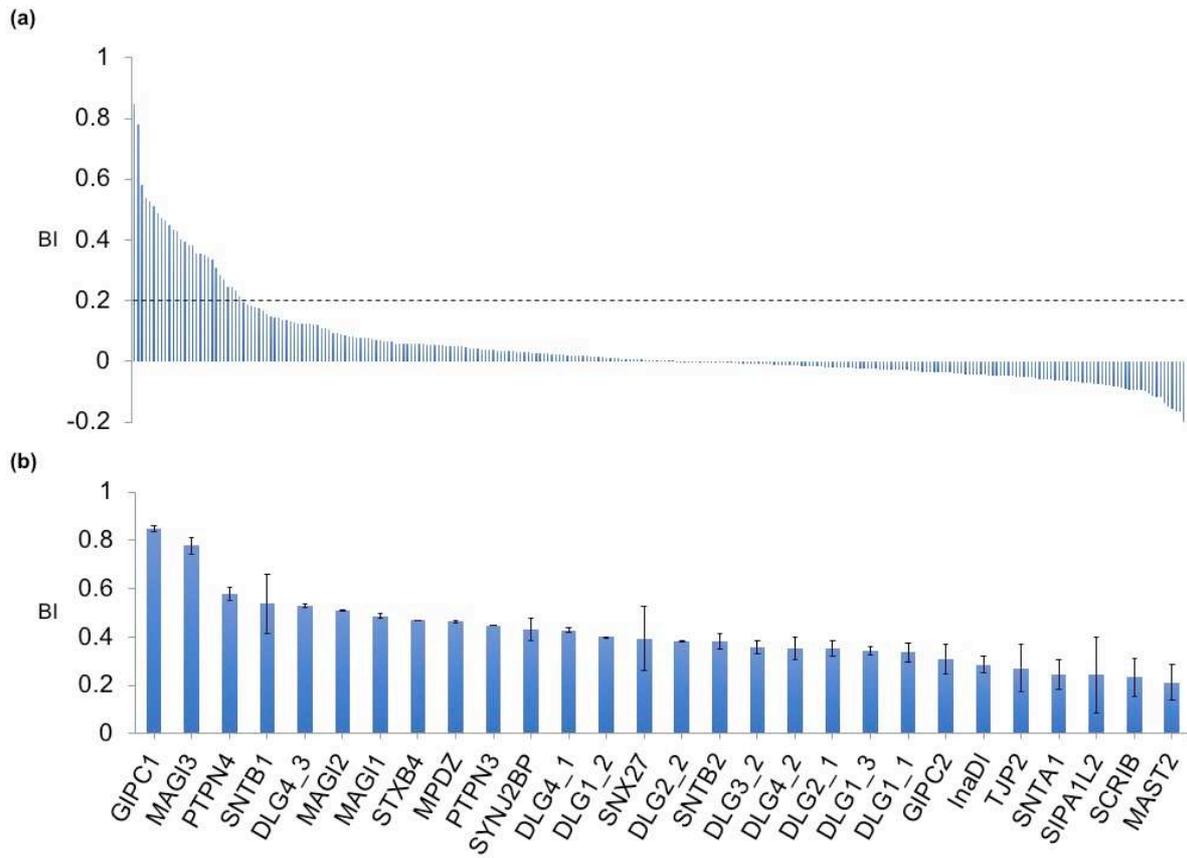


Figure 2

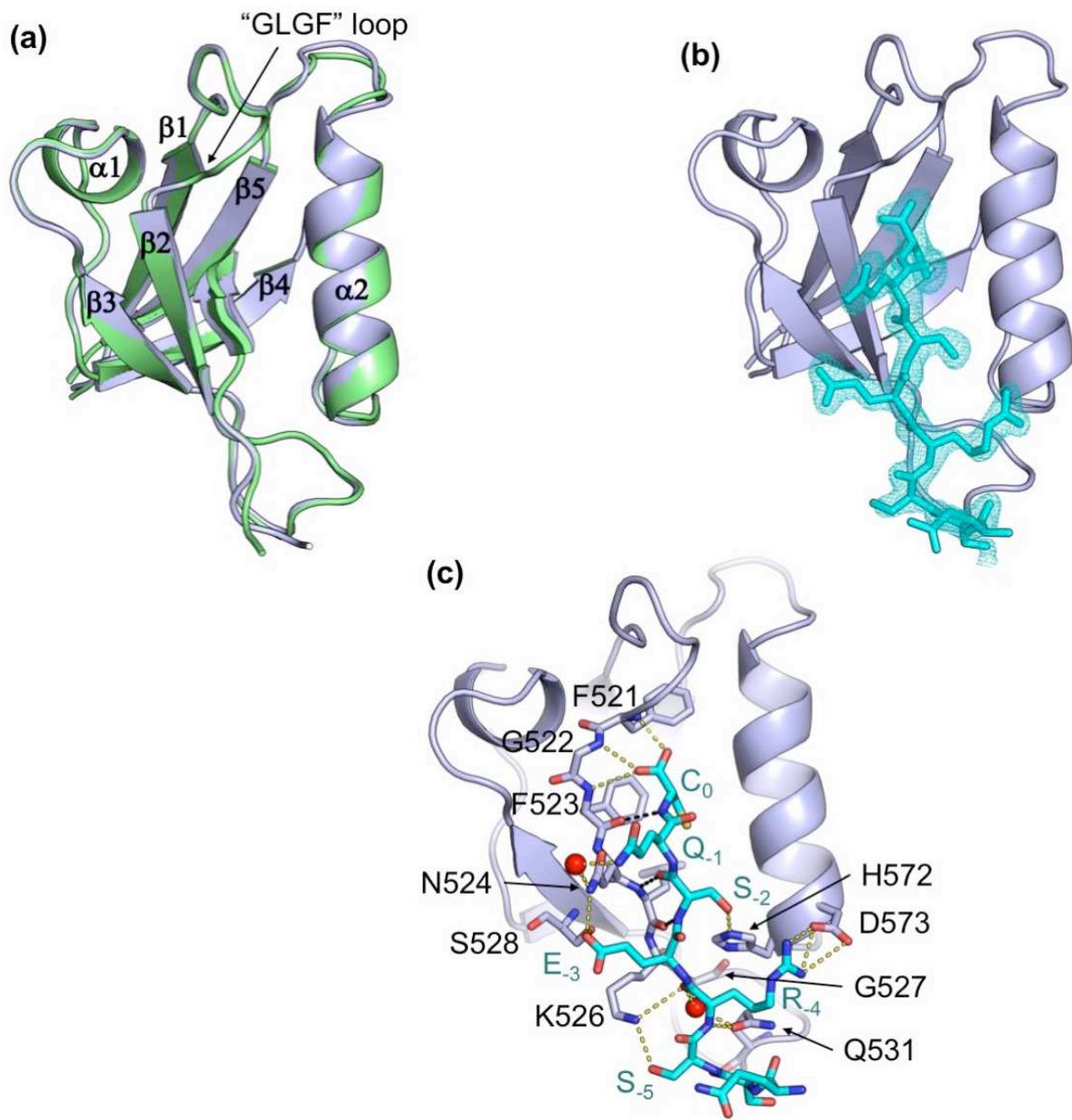


Figure 3

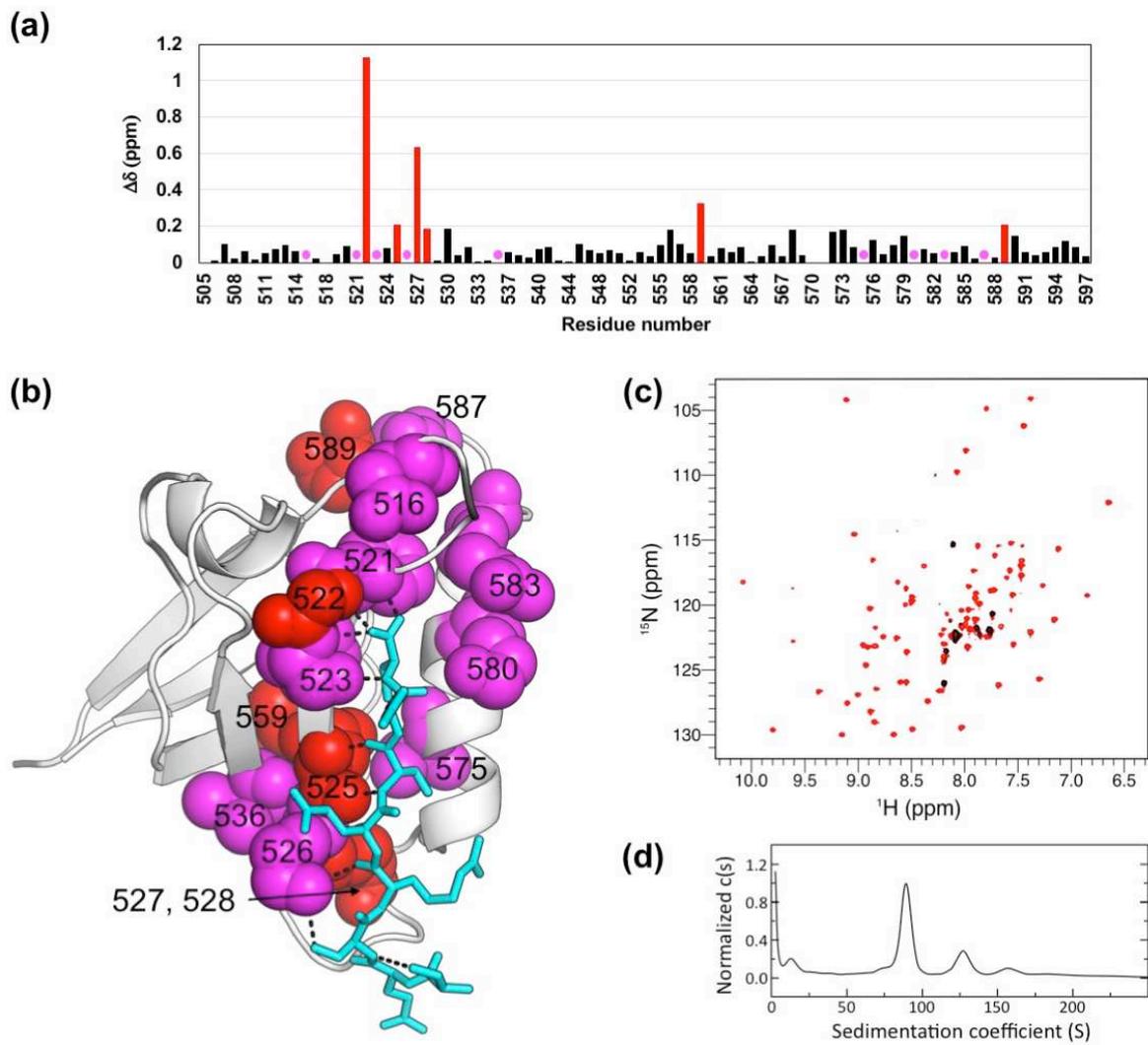


Figure 4

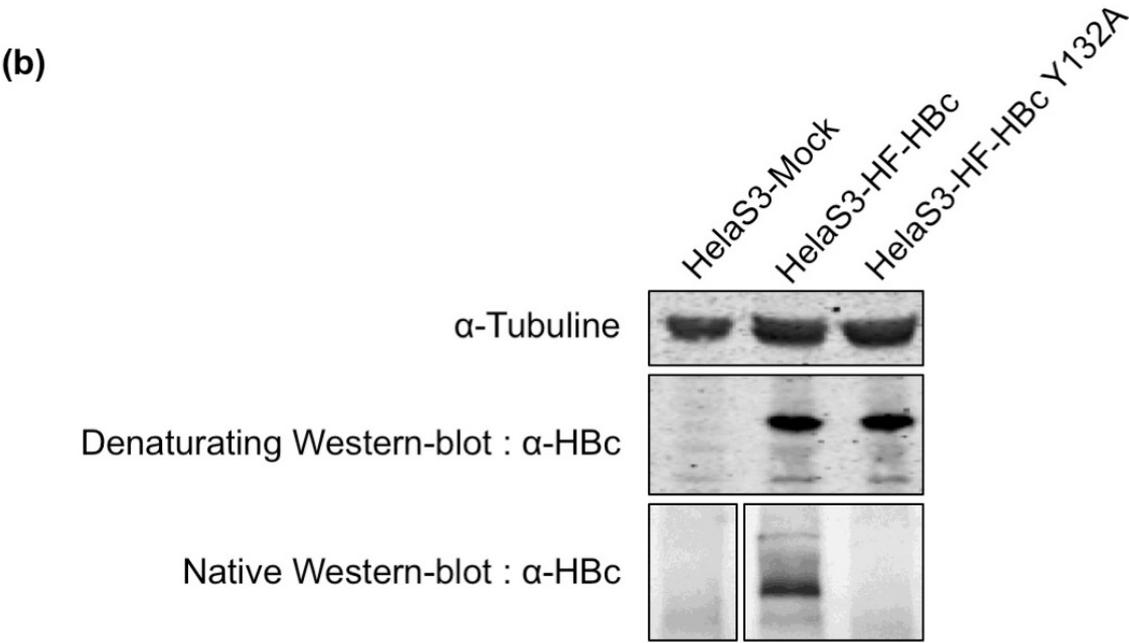
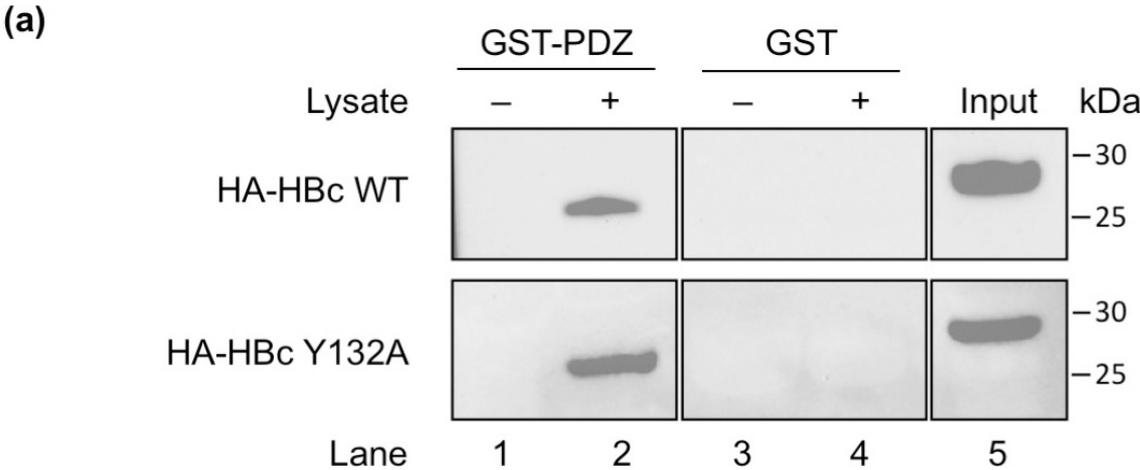
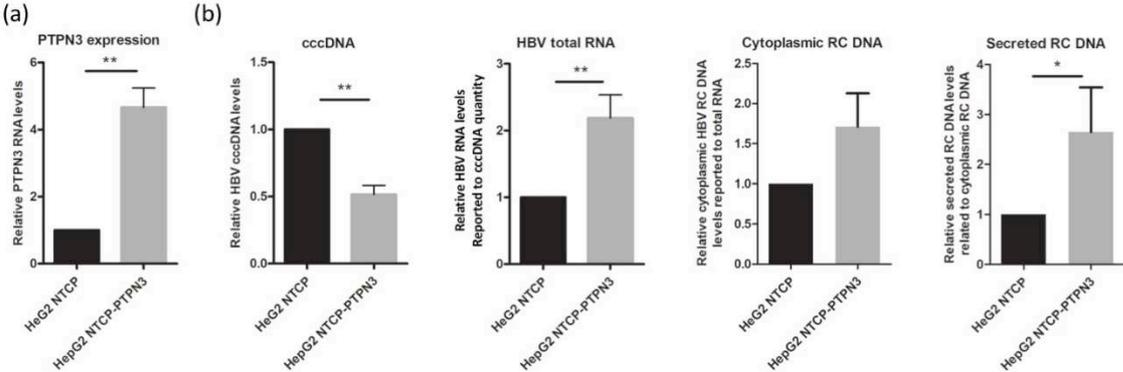


Figure 5



7.3 Comments

In this work, we began exploring the interplay between PTPN3 and HBV. After confirming that PTPN3 is able to bind HBc in both its capsid and dimeric forms, we obtained the first data about the multiple effects of PTPN3 in HBV infection. One possible explanation we propose for these effects is that PTPN3 modulates capsid stability by altering its phosphorylation state. HBc presents several tyrosines that are susceptible of being phosphorylated, including the tyrosine 132 (Y132), which is crucial for capsid assembly³¹⁸. Indeed, the phenolic side chain of Y132 inserts into a hydrophobic pocket of a neighboring subunit, contributing to around 10% of the surface buried by interdimer hydrophobic contacts. Mutation of Y132 to alanine renders the protein unable to form capsids. We hypothesized that capsid assembly or stability could be regulated by Y132 phosphorylation, and PTPN3 could be implicated by specifically dephosphorylating this residue. However, there is currently no data on tyrosine phosphorylation in HBc. To investigate this, we attempted to identify phosphorylated tyrosines in HBc. We used our GST-PTPN3-PDZ construct to purify wild type HA-tagged HBc from HBV-infected HepG2 NTCP cells, and subsequently tried to identify phosphorylated tyrosines by mass spectrometry. However, due to the low relative abundance of tyrosine phosphorylation and that we lacked the tools to perform an enrichment of this type of samples, our attempts were not successful. As an alternative, we also tried to obtain this result by Western blot. Using the same purification technique, we resolved the proteins by SDS-PAGE and blotted them on PVDF membranes. Our plan was to use anti-phosphotyrosine antibody to detect Tyr-phosphorylated species, and then identify the band corresponding to HBc by re-blotting with anti-HA antibody. We also pre-treated one of the HBc samples with a large amount of PTPN3-Linker-PTP to assess if PTPN3 was able to specifically dephosphorylate HBc. Although PTPN3 was able to significantly reduce the levels of tyrosine phosphorylation in the samples, the blotting was very noisy and we could not obtain a clear band for HBc. We were unable to optimize this approach due to time constraints. However, given the significance on Y132 in capsid assembly, it will be exciting to further pursue this line of research in the future.

Concluding remarks

8 Discussion and perspectives

Protein interaction modules evolved hand in hand with multi-cellularity, becoming essential for enzyme function specialization by enhancing specificity and function. These modules determine interactions that add to the complexity of functions associated to proteins whose catalytic domains *per se* would be unable to achieve. It is clear from the large amounts of evidence gathered so far that protein interaction modules represent a particularly attractive target for pathogenic microorganisms. The recognition of many protein-protein interaction domains involves short unfolded sequences exposed in larger intrinsically disordered regions of proteins. It has been shown that viral proteins are more prone to target existing endogenous protein-protein interaction interfaces rather than creating new interfaces. The protein-protein interactions mimicked by viral proteins, such as PDZ/PBM interactions, are frequently involved in multiple and transient regulatory interactions³⁵⁸. As it was previously mentioned, the short sequence needed to establish an interaction with a PDZ domain and the fact that PBMs are located in flexible or disordered regions makes the appearance of these motifs relatively easy from an evolutionary point of view.

In this thesis, I have developed the study of the human protein tyrosine phosphatase PTPN3, with a particular focus on its PDZ domain and its PDZ-mediated interactions. For this, I have employed an integrative and multidisciplinary approach.

The first part of this project focused on studying the specificity determinants of PTPN3 for its PBM ligands. My work on PTPN3-PDZ contributed to the first structural data on this domain. The structure of the domain provides insights into the structural determinants of ligand binding. By using NMR spectroscopy, we identified short- and long-range effects that occur upon the formation of the complex with the HPV 16 E6 PBM, including a perturbation pathway that appears to communicate the binding cleft with distal regions in the domain. This could be an indication of allostery in the PDZ domain of PTPN3. A few examples demonstrate that PDZ domains are capable of supporting allosteric mechanisms^{171,359-361}. In PTPN4, and most likely in PTPN3, an allosteric mechanism involving transient interactions between the PDZ and PTP domains regulates the activity of the catalytic domain. The dynamics of the PDZ domains could be relevant in this regulation. Additionally, the observation that PBM binding stabilized the monomeric form of PTPN3-PDZ suggests a modification of the domain's dynamics upon complex formation, although we

were unable to correlate these effects to a regulation of PTPN3 catalytic inhibition by its PDZ domain.

Several viruses and cellular proteins target the PDZ domain on PTPN3 through PBMs, potentially regulating homeostasis or affecting the function of PTPN3. I was interested in studying the basis of the specificity of recognition of such short, unstructured motifs. Therefore, I further characterized the determinants of PDZ ligand recognition of PTPN3 by solving the crystal structures of the complexes of PTPN3-PDZ with PBM peptides derived from viral and cellular PTPN3 partners. I found that all of these ligands establish a similar bonding pattern that involves residues that are conserved in the PTPN3 and PTPN4 PDZ domains, but less often in the PDZome, and thus could represent determinants for PBM ligand sequence preference of the NT5 phosphatase subfamily. Moreover, I used proteomics data to explore the sequence consensus of the cellular partners that PTPN3-PDZ preferentially binds, and found that these data was in agreement with the observations derived from the crystal structures. Some of the conserved positions in PTPN3 and PTPN4 mediate contacts beyond the canonical 3-residue PBM that are not required for the interactions to occur, but possibly favour the selection of some PBM partners over others by increasing their interaction affinities. We have seen that PDZ domains are promiscuous protein-protein interaction modules that bind their partners with low-to-medium affinities (1-100 μ M), which is related to the transient nature of signalling interactions. The specific polar bonds and hydrophobic contacts that the preferred PTPN3 ligands establish via their positions -3 and -4 are likely to enhance their binding over other potential PBM-containing partners. Additionally, it is likely that a further degree of selectivity is achieved thanks to the spatial segregation of the protein by its FERM domain, which targets the phosphatase to the interface of the membrane and the cytoskeleton, promoting interaction with certain ligands or substrates over others. This was supported by the observation that both the FERM and PTP domains of PTPN3 are required for attenuation of HBV genome expression⁶⁸. It is interesting to consider that from the three isoforms of PTPN3 that have been described, two are likely to lack this spatial segregation due to their incomplete FERM domains. Importantly, these isoforms are likely to be more active than full-length PTPN3, as suggested by *in vitro* limited proteolysis studies⁵⁶. Unfortunately, there is currently no data on the subcellular location or the physiological relevance of these isoforms. One can only hypothesize about their potential role, and the relevance of their PDZ domains for selecting substrates or anchoring these enzymes to signalling complexes. To date, the role of PTPN3 in cell physiology has been studied with 'classic' methodologies, either by overexpressing the protein in cell or by decreasing its

expression levels using techniques such as siRNA or shRNA. It would be interesting to use genetic engineering techniques to generate cell lines that express only a particular isoform of PTPN3 at endogenous levels to avoid artefacts of overexpression, and this way dissect the individual roles of the splice variants of this protein.

The second part of this project focused on studying the role of PTPN3 in cell signalling and to understand the effects of PTPN3 targeting during HBV infection through the viral HBc protein. I investigated the potential interactome of the HBc PBM, and identified novel potential PDZ protein partners. Many of these protein partners are scaffolding proteins that are involved in cell polarity and/or cell-cell contacts. It was interesting to find that among the pool of the HBc targets, a large majority has been reported as targets of other viruses, suggesting that these interactions are potentially relevant for HBV infection. Despite the distinct replication cycles of these viruses, they all target at least some members of the same subset of cellular PDZ proteins. It is likely that facing similar cellular environments has forced these viruses to adopt analogous strategies to favour their dissemination by hijacking the same cellular pathways or processes.

PTPN3 was identified as one of the binders of the HBc PBM, and I was interested in understanding its role in viral infection. By overexpressing PTPN3 in HBV-infected hepatocytes, we obtained the first indications that PTPN3 has an active role in the viral life cycle and impacts multiple virological parameters. Given the central importance of HBc in HBV biology and its interaction with PTPN3, we hypothesized that the effects on the viral processes could be due to a modulation of the stability of the viral capsid by PTPN3. Indeed, I confirmed that PTPN3 binds the PBM of HBc when HBc is forming dimers or capsids, indicating that this interaction could occur in different steps of the viral life cycle. HBc contains phosphorylated tyrosines that are potential substrates of PTPN3. The HBc Y132 is well known to drive capsid assembly by burying its hydrophobic side chain³¹⁸. Binding of the HBc PBM could be a mechanism to favour dephosphorylation in a specific stage of capsid maturation, as it has been suggested that HBc CTD exposure could be regulated by the nucleic acid content of the capsid³¹⁴. It is possible that phosphorylation of Y132 represents a destabilizing factor for the viral capsid. Therefore, PTPN3 overexpression and increased Y132 dephosphorylation could then be considered to increase capsid stability. Currently there is no data on tyrosine phosphorylation in HBc, so this will be an interesting path to further explore. Another possibility is that interaction of HBc with the full-length PTPN3 prevents capsid assembly. When I probed the interaction between PTPN3 and the viral capsid I used the PTPN3-PDZ construct and not the full-length protein. It is necessary to evaluate whether

the binding of full-length PTPN3 causes a steric hindrance for the assembly of the viral capsid. This effect has been proposed for the SRPK kinase³¹⁴. The authors of this work speculate that SRPK could bind HBc and act as a chaperone to sterically prevent HBc self-assembly. Additionally, SRPK phosphorylates serine residues in the CTD of HBc *in vivo*. It is therefore attractive to postulate a hypothetical assembly regulation mechanism that combines the action of SRPK and/or other kinases and PTPN3. These proteins could be acting as molecular chaperones, but also their activities might be involved in the regulation, for example by modulating the phosphorylation state of Y132.

Finally, another relevant development of this thesis project is the first data on the potential interactome of PTPN3-PDZ. Until now, the PDZ ligands of PTPN3 had not been explored. We provided the first data on cellular partners of PTPN3 by its PDZ domain. This is a first step towards uncovering the roles of this protein in cell signalling. The interactions mediated by the PDZ domain of PTPN3 can be a mechanism to enhance the dephosphorylation of specific substrates. For example, the PDZ/PBM interaction is required for dephosphorylation of the PTPN3 substrate p38 γ ⁷⁰. Alternatively, interactions with other partners that are not substrates could place PTPN3 in multiprotein signalling complexes, where the efficiency of signal transduction can be greatly increased, as discussed for the phototransduction cascade in *Drosophila*^{150,157}. A first approach could be to search for other protein-protein interaction modules in the PTPN3-PDZ interactors. In line with this, it will be interesting to explore the interactions between the different partners. Then, as our affinity data supported a possible competition between viral and cellular ligands of PTPN3, we focused on searching for potential links of the PTPN3 ligands with HBV or HCC. We found that a number of PTPN3 partners have been implicated in HBV infection and/or HCC. Further validation will be required to determine which of these interactions are physiologically relevant. In the case of the proteins reported in HCC, all of them are overexpressed in this cancer, suggesting that they participate to the oncogenic activity. For those implicated in HBV infection, it is possible that the HBc PBM disrupts their normal interactions with PTPN3 by competition for binding. This opens interesting possibilities. For example, CNP appears to be involved in cellular response against HBV infection, as it has been found to block viral protein synthesis and reduce viral RNAs. It is possible that PTPN3 is involved in these effects by participating in a signalling pathway of interferon-stimulated genes. HBc could block this process by targeting the PDZ domain of PTPN3, displacing CNP. Similarly, HBc could displace another PBM partner of PTPN3, PTB. As I described in section 7.2, PTB binds the regulatory protein PRE, and both proteins enhance the nuclear export of unspliced

RNAs (such as the viral RNAs). By targeting and disrupting the PTPN3-PTB interaction, HBV could potentially favour the export of its RNAs, and thus enhance virus production.

In conclusion, the combination of structural, biophysical, proteomic and biochemical approaches allowed me to study the functions and interactions of PTPN3 from a broad and multidisciplinary perspective. I obtained a detailed structural understanding of its PDZ-mediated interactions, but also a general idea of its potential implications in HBV infection and cellular signalling. Further work is needed to validate the potential interactions I identified and to test the hypotheses that these initial observations on the physiological role of PTPN3 generated.

Appendices

9 Appendices

9.1 Table: The extended human PTPome.

	Official gene symbol/ID	Protein names	Catalytic signature motif	Specificity	
Classic PTPs (class I Cys-based) ^a (subclass I) (37) (HCSxGxGR)					
Classic RPTPs (class I Cys-based) ² (PTP domain) (20)					
1	PTPRA/5786	RPTP α	HCSAGVGR ^b	pTyr	
2	PTPRB/5787	RPTP β			
3	PTPRC/5788	CD45			
4	PTPRD/5789	RPTP δ			
5	PTPRE/5791	RPTP ϵ			
6	PTPRF/5792	LAR			
7	PTPRG/5793	RPTP γ			
8	PTPRH/5794	RPTP η , SAP-1			
9	PTPRJ/5795	DEP1, CD148			
10	PTPRK/5796	RPTP κ			
11	PTPRM/5797	RPTP μ			
12	PTPRN/5798	IA-2			Phosphatase inactive ^c
13	PTPRN2/5799	IA-2 β , phogrin			PIPs
14	PTPRO/5800	GLEPP1			pTyr
15	PTPRQ/374462	PTPRQ, PTPS31	HCSAGVGR	PIPs (D3 and D5 positions)	
16	PTPRR/5801	PTP-SL, PTPBR7	HCSAGIGR	pTyr	
17	PTPRS/5802	RPTP σ	HCSAGVGR		
18	PTPRT/11122	RPTP ρ	HCSAGAGR		
19	PTPRU/10076 ^d	PCP-2, PTPJ	HCSAGTGR		
20	PTPRZ1/5803	RPTP ζ , RPTP β/ζ , phosphacan	HCSAGVGR		
Classic non-RPTPs (class I Cys-based) ^a (PTP domain) (17)					
21	PTPN1/5770	PTP1B	HCSAGIGR	pTyr	
22	PTPN2/5771	TCPTP			
23	PTPN3/5774	PTPH1			
24	PTPN4/5775	PTP-MEG1			
25	PTPN5/84867	STEP			
26	PTPN6/5777	PTP1C, SHP1			
27	PTPN7/5778	HePTP, LC-PTP			

28	PTPN9/5780	PTP-MEG2		
29	PTPN11/5781	PTP1D, SHP2		
30	PTPN12/5782	PTP-PEST	HCSAGCGR	
31	PTPN13/5783	PTP-BL, PTP-BAS, FAP1	HCSAGIGR	
32	PTPN14/5784	PEZ, PTP36	HCSAGVGR	
33	PTPN18/26469	BDP1	HCSAGCGR	
34	PTPN20A, 20B/653129, 26095 ^e	Typ (copyA, B) ^e	HCSAGIGR	
35	PTPN21/11099	PTPD1	HCSAGVGR	
36	PTPN22/26191	LYP, PEP	HCSAGCGR	
37	PTPN23/25930	HDPTP	HCSSGVGR	
Dual-specificity VH1-like PTPs (DUSPs) (class I Cys-based) ^a (subclass II) (64) (HCxxGxxR) ^f				
Dual-specificity MAPK phosphatases (MKPs) (class I Cys-based) ^a (PTP-DSP domain) (11)				
38	DUSP1/1843	MKP-1		
39	DUSP2/1844	PAC-1	HCQAGISR	
40	DUSP4/1846	MKP-2		
41	DUSP5/1847	B23, hVH3	HCEAGISR	
42	DUSP6/1848	MKP-3	HCLAGISR	
43	DUSP7/1849	MKP-X		
44	DUSP8/1850	hVH5	HCLAGISR	pSer, pThr, pTyr
45	DUSP9/1852	MKP-4	HCLAGVSR	
46	DUSP10/11221	MKP-5	HCQAGVSR	
47	DUSP16/80824	MKP-7	HCLAGISR	
48	STYXL1/51657	MK-STYX, DUSP24	FSTQGISR	Phosphatase inactive
Dual-specificity atypical phosphatases (atypical DUSPs) (class I Cys-based) ^a (PTP-DSP domain) (20)				
Small-size atypical DUSPs (15)				
49	DUSP3/1845	VHR	HCREGYSR	
50	DUSP13/51207 ^g	DUSP13A ^g , BEDP, MDSP	HCVVGVS	
51	DUSP13/51207 ^g	DUSP13B ^g , TMDP	HCAMGVSR	
52	DUSP14/11072	MKP-6	HCAAGVSR	
53	DUSP15/128853	VHY	HCFAGISR	
54	DUSP18/150290	LMWDSP20	HCAAGVSR	
55	DUSP19/142679	LMWDSP3, SKRP1, LDP-2	HCNAGVSR	pSer, pThr, pTyr
56	DUSP21/63904	LMWDSP21	HCMAGVSR	
57	DUSP22/56940	VHX	HCLAGVSR	
58	DUSP23/54935	VHZ	HCALGFGR	
59	DUSP26/78986	MKP-8	HCAVGVS	
60	DUPD1/338599 ^h	DUSP27, DUPD1	HCVMGRSR	

61	DUSP28/285193	VHP	YCKNGRSR	? = unknown substrate
62	PTPMT1/114971	DUSP23, MOSP, PLIP	HCKAGRSR	Phosphatidylglycerophosphate, PIPs
63	STYX/6815	STYX	HGNAGISR	Phosphatase inactive
Other atypical DUSPs (5)				
64	RNGTT/8732	HCE1	HCTHGFNR	5'CAP-mRNA
65	DUSP11/8446	PIR1	HCTHGLNR	
66	DUSP12/11266	HYVH1	HCHAGVSR	pSer, pThr, pTyr?
67	EPM2A/7957	Laforin	HCNAGVGR	Phosphoglycogen, pSer
68	DUSP27/92235 ⁱ		SSEMGISR	Phosphatase inactive
Slingshots (class I Cys-based) ^a (PTP-DSP domain) (3)				
69	SSH1/54434	SSH1	HCKMGVSR	pSer
70	SSH2/85464	SSH2		
71	SSH3/54961	SSH3		
Phosphatases of regenerating liver (PRLs) (class I Cys-based) ^a (PTP-DSP domain) (3)				
72	PTP4A1/7803	PRL-1	HCVAGLGR	pTyr
73	PTP4A2/601584	PRL-2		? = unknown substrate
74	PTP4A3/606449	PRL-3		pSer, pThr, pTyr, PtdIns(4,5)P ₂ (D5 position)
CDC14s (class I Cys-based) ^a (PTP-DSP domain) (4)				
75	CDC14A/8556	CDC14A	HCKAGLGR	pSer, pThr
76	CDC14B/8555	CDC14B		
77	CDKN3/1033	CDKN3, KAP	HCYGGGLGR	
78	PTPDC1/138639	PTP9Q22	HCHAGLGR	? = unknown substrate
PTEN-like (class I Cys-based) ^a (PTP-DSP domain) (8)				
79	PTEN/5728	PTEN	HCKAGKGR	PtdIns(3,4,5)P ₃ (D3 position), pSer, pThr, pTyr
80	TPTE/7179 ^j	TPTE	HCKGGTDR	Phosphataseinactive ^c
81	TPTE2/93492	TPIP, TPTE2, PTEN2	HCKGGKGR	PtdIns(3,4,5)P ₃ , PtdIns(3,4)P ₃ (D3 position), PtdIns(4,5)P ₂ (D5 position)
82	TNS1/7145	Tensin	HNKGNRGR	Phosphatase inactive
83	TNS3/64759	Tensin 3	HCRGGKGR	? = unknown substrate, phosphatase inactive ^c
84	TENC1/23371	Tensin 2, C1-TEN	YCKGNK GK	pTyr, PtdIns(3,4,5)P ₃ k
85	DNAJC6/9829 ^l	Auxilin	HCLDGRAA	Phosphatase inactive
86	GAK/2580 ^{l,m}	Auxilin 2, GAK	HCMDGRAA	
MTMs (class I Cys-based) ^a (PTP-MTM domain) (15)				
87	MTM1/4534	MTM1, myotubularin	HCSDGWDR	PtdIns(3,5)P ₂ , PtdIns(3)P (D3 position)
88	MTMR1/8776	MTMR1		
89	MTMR2/8898	MTMR2, CM4TB		
90	MTMR3/8897	MTMR3		
91	MTMR4/9110	MTMR4		PtdIns(3,5)P ₂ , PtdIns(3)P (D3 position)

				position), pSer
92	SBF1/6305	MTMR5	GLEDGWDI	Phosphatase inactive
93	MTMR6/9107	MTMR6	HCSDGWDR	PtdIns(3,5) P_2 , PtdIns(3) P (D3 position)
94	MTMR7/9108	MTMR7		
95	MTMR8/55613	MTMR8		
96	MTMR9/66036	MTMR9	HGTEGTDS	Phosphatase inactive
97	MTMR10/54893	MTMR10	QEEEGRDL	
98	MTMR11/10903	MTMR11	QERGDRDL	
99	MTMR12/606501	MTMR12	LEENASDL	
100	SBF2/81846	MTMR13	CLEEGWDI	
101	MTMR14/64419	MTMR14, MIP	HCISGWDR	PtdIns(3,5) P_2 , PtdIns(3) P (D3 position)
SAC phosphoinositide phosphatases (class I Cys-based) (subclass III) (SAC domain) (5) (CxxxxxR)				
102	SACM1L/22908	SAC1	NCMDCLDR	PtdIns(4) P
103	INPP5F/22876	SAC2		PtdIns(4,5) P_2 , PtdIns(3,4,5) P_3 (D5 position)
104	FIG 4/9896	SAC3	NCVDCLDR	PtdIns(3,5) P_3 (D5 position)
105	SYNJ1/8867	Synaptojanin 1, INPP5G	SCERAGTR	PtdIns(3) P , PtdIns(4) P , PtdIns(3,5) P_2
106	SYNJ2/8871	Synaptojanin 2, INPP5H	SCERTGTR	
Paladin (class I Cys-based) (subclass IV) (PTP-like phytase) ⁿ (1) (CxxGxxR)				
107	PALD1/27143	Paladin	SCQMGVGR	? = unknown substrate
			SCLSGQGR ⁿ	
INPP4 inositol polyphosphate phosphatases 4' (class I Cys-based) (subclass V) (<i>L. pneumophila</i> SidP-like; <i>M. tuberculosis</i> PtpB-like; INPP4 domain) (2) (CxxxxxR)				
108	INPP4A/3631	INPP4A	SCKSAKDR	PtdIns(3,4) P_2 (D4 position), PtdIns(3,4,5) P_3
109	INPP4B/8821	INPP4B	CCKSAKDR	
TMEM55 inositol polyphosphate phosphatases 4' (class I Cys-based) (subclass VI) (BopB-like; TMEM55 domain) (2) (CxxxxxR)				
110	TMEM55A/55529	TMEM55A	ICKDTSRR	PtdIns(4,5) P_2 (D4 position)
111	TMEM55B/90809	TMEM55B	ICKVTSQR	
LMW-PTP (class II Cys-based) ^a (arsenate reductase domain) (1) (CxxxxxR)				
112	ACPI/52	LMW-PTP, red cell phosphatase	VCLGNICR	pTyr
SSU72 (class II Cys-based) ^o (arsenate reductase domain) (1)				
113	SSU72/29101	SSU72	VCSSNQNR	pSer
CDC25 (class III Cys-based) ^a (rhodanese domain) (3) (CxxxxxR)				
114	CDC25A/993	CDC25A	HCEFSSER	pThr, pTyr
115	CDC25B/994	CDC25B		
116	CDC25C/995	CDC25C		
HAD-EYAs (Asp-based) ^a [Rossmannoid (ED) domain] (4)				

117	EYA1/2138	EYA1	WDLDET	pSer, pTyr
118	EYA2/2139	EYA2		
119	EYA3/2140	EYA3		
120	EYA4/2070	EYA4		
UBASH3 HIS-PGAM phosphatases (TULAs) (branch 1 His-based) ^p (HP domain) (2)				
121	UBASH3A/53347	TULA, Sts-2	RHGE	pTyr
122	UBASH3B/84959	TULA-2, Sts-1		
ACP HIS-acid phosphatases (branch 2 His-based) ^p (HP domain) (3)				
123	ACPP/55	PAP, PAcP	RHGD	Nonproteinaceous, pSer, pThr, pTyr
124	ACP2/53 ^q	LAP		? = unknown substrate
125	ACPT/93650	ACPT		Nonproteinaceous, pTyr

^a Classification according to Alonso *et al.*⁴. ^b In the case of RPTPs, the catalytic signature motif for the PTP D1 domain is provided. ^c Note that some PTPs documented as inactive phosphatases contain all essential residues in their catalytic signature motif, making it possible that these enzymes are active towards unknown substrates. ^d PTPRU is named PTP λ in chicken. ^e PTPN20A (NM_001042387.1 \rightarrow NP_001035846.1) and PTPN20B (NM_001042357.3 \rightarrow NP_001035816.1) are two duplicated identical genes. ^f Note that the catalytic signature motif is not conserved in the catalytically inactive pseudophosphatases. ^g DUSP13A (NM_001007271.1 \rightarrow NP_001007272.1) and DUSP13B (NM_016364.3 \rightarrow NP_057448.3) are different genes located in the same locus. ^h DUPD1 (NM_001003892.1 \rightarrow NP_001003892.1) is a small-size active phosphatase. ⁱ DUSP27 (NM_001080426.1 \rightarrow NP_001073895.1) is a large-size inactive phosphatase. ^j Absent in mouse. ^k Note that TENC1 has been documented as an active PTP, although it lacks the conserved Arg in the catalytic signature motif. ^l Auxilins are included as PTPome members based on the presence of a PTEN-like structural domain. ^m GAK (cyclin G-associated kinase) contains an active Ser/Thr kinase domain. ⁿ PALD1 contains two CxxxxR motifs within two domains annotated as PTP-like_{phytase/DSPc} superfamily. PALD1 has been proposed to be catalytically inactive, although the evidence is indirect. ^o Classification based on structural similarity to ACP1 [124]. ^p Classification according to Rigden. ^q ACP2 is included as a PTPome member on the basis of its similarity to ACPP and ACPT. Adapted from Alonso and Pulido, 2016¹⁴.

9.2 Table: Summary of Reactome pathway participation of PDZ-containing proteins

Reactome pathway group	Number of proteins ^a	Most enriched protein(s) ^b	Number of pathways ^c
Cell cycle	3	PSMD9	13
Cell–cell communication	8	–	1
Cellular responses to external stimuli	1	PSMD9	1
DNA replication	1	PSMD9	3
Developmental biology	12	LIMK1	3
Disease: diseases of signal transduction	4	CNKSR1, CNKSR2	5
Disease: disorders of transmembrane transporters	1	PSMD9	1
Disease: infectious disease	2	PSMD9	2
Extracellular matrix organization	2	–	1
Gene expression (transcription): RNA polymerase II transcription	2	PSMD9	3
Hemostasis	2	–	1
Immune system: adaptive immune system	5	PSMD9	5
Immune system: cytokine signaling in immune system	4	PSMD9	3
Immune system: innate immune system	9	PSMD9	3
Metabolism of RNA	1	PSMD9	1
Metabolism of proteins: post-translational protein modification	3	PSMD9	3
Metabolism: metabolism of lipids	2	–	1
Metabolism: metabolism of polyamines	1	PSMD9	1
Metabolism: metabolism of water-soluble vitamins and cofactors	1	PDZD11	2
Muscle contraction: ion homeostasis	1	NOS1	1
Neuronal system: neurexins and neuroligins	15	–	1
Neuronal system: neurotransmitter receptors and postsynaptic signal transmission	6	DLG4	5
Neuronal system: neurotransmitter release cycle	6	RIMS1	6
Neuronal system: synaptic adhesion-like molecules	3	–	1
Programmed cell death: apoptosis	3	–	1
Signal transduction: death receptor signalling	5	–	1
Signal transduction: GPCR downstream	7	–	1

Reactome pathway group	Number of proteins ^a	Most enriched protein(s) ^b	Number of pathways ^c
signalling			
Signal transduction: intracellular signaling by second messengers	2	–	1
Signal transduction: MAPK family signaling cascades	5	PSMD9	2
Signal transduction: signaling by Hedgehog	2	PSMD9	5
Signal transduction: signaling by Hippo	3	–	1
Signal transduction: signaling by receptor tyrosine kinases	3	ERBIN	2
Signal transduction: signaling by Rho GTPases	18	LIMK1	2
Signal transduction: signaling by WNT	6	DVL2	8
Signal transduction: TGF- β receptor signaling in EMT (epithelial to mesenchymal transition)	2	–	1

^a Total number of PDZ-containing proteins participating in particular pathway group; ^b PDZ-containing protein known to participate in most of the pathways in particular pathway group; ^c The number of pathways in the pathway group the protein is involved in. Adapted from Christensen *et al.* 2019³⁶²

9.3 Table: Main functions and molecular targets of E6 and E7 HPV oncoproteins

High-risk viral protein	Cellular location	Function	Main targets	Effect on target expression
E6	Nucleus and cytoplasm	Escaping cell death	P53 protein	↓
			Procaspace 8 protein	↓
			Bak protein	↓
			TNR1	×
			Fas/Fas ligand death pathway	×
			NF-κB; cIAP-2	↑
		Deregulation of cell cycle	P300/CBP complex protein	↓
			miR34a	↓
		Immune system modulation	IRF3	↓
			IFNα	↓
			IFNκ	↓
		Cell immortalization	NFX1-91	↓
			Myc	↑
			Sp1	↑
		Genomic instability	APOBEC3	↑
			XRCC1	↓
		Cell invasion	Dlg	↓
			SCRIB	↓
			MAGI-1, MAGI-2 and MAGI-3	↓
			PAR3	↓
Fibulin-1	↓			
miR-23b	↓			
Paxillin disruption	↓			
E7	Nucleus	Deregulation of cell cycle	pRB protein	↓
			p107/p130	↓
			p21 p27	↓
			Claspin	↓
			E2F6	↓
		Immune system modulation	TLR9	×
			Cgas-STING	×
		Cell invasion	MMP-9	↑
		Genomic instability	Abnormal centrosome synthesis	×
			γ-Tubulin	×

			CDK2	↑
E6/E7	Nucleus and cytoplasm	Deregulation of cellular energetics	Aerobic glycolysis	↑
			mTORC1	↑
			GLUT-1	↑
		Genomic instability: epigenetic deregulation	DNMT1	↑
			DNMT3A	
			DNMT3B	
			E-cadherin	↓
			CXCL14	↓
			CCNA1	↓
			CBP/p300	↓
		Inflammation promotion	TIP60	↓
			ADA3	↓
		Inflammation promotion	IL-6	↑
			IL18	↑
		Angiogenic switch	Maspin	↓
			Thrombospondin-1	↓
VEGF	↑			
IL-8	↑			
Cell Invasion	TIM-2	↓		
	RECK	↓		

Adapted from Estêvão *et al.* 2019²⁶⁷.

9.4 Table: Data collection and refinement statistics for structures in section 6.2

	PTPN3-PDZ _{Next} complex			
	PBM-HBc	PBM-16E6	PBM-18E6	PBM-TACE
Data collection				
Space group	P 31 2 1	P 1 21 1	P 61 2 2	P 31 2 1
Unit cell (a, b, c) (Å)	75.58, 75.58, 46.59	46.62, 77.43, 130.03	82.34, 82.29, 139.35	76.79, 76,79, 42,26
a, b, g (°)	90.00, 90.00, 120.00	90.00, 90.14, 90.00	90.08, 89.92, 60.02	90.00, 90.00, 120.00
Resolution (Å)	46.59-1.86 (1.96-1.86)	46.62-2.19 (2.23-2.19)	41.12-1.88 (1.99-1.88)	66.58-1.70 (1.80-1.70)
No. of reflections (total/unique)	67510/13198	324034/47024	291781/23310	244786/17573
Redundancy	5.1 (5.0)	6.9 (6.5)	12.5 (12.7)	13.9 (13.7)
Completeness (%)	99.4 (97.8)	99.1 (95.7)	98.8 (95.8)	99.4 (98.4)
<i>I</i> /σ(<i>I</i>)	17.38 (2.70)	8.17 (1.11)	18.12 (2.24)	12.51 (1.77)
<i>R</i> _{meas} [¶] (%)	5.6 (57.5)	0.14 (1.71)	9.4 (126.7)	10.8 (131.1)
CC(1/2)	99.9 (97.7)	99.6 (72.6)	99.9 (90.8)	99.7 (91.8)
Refinement				
Resolution (Å)	29.35-1.86	43.34-2.19	35.65-2.1	66.58-1.70
No. reflections	13066	46869	16822	16687
<i>R</i> _{work} [†] / <i>R</i> _{free} [‡]	0.197/0.236	0.195/0.245	0.192/0.227	0.1676/0.2026
No. of protein atoms/ligand atoms	749/63	4402/350	1471/156	737/46
No. of solvent/hetero- atoms	68/1	235/9	69/1	21/2
Rmsd bond lengths (Å)	0.007	0.008	0.007	0.029
Rmsd bond angles (°)	0.836	0.949	0.815	2.648
<i>Wilson B</i> -factors	28.6	47.6	32.91	42.01
Ramachandran plot (favoured/outliers)*	96.9/0.0	97.4/0.4	98.97/0.00	96.84/0.00
PDB code	6T36	6HKS	YYYY	ZZZZ

Values in parenthesis correspond to the highest resolution bin. [¶] $R_{\text{meas}} = \sum_h (\frac{n_h}{\Omega_h} - 1)^{1/2} \sum_i \langle |I_h| - I_{h,i} \rangle / \sum_h \sum_i I_{h,i}$, where *n* denotes multiplicity. [†] $R_{\text{work}} = \sum_h ||\text{Fobs}_h| - |\text{Fcalc}_h|| / \sum_h |\text{Fobs}_h|$, where *Fobs_h* and *Fcalc_h* are the observed and calculated structure factors, respectively. No *I*/σ cut-off was applied. [‡] *R*_{free} is the R value obtained for a test set of reflections consisting of a randomly selected 5% subset of the data set excluded from refinement. * Categories were defined by MolProbity.

9.5 Table: PTPN3-PDZ partners fished by pull-down from HeLaS3 cell lysates

PDB ID	Protein name	PBM	Class
P48960	CD97 antigen	SESGI	I
P12814	Alpha-actinin-1	GESDL	I
P11169	Solute carrier family 2, facilitated glucose transporter member 3	TTTNV	I
O96008	Mitochondrial import receptor subunit TOM40 homolog	GLTIG	I
Q8TBC3	SH3KBP1-binding protein 1	NETSF	I
Q96TA1	Niban-like protein 1	VQTEF	I
P81605	Dermeidin	LDSVL	I
P53985	Monocarboxylate transporter 1	EESPV	I
O00592	Podocalyxin	EDTHL	I
Q13151	Heterogeneous nuclear ribonucleoprotein A0	GGSSF	I
P62701	40S ribosomal protein S4, X isoform	KQSSG	I
P26599	Polypyrimidine tract-binding protein 1	SKSTI	I
P27816	Microtubule-associated protein 4	QETSI	I
P02545	Prelamin-A/C	NCSIM	I
P55060	Exportin-2	SVTLL	I
P23219	Prostaglandin G/H synthase 1	PSTEL	I
Q9HB71	Calcyclin-binding protein	GDTEF	I
Q9BSE5	Agmatinase, mitochondrial	KVTTV	I
O15427	Monocarboxylate transporter 4	PETSV	I
P35354	Prostaglandin G/H synthase 2	RSTEL	I
O95336	6-phosphogluconolactonase	KHSTL	I
P49189	4-trimethylaminobutyraldehyde dehydrogenase	VESAF	I
Q9Y6M5	Zinc transporter 1	PESSL	I
P20020	Plasma membrane calcium-transporting ATPase 1	LETSL	I
Q9Y6M7	Sodium bicarbonate cotransporter 3	AETSL	I
P52569	Cationic amino acid transporter 2	KTSEF	I
Q02543	60S ribosomal protein L18a	PNTFF	I
P62829	60S ribosomal protein L23	AGSIA	I
O43707	Alpha-actinin-4	GESDL	I
Q96T17	MAP7 domain-containing protein 2	LNTFC	I
Q8TAA9	Vang-like protein 1	SETSV	I
Q15758	Neutral amino acid transporter B(0)	KESVM	I
Q15046	Lysine--tRNA ligase	VGTSV	I
Q96BN8	Ubiquitin thioesterase otulin	EETSL	I
Q5JVF3	PCI domain-containing protein 2	LSTVC	I
P09543	2',3'-cyclic-nucleotide 3'-phosphodiesterase	SCTII	I
O14734	Acyl-coenzyme A thioesterase 8	SESKL	I
P28331	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	EPSIC	I
P23634	Plasma membrane calcium-transporting ATPase 4	LETSV	I
Q9Y2T3	Guanine deaminase	FSSSV	I

Q13765	Nascent polypeptide-associated complex subunit alpha	MELTM	II
Q7Z2K6	Endoplasmic reticulum metallopeptidase 1	DLFVF	II
Q9UBM7	7-dehydrocholesterol reductase	LPGIF	II
Q16718	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5	WKWPI	II
Q01813	ATP-dependent 6-phosphofructokinase, platelet type	QPWSV	II
P62750	60S ribosomal protein L23a	KIGII	II
P21796	Voltage-dependent anion-selective channel protein 1	LEFQA	II
Q12846	Syntaxin-4	VTVVG	II
Q14165	Malectin	CLCRL	II
Q9P0L0	Vesicle-associated membrane protein-associated protein A	GKFIL	II
Q9P2X0	Dolichol-phosphate mannosyltransferase subunit 3	RGLRF	II
P45880	Voltage-dependent anion-selective channel protein 2	LELEA	II
P47897	Glutamine--tRNA ligase	DPGKV	II
P25705	ATP synthase subunit alpha, mitochondrial	AGFEA	II
Q29963	HLA class I histocompatibility antigen, Cw-6 alpha chain	IACKA	II
Q9H0U4	Ras-related protein Rab-1B	GGGCC	II
P62081	40S ribosomal protein S7	PEFQL	II
P08237	ATP-dependent 6-phosphofructokinase, muscle type	GEAAV	II
P10809	60 kDa heat shock protein, mitochondrial	GGGMF	II
O43399	Tumor protein D54	DPAPF	II
Q9NRW1	Ras-related protein Rab-6B	GGCSC	II
P11488	Guanine nucleotide-binding protein G(t) subunit alpha-1	DCGLF	II
Q15165	Serum paraoxonase/arylesterase 2	LYCEL	II
O75396	Vesicle-trafficking protein SEC22b	RFWWL	II
Q969Q0	60S ribosomal protein L36a-like	QVIQF	II
Q59GN2	Putative 60S ribosomal protein L39-like 5	TKLGL	II
P22234	Multifunctional protein ADE2	RECNL	II
P16070	CD44 antigen	MKIGV	II
P62277	40S ribosomal protein S13	SALVA	II
O95292	Vesicle-associated membrane protein-associated protein B/C	GKIAL	II
P61619	Protein transport protein Sec61 subunit alpha isoform 1	GALLF	II
Q9Y6C9	Mitochondrial carrier homolog 2	LKMLI	II
O43175	D-3-phosphoglycerate dehydrogenase	FQFHF	II
Q9NZ01	Very-long-chain enoyl-CoA reductase Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	IPFLL	II
P21912		KKASV	II
P60866	40S ribosomal protein S20	TIADA	II
P46778	60S ribosomal protein L21	YEFMA	II
Q13347	Eukaryotic translation initiation factor 3 subunit I	FEFEA	II
Q9NRP0	Oligosaccharyltransferase complex subunit OSTC	GylMG	II
P00403	Cytochrome c oxidase subunit 2	PVFTL	II
Q96P70	Importin-9	QTIGI	II
P27708	CAD protein	VLGRF	II
P41252	Isoleucine--tRNA ligase, cytoplasmic	TTADF	II
P54709	Sodium/potassium-transporting ATPase subunit beta-3	ITARA	II
Q05193	Dynamin-1	PPFDL	II

P14854	Cytochrome c oxidase subunit 6B1	FPGKI	II
Q9Y277	Voltage-dependent anion-selective channel protein 3	FELEA	II
Q9UFH2	Dynein heavy chain 17, axonemal	LLLQV	II
Q5JTV8	Torsin-1A-interacting protein 1	RGICL	II
P31930	Cytochrome b-c1 complex subunit 1, mitochondrial	FWLRF	II
O14975	Very long-chain acyl-CoA synthetase	KTLKL	II
P01834	Immunoglobulin kappa constant	NRGEC	II
P62805	Histone H4	YGFGG	II
Q02978	Mitochondrial 2-oxoglutarate/malate carrier protein	LFLSG	II
Q96HY6	DDRGK domain-containing protein 1	AQAPA	II
Q9BUF5	Tubulin beta-6 chain	EEIDG	II
P00338	L-lactate dehydrogenase A chain	KELQF	II
O75832	26S proteasome non-ATPase regulatory subunit 10	RMVEG	II
P06703	Protein S100-A6	EALKG	II
O00303	Eukaryotic translation initiation factor 3 subunit F	KLVNL	II
P01891	HLA class I histocompatibility antigen, A-68 alpha chain	TACKV	II
P51571	Translocon-associated protein subunit delta	SHIQA	II
O15260	Surfeit locus protein 4	KKKEW	III
P63261	Actin, cytoplasmic 2	HRKCF	III
P06748	Nucleophosmin	WRKSL	III
P26640	Valine--tRNA ligase	FQKML	III
P12235	ADP/ATP translocase 1	IKKYV	III
P12236	ADP/ATP translocase 3	LKKVI	III
P49327	Fatty acid synthase	SVREG	III
P31327	Carbamoyl-phosphate synthase [ammonia], mitochondrial	AGKAA	III
O14684	Prostaglandin E synthase	AARHL	III
Q07812	Apoptosis regulator BAX	WKKMG	III
Q9H9B4	Sideroflexin-1	FNKGL	III
O75955	Flotillin-1	PLRTA	III
Q71UI9	Histone H2A.V	QQKTA	III
P49257	Protein ERGIC-53	AKKFF	III
P62424	60S ribosomal protein L7a	ATKLG	III
P27105	Erythrocyte band 7 integral membrane protein	HSHLG	III
Q9H845	Acyl-CoA dehydrogenase family member 9, mitochondrial	LDRTC	III
Q7L014	Probable ATP-dependent RNA helicase DDX46	RYKVL	III
P61978	Heterogeneous nuclear ribonucleoprotein K	SGKFF	III
Q9Y6M9	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9	RERPM	III
P09651	Heterogeneous nuclear ribonucleoprotein A1	SGRRF	III
P62263	40S ribosomal protein S14	RGRRL	III
P61026	Ras-related protein Rab-10	KSKCC	III
Q9NX63	MICOS complex subunit MIC19	LEKGG	III
P20674	Cytochrome c oxidase subunit 5A, mitochondrial	GLDKV	IV
P48047	ATP synthase subunit O, mitochondrial	MREIV	IV

10 Index of figures

Figure 1.1: Classification of the Tyr phosphatase superfamily.....	10
Figure 1.2: Structure of a classical class I PTP catalytic domain bound to tungstate.....	12
Figure 1.3: Scheme of the catalytic mechanism of classic phosphatases.....	13
Figure 1.4: The pTyr-loop.....	14
Figure 1.5: Classification of classical PTPs based on their subcellular localization.....	15
Figure 1.6: Modular architecture of PTPN3.....	19
Figure 1.7: PTPN3 phosphorylation sites.....	20
Figure 1.8: Structural organization of the FERM domain.....	22
Figure 1.9: Subcellular location of PTPN3 and PTPN4 in Jurkat T cells.....	22
Figure 1.10: Comparison of the structures of the unbound, wild type PTPN3-PTP and its substrate-trapping mutant.....	24
Figure 2.1: Domain organizations of a selected set of PDZ domain proteins.....	35
Figure 2.2: Organization of the INAD signaling complex.....	36
Figure 2.3: Protein complex organization in the postsynaptic density.....	38
Figure 2.4: Structures of PDZ domains and their interaction with peptide ligands.....	39
Figure 2.5: Recognition of position -2 by different subclasses of PDZ domains.....	42
Figure 2.6: Specific interactions determine selectivity in positions -1 and -3 for ZO1 PDZ1 and Erbin PDZ.....	43
Figure 2.7: PDZ/PBM interactions beyond position -3 contribute to binding affinity.....	45
Figure 2.8: Structural Comparison of Par-6 PDZ internal and C-terminal binding.....	46
Figure 2.9: The structure of the nNOS PDZ–syntrophin PDZ heterodimer.....	47
Figure 2.10: two PDZ dimerization mechanisms.....	49
Figure 2.11: Schematic diagrams of the distinct target-binding properties conferred by the supramodular organization of GRIP1 and ZO-1 PDZ domains.....	50
Figure 2.12: PDZ-phospholipid interactions.....	52
Figure 2.13: Regulation of PDZ–target interactions by phosphorylation and auto-inhibition.....	54
Figure 2.14: Allosteric conformational regulation of PDZ domains.....	55
Figure 3.1: Cellular PDZ proteins that function in cell polarity establishment and cell junction formation are commonly targeted by viruses.....	58
Figure 3.2: Model of infection by a virulent or an attenuated strain of RABV and associated proteins.....	62
Figure 3.3: The essential steps in the HPV life cycle.....	64
Figure 3.4 the HPV life cycle is coupled to the differentiation program of the host epithelia.....	65
Figure 3.5 Phylogenetic tree of 118 papillomavirus types.....	66
Figure 3.6: Structure of the high risk HPV E6 protein.....	69
Figure 3.7: The HBV life cycle.....	75
Figure 3.8: HBc protein domain organization, phosphorylation sites and sequence features.....	76
Figure 3.9: X-ray structures of the HBc assembly domain.....	77
Figure 3.10: Direct observation of the HBc CTD segment protruding outside the core.....	78
Figure 5.1: Schematic representation of the PTPN3 constructs.....	95
Figure 5.2: Regulation of the phosphatase activity of PTPN3 by its PDZ domain.....	97
Figure 5.3: Sequence alignment of the PDZ-PTP linker of PTPN3 and PTPN4 homologs ..	101
Figure 5.4. Secondary structure prediction for the PTPN3 PDZ-PTP linker.....	101
Figure 5.5: Sensitivity of PTPN3-Bidomain to proteolysis.....	102
Figure 5.6: ¹ H- ¹⁵ N TROSY spectra of PTPN3-Bidomain.....	103
Figure 6.1: Sequence alignment of the PDZ domains of PTPN3 PTPN4.....	106

Figure 6.2: Detail of the bonding network of PTPN3-PDZ and the PBM peptides of HPV16 E6, HPV 18 E6, HBc and TACE	108
Figure 6.3: Sequence conservation of the PDZome	109
Figure 6.4: Sequence alignment of PTPN3 and PTPN4 homologs.....	110
Figure 6.5: Structures of PTPN4-PDZ Bound to Cyt13-att and GluN2A-16	111
Figure 6.6: Surface hydrophobicity of PTPN3-PDZ.....	113
Figure 6.7: Crystal structure of the PDZ6 domain of GRIP bound to the C-terminal peptide of liprin- α	114
Figure 6.8 Frequency plot of the C-terminal residues of the class I and class II PTPN3-PDZ partners identified by pull-down.	115

11 Index of tables

Table 1.1: Protein tyrosine phosphatases involved in lymphocyte activation	9
Table 1.2: Protein tyrosine phosphatases involved in cancer.....	17
Table 1.3: PTPN3 phosphorylation sites.....	20
Table 1.4: Implications of PTPN3 in cancer	33
Table 3.1: Known cellular PDZ protein targets of viral proteins.....	57
Table 3.2: Known PDZ domain-containing targets of E6.....	68
Table 5.1: kinetic parameters of hydrolysis of pNPP by PTPN3 constructs.....	97
Table 5.2: Affinities of the synthetic PBM peptides.....	98
Table 6.1: Parameters of the crystal structures of PTPN3-PDZ-peptide complexes	106
Table 6.2: PTPN3-PDZ Class I PBM-containing partners	116

12 List of abbreviations

AdV	Adenovirus
AJ	Adherens junction
AMPAR	α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptors
ANK	Ankyrin repeats
CAL	Cystic fibrosis transmembrane regulator-associated ligand
CaM	Calmodulin
CAM	Core allosteric modulators
CaMK	Calmodulin-dependent kinase
cccDNA	Covalently closed circular DNA
CFTR	Cystic fibrosis transmembrane conductance regulator
CRIP1	Cysteine-rich PDZ-binding protein
CTD	C-terminal domain
DLG	Discs large homolog
DSPs	Dual specificity phosphatases
E6AP	Ubiquitin ligase E6 associated protein
EGFR	Epidermal growth factor receptor
ER	Estrogen receptor α
GHR	Growth hormone receptor
GK	Guanylate kinase
GKAP	Guanylate kinase-associated protein
GLGF	Gly-Leu-Gly-Phe
GluR2	Glutamate receptor 2
GRIP	Glutamate receptor interacting protein
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HFKs	Human foreskin keratinocytes
HPV	Human papillomavirus
HTLV-1	Human T-lymphotropic virus type 1
JAK	Janus kinase
LMW-PTP	Low molecular weight protein tyrosine phosphatase
MAGI3	Membrane-associated guanylate kinase-related 3
MAGUK	Membrane associated guanylate kinases
MAPK	Mitogen-activated protein kinase
MAST2	Microtubule-associated serine/threonine-protein kinase 2
MUPP1	Multiple PDZ-containing protein 1
Nav1.5	Voltage-gated sodium channels
NavCTD	Voltage-gated sodium channels C-terminal domain
NHERF1	Na ⁺ /H ⁺ Exchanger Regulatory Factor
NLS	Nuclear localization signals
NMDA	N-methyl-D-aspartate
NSCLC	Non-small cell lung cancer
NTCP	Sodium Taurocholate Co-transporter Polypeptide
NTPTs	Non transmembrane protein tyrosine phosphatases
ORF	Open reading frame
p-Ser	Phospho-serine
p-Thr	Phospho-threonine
p-Tyr	Phospho-tyrosine

Par	Partitioning defective
PBM	PDZ-binding motif
PDZ2	PDZ domain 2
pgRNA	Pre-genomic RNA
PH	Pleckstrin homology
PI3K	Phosphatidylinositol-3-kinase
PICK1	Protein interacting with C kinase 1
PIP	Phospho-phosphatidylinositol
PKC	Protein kinase C
PLC	Phospholipase C
pRb	Retinoblastoma protein
PSD-95	Postsynaptic density protein 95
PTEN2	Phosphatase and tensin homolog deleted on chromosome 10
PTKs	Protein tyrosine kinases
PTPN	Protein tyrosine phosphatase non-receptor
PTPs	Protein tyrosine phosphatases
RABV	Rabies virus
rcDNA	Relaxed circular DNA
ROS	Reactive-oxygen species
RPTPs	Receptor-like protein tyrosine phosphatases
SAM	sterile-alpha motif
SH	Src homology
SH3	Src-homology domain 3
STAT	Signal transducer and activator of transcription
TACE/ADAM17	Tumor necrosis factor α -convertase
TAM	Tamoxifen
TCR	T cell receptor
tER	Transitional endoplasmic reticulum
TGF β	Tumor growth factor β
TJ	Tight junction
TKIs	Tyrosine kinase inhibitors
TNBC	Triple-negative breast cancers
TNF- α	Tumor necrosis factor α
TRP	Transient receptor potential channel
T β RI	TGF β type I receptor
VCP	Valosin-containing protein
ZO-1	Zonula occludens 1

Bibliography

1. Cohen, P. The role of protein phosphorylation in human health and disease. *Eur. J. Biochem.* **268**, 5001–5010 (2001).
2. Hunter, T. The role of tyrosine phosphorylation in cell growth and disease. *Harvey Lect.* **94**, 81–119
3. Larsen, M., Tremblay, M. L. & Yamada, K. M. Phosphatases in cell–matrix adhesion and migration. *Nat. Rev. Mol. Cell Biol.* **4**, 700–711 (2003).
4. Alonso, A. *et al.* Protein Tyrosine Phosphatases in the Human Genome. *Cell* **117**, 699–711 (2004).
5. Mustelin, T., Vang, T. & Bottini, N. Protein tyrosine phosphatases and the immune response. *Nat. Rev. Immunol.* **5**, 43–57 (2005).
6. Tonks, N. K. Protein tyrosine phosphatases: from genes, to function, to disease. *Nat Rev Mol Cell Biol* **7**, 833–846 (2006).
7. Hallé, M., Tremblay, M. L. & Meng, T.-C. Protein Tyrosine Phosphatases: Emerging Regulators of Apoptosis. *Cell Cycle* **6**, 2773–2781 (2007).
8. Pao, L. I., Badour, K., Siminovitch, K. A. & Neel, B. G. Nonreceptor Protein-Tyrosine Phosphatases in Immune Cell Signaling. *Annu. Rev. Immunol.* **25**, 473–523 (2007).
9. Hunter, T. Tyrosine phosphorylation: thirty years and counting. *Curr. Opin. Cell Biol.* **21**, 140 (2009).
10. Rhee, I. & Veillette, A. Protein tyrosine phosphatases in lymphocyte activation and autoimmunity. *Nat. Immunol.* **13**, 439–447 (2012).
11. Delves, P. J., Martin, S. J., Burton, D. R. & Roitt, I. M. (Ivan M. Roitt's essential immunology.
12. Alonso, A. & Pulido, R. CORRECTIONS - The extended human PTPome: a growing tyrosine phosphatase family. *FEBS J.* **283**, 2197–2201 (2016).
13. Alonso, A., Nunes-Xavier, C. E., Bayón, Y. & Pulido, R. The Extended Family of Protein Tyrosine Phosphatases. in 1–23 (Humana Press, New York, NY, 2016). doi:10.1007/978-1-4939-3746-2_1
14. Alonso, A. & Pulido, R. The extended human PTPome: a growing tyrosine phosphatase family. *FEBS J.* **283**, 1404–1429 (2016).
15. Zhang, Z. Y. & Dixon, J. E. Active site labeling of the Yersinia protein tyrosine phosphatase: The determination of the pKa of the active site cysteine and the function of the conserved histidine 402. *Biochemistry* **32**, 9340–9345 (1993).
16. Tautz, L., Critton, D. A. & Grotegut, S. Protein Tyrosine Phosphatases: Structure, Function, and Implication in Human Disease. in 179–221 (Humana Press, Totowa, NJ, 2013). doi:10.1007/978-1-62703-562-0_13
17. Böhmer, F., Szedlaczek, S., Taberner, L., Östman, A. & den Hertog, J. Protein tyrosine phosphatase structure-function relationships in regulation and pathogenesis. *FEBS J.* **280**, 413–431 (2013).
18. Streuli, M., Krueger, N. X., Thai, T., Tang, M. & Saito, H. Distinct functional roles of the two intracellular phosphatase like domains of the receptor-linked protein tyrosine phosphatases LCA and LAR. *EMBO J.* **9**, 2399–407 (1990).
19. Felberg, J. & Johnson, P. Characterization of recombinant CD45 cytoplasmic domain proteins. Evidence for intramolecular and intermolecular interactions. *J. Biol. Chem.* **273**, 17839–45 (1998).
20. Toledano-Katchalski, H. *et al.* Dimerization in vivo and inhibition of the nonreceptor form of protein tyrosine phosphatase epsilon. *Mol. Cell. Biol.* **23**, 5460–71 (2003).
21. Brady-Kalnay, S. M. & Tonks, N. K. Protein tyrosine phosphatases as adhesion receptors. *Curr. Opin. Cell Biol.* **7**, 650–657 (1995).
22. Andersen, J. N. *et al.* A genomic perspective on protein tyrosine phosphatases: gene structure,

- pseudogenes, and genetic disease linkage. *FASEB J.* **18**, 8–30 (2004).
23. Elson, A. & Leder, P. Identification of a cytoplasmic, phorbol ester-inducible isoform of protein tyrosine phosphatase epsilon. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 12235–9 (1995).
 24. Si, X., Zeng, Q., Ng, C. H., Hong, W. & Pallen, C. J. Interaction of farnesylated PRL-2, a protein-tyrosine phosphatase, with the beta-subunit of geranylgeranyltransferase II. *J. Biol. Chem.* **276**, 32875–82 (2001).
 25. Tautz, L., Pellecchia, M. & Mustelin, T. Targeting the PTPome in human disease. *Expert Opin. Ther. Targets* **10**, 157–177 (2006).
 26. Cantley, L. C. & Neel, B. G. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc. Natl. Acad. Sci.* **96**, 4240–4245 (1999).
 27. Guldberg, P. *et al.* Disruption of the MMAC1/PTEN gene by deletion or mutation is a frequent event in malignant melanoma. *Cancer Res.* **57**, 3660–3 (1997).
 28. Perren, A. *et al.* Immunohistochemical Evidence of Loss of PTEN Expression in Primary Ductal Adenocarcinomas of the Breast. *Am. J. Pathol.* **155**, 1253–1260 (1999).
 29. Reifenger, J. *et al.* Allelic losses on chromosome arm 10q and mutation of the PTEN (MMAC1) tumour suppressor gene in primary and metastatic malignant melanomas. *Virchows Arch.* **436**, 487–93 (2000).
 30. Keniry, M. & Parsons, R. The role of PTEN signaling perturbations in cancer and in targeted therapy. *Oncogene* **27**, 5477–5485 (2008).
 31. Wang, Z. *et al.* Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* **300**, 949 (2003).
 32. van Puijenbroek, M. *et al.* Mass Spectrometry-Based Loss of Heterozygosity Analysis of Single-Nucleotide Polymorphism Loci in Paraffin Embedded Tumors Using the MassEXTEND Assay. *J. Mol. Diagnostics* **7**, 623–630 (2005).
 33. Ruivenkamp, C. *et al.* LOH of PTPRJ occurs early in colorectal cancer and is associated with chromosomal loss of 18q12–21. *Oncogene* **22**, 3472–3474 (2003).
 34. Mena-Duran, A. V. *et al.* SHP1 expression in bone marrow biopsies of myelodysplastic syndrome patients: a new prognostic factor. *Br. J. Haematol.* **129**, 791–794 (2005).
 35. Zhang, Q. *et al.* STAT3- and DNA methyltransferase 1-mediated epigenetic silencing of SHP-1 tyrosine phosphatase tumor suppressor gene in malignant T lymphocytes. *Proc. Natl. Acad. Sci.* **102**, 6948–6953 (2005).
 36. Sun, T. *et al.* Activation of Multiple Proto-oncogenic Tyrosine Kinases in Breast Cancer via Loss of the PTPN12 Phosphatase. *Cell* **144**, 703–718 (2011).
 37. Noguchi, T., Matozaki, T., Horita, K., Fujioka, Y. & Kasuga, M. Role of SH-PTP2, a protein-tyrosine phosphatase with Src homology 2 domains, in insulin-stimulated Ras activation. *Mol. Cell. Biol.* **14**, 6674–6682 (1994).
 38. Salmond, R. J. & Alexander, D. R. SHP2 forecast for the immune system: fog gradually clearing. *Trends Immunol.* **27**, 154–160 (2006).
 39. Tartaglia, M. *et al.* Genetic evidence for lineage-related and differentiation stage-related contribution of somatic PTPN11 mutations to leukemogenesis in childhood acute leukemia. *Blood* **104**, 307–13 (2004).
 40. Tartaglia, M. *et al.* Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat. Genet.* **34**, 148–150 (2003).
 41. Loh, M. L. *et al.* Mutations in PTPN11 implicate the SHP-2 phosphatase in leukemogenesis. *Blood* **103**, 2325–31 (2004).
 42. Bentires-Alj, M. & Neel, B. G. Protein-Tyrosine Phosphatase 1B Is Required for HER2/Neu-Induced Breast Cancer. *Cancer Res.* **67**, 2420–2424 (2007).

43. Saha, S. A Phosphatase Associated with Metastasis of Colorectal Cancer. *Science (80-.)*. **294**, 1343–1346 (2001).
44. Sur, S. & Agrawal, D. K. Phosphatases and kinases regulating CDC25 activity in the cell cycle: clinical implications of CDC25 overexpression and potential treatment strategies. *Mol. Cell. Biochem.* **416**, 33–46 (2016).
45. He, R.-J., Yu, Z.-H., Zhang, R.-Y. & Zhang, Z.-Y. Protein tyrosine phosphatases as potential therapeutic targets. *Acta Pharmacol. Sin.* **35**, 1227–46 (2014).
46. Loda, M. *et al.* Expression of mitogen-activated protein kinase phosphatase-1 in the early phases of human epithelial carcinogenesis. *Am. J. Pathol.* **149**, 1553–64 (1996).
47. Nikolaienko, R. M., Agyekum, B. & Bouyain, S. Receptor protein tyrosine phosphatases and cancer: new insights from structural biology. *Cell Adh. Migr.* **6**, 356–64 (2012).
48. Östman, A., Hellberg, C. & Böhmer, F. D. Protein-tyrosine phosphatases and cancer. *Nat. Rev. Cancer* **6**, 307–320 (2006).
49. Hendriks, W. J. A. J. *et al.* Protein tyrosine phosphatases in health and disease. *FEBS J.* **280**, 708–730 (2013).
50. Bottini, N. *et al.* A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat. Genet.* **36**, 337–338 (2004).
51. Begovich, A. B. *et al.* A Missense Single-Nucleotide Polymorphism in a Gene Encoding a Protein Tyrosine Phosphatase (PTPN22) Is Associated with Rheumatoid Arthritis. *Am. J. Hum. Genet.* **75**, 330–337 (2004).
52. Kyogoku, C. *et al.* Genetic Association of the R620W Polymorphism of Protein Tyrosine Phosphatase PTPN22 with Human SLE. *Am. J. Hum. Genet.* **75**, 504–507 (2004).
53. Goebel-Goody, S. M. *et al.* Therapeutic implications for striatal-enriched protein tyrosine phosphatase (STEP) in neuropsychiatric disorders. *Pharmacol. Rev.* **64**, 65–87 (2012).
54. Yang, Q. & Tonks, N. K. Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with homology to the cytoskeletal-associated proteins band 4.1, ezrin, and talin. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5949–53 (1991).
55. Itoh, F. *et al.* Expression and chromosomal assignment of PTPH1 gene encoding a cytosolic protein tyrosine phosphatase homologous to cytoskeletal-associated proteins. *Int. J. Cancer* **55**, 947–951 (1993).
56. Zhang, S. H., Eckberg, W. R., Yang, Q., Samatar, A. A. & Tonks, N. K. Biochemical characterization of a human band 4.1-related protein-tyrosine phosphatase, PTPH1. *Journal of Biological Chemistry* **270**, 20067–20072 (1995).
57. Gjørloff-Wingren, A. *et al.* Subcellular localization of intracellular protein tyrosine phosphatases in T cells. *Eur. J. Immunol.* **30**, 2412–2421 (2000).
58. Zhang, S.-H., Kobayashi, R., Graves, P. R., Piwnica-Worms, H. & Tonks, N. K. Serine Phosphorylation-dependent Association of the Band 4.1-related Protein-tyrosine Phosphatase PTPH1 with 14-3-3 β Protein. *J. Biol. Chem.* **272**, 27281–27287 (1997).
59. Hou, S. *et al.* p38 γ Mitogen-activated protein kinase signals through phosphorylating its phosphatase PTPH1 in regulating ras protein oncogenesis and stress response. *J. Biol. Chem.* **287**, 27895–905 (2012).
60. Ma, S. *et al.* Tyrosine dephosphorylation enhances the therapeutic target activity of epidermal growth factor receptor (EGFR) by disrupting its interaction with estrogen receptor (ER). *Oncotarget* **6**, 13320–13333 (2015).
61. Cau, Y., Valensin, D., Mori, M., Draghi, S. & Botta, M. Structure, Function, Involvement in Diseases and Targeting of 14-3-3 Proteins: An Update. *Curr. Med. Chem.* **25**, 5–21 (2018).
62. Chishti, A. H. *et al.* The FERM domain: a unique module involved in the linkage of cytoplasmic proteins to the membrane. *Trends Biochem. Sci.* **23**, 281–282 (1998).

63. Baines, A. J., Lu, H.-C. & Bennett, P. M. The Protein 4.1 family: Hub proteins in animals for organizing membrane proteins. *Biochim. Biophys. Acta - Biomembr.* **1838**, 605–619 (2014).
64. Patrignani, C. *et al.* Knockout mice reveal a role for protein tyrosine phosphatase H1 in cognition. *Behav Brain Funct* **4**, 1–14 (2008).
65. Gu, M. & Majerus, P. W. The properties of the protein tyrosine phosphatase PTPMEG. *J. Biol. Chem.* **271**, 27751–9 (1996).
66. Hou, S. *et al.* PTPH1 dephosphorylates and cooperates with p38 γ MAPK to increase Ras oncogenesis through PDZ-mediated interaction. *Cancer Res.* **148**, 895–901 (2010).
67. Töpffer, S., Müller-Schiffmann, A., Matentzoglou, K., Scheffner, M. & Steger, G. Protein tyrosine phosphatase H1 is a target of the E6 oncoprotein of high-risk genital human papillomaviruses. *J. Gen. Virol.* **88**, 2956–2965 (2007).
68. Hsu, E.-C. *et al.* Suppression of hepatitis B viral gene expression by protein-tyrosine phosphatase PTPN3. *J. Biomed. Sci.* **14**, 731–44 (2007).
69. Barr, A. J. *et al.* Large-scale structural analysis of the classical human protein tyrosine phosphatome. *Cell* **136**, 352–63 (2009).
70. Chen, K.-E. *et al.* Reciprocal allosteric regulation of p38 γ and PTPN3 involves a PDZ domain-modulated complex formation. *Sci. Signal.* **7**, ra98 (2014).
71. Pulido, R., Zúñiga, A. & Ullrich, A. PTP-SL and STEP protein tyrosine phosphatases regulate the activation of the extracellular signal-regulated kinases ERK1 and ERK2 by association through a kinase interaction motif. *EMBO J.* **17**, 7337–50 (1998).
72. Camps, M., Nichols, A. & Arkininstall, S. Dual specificity phosphatases: a gene family for control of MAP kinase function. *FASEB J.* **14**, 6–16 (2000).
73. Yu, Z.-H. & Zhang, Z.-Y. Regulatory Mechanisms and Novel Therapeutic Targeting Strategies for Protein Tyrosine Phosphatases. *Chem. Rev.* **118**, 1069–1091 (2018).
74. Karisch, R. *et al.* Global Proteomic Assessment of the Classical Protein-Tyrosine Phosphatome and “Redoxome”. *Cell* **146**, 826–840 (2011).
75. Aranda, E., López-Pedraza, C., De La Haba-Rodríguez, J. R. & Rodríguez-Ariza, A. Nitric oxide and cancer: the emerging role of S-nitrosylation. *Curr. Mol. Med.* **12**, 50–67 (2012).
76. Zhang, S. H., Liu, J., Kobayashi, R. & Tonks, N. K. Identification of the cell cycle regulator VCP (p97/CDC48) as a substrate of the band 4.1-related protein-tyrosine phosphatase PTPH1. *J. Biol. Chem.* **274**, 17806–12 (1999).
77. Bogdan, S. & Klämbt, C. Epidermal growth factor receptor signaling. *Curr. Biol.* **11**, R292–R295 (2001).
78. Tarcic, G. *et al.* An Unbiased Screen Identifies DEP-1 Tumor Suppressor as a Phosphatase Controlling EGFR Endocytosis. *Curr. Biol.* **19**, 1788–1798 (2009).
79. Yuan, T., Wang, Y., Zhao, Z. J. & Gu, H. Protein-tyrosine phosphatase PTPN9 negatively regulates ErbB2 and epidermal growth factor receptor signaling in breast cancer cells. *J. Biol. Chem.* **285**, 14861–70 (2010).
80. Xu, Y., Tan, L.-J., Grachtchouk, V., Voorhees, J. J. & Fisher, G. J. Receptor-type protein-tyrosine phosphatase-kappa regulates epidermal growth factor receptor function. *J. Biol. Chem.* **280**, 42694–700 (2005).
81. Tiganis, T. Protein Tyrosine Phosphatases: Dephosphorylating the Epidermal Growth Factor Receptor. *IUBMB Life (International Union Biochem. Mol. Biol. Life)* **53**, 3–14 (2002).
82. Eden, E. R., White, I. J., Tsapara, A. & Futter, C. E. Membrane contacts between endosomes and ER provide sites for PTP1B–epidermal growth factor receptor interaction. *Nat. Cell Biol.* **12**, 267–272 (2010).
83. Ali, N. *et al.* Recruitment of UBPY and ESCRT Exchange Drive HD-PTP-Dependent Sorting of EGFR to the MVB. *Curr. Biol.* **23**, 453–461 (2013).

84. Musgrove, E. A. & Sutherland, R. L. Biological determinants of endocrine resistance in breast cancer. *Nat. Rev. Cancer* **9**, 631–643 (2009).
85. Castoria, G. *et al.* Tyrosine phosphorylation of estradiol receptor by Src regulates its hormone-dependent nuclear export and cell cycle progression in breast cancer cells. *Oncogene* **31**, 4868–4877 (2012).
86. Sun, J., Zhou, W., Kaliappan, K., Nawaz, Z. & Slingerland, J. M. ER α Phosphorylation at Y537 by Src Triggers E6-AP-ER α Binding, ER α Ubiquitylation, Promoter Occupancy, and Target Gene Expression. *Mol. Endocrinol.* **26**, 1567–1577 (2012).
87. Skliris, G. P., Nugent, Z., Watson, P. H. & Murphy, L. C. Estrogen Receptor Alpha Phosphorylated at Tyrosine 537 is Associated with Poor Clinical Outcome in Breast Cancer Patients Treated with Tamoxifen. *Horm. Cancer* **1**, 215–221 (2010).
88. Massarweh, S. *et al.* Tamoxifen Resistance in Breast Tumors Is Driven by Growth Factor Receptor Signaling with Repression of Classic Estrogen Receptor Genomic Function. *Cancer Res.* **68**, 826–833 (2008).
89. Fan, P., Wang, J., Santen, R. J. & Yue, W. Long-term Treatment with Tamoxifen Facilitates Translocation of Estrogen Receptor α out of the Nucleus and Enhances its Interaction with EGFR in MCF-7 Breast Cancer Cells. *Cancer Res.* **67**, 1352–1360 (2007).
90. Suresh, P. S., Ma, S., Migliaccio, A. & Chen, G. Protein-Tyrosine Phosphatase H1 Increases Breast Cancer Sensitivity to Antiestrogens by Dephosphorylating Estrogen Receptor at Tyr537. *Mol. Cancer Ther.* **13**, 230–238 (2014).
91. Dawson, S. J., Provenzano, E. & Caldas, C. Triple negative breast cancers: Clinical and prognostic implications. *Eur. J. Cancer* **45**, 27–40 (2009).
92. Diluvio, G. *et al.* NOTCH3 inactivation increases triple negative breast cancer sensitivity to gefitinib by promoting EGFR tyrosine dephosphorylation and its intracellular arrest. *Oncogenesis* **7**, 42 (2018).
93. Li, M.-Y. *et al.* Protein tyrosine phosphatase PTPN3 inhibits lung cancer cell proliferation and migration by promoting EGFR endocytic degradation. *Oncogene* **34**, 3791–3803 (2015).
94. Yin, N. *et al.* The K-Ras effector p38 γ MAPK confers intrinsic resistance to tyrosine kinase inhibitors by stimulating EGFR transcription and EGFR dephosphorylation. *J. Biol. Chem.* **292**, 15070–15079 (2017).
95. Dehkhoda, F., Lee, C. M. M., Medina, J. & Brooks, A. J. The Growth Hormone Receptor: Mechanism of Receptor Activation, Cell Signaling, and Physiological Aspects. *Front. Endocrinol. (Lausanne)*. **9**, 35 (2018).
96. Pasquali, C. *et al.* Identification of Protein Tyrosine Phosphatases with Specificity for the Ligand-Activated Growth Hormone Receptor. *Mol. Endocrinol.* **17**, 2228–2239 (2003).
97. Pilecka, I. *et al.* Protein-tyrosine phosphatase H1 controls growth hormone receptor signaling and systemic growth. *J. Biol. Chem.* **282**, 35405–35415 (2007).
98. Jentsch, S. & Rumpf, S. Cdc48 (p97): a ‘molecular gearbox’ in the ubiquitin pathway? *Trends Biochem. Sci.* **32**, 6–11 (2007).
99. Meyer, H., Bug, M. & Bremer, S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat. Cell Biol.* **14**, 117–123 (2012).
100. Alexandru, G. *et al.* UBXD7 binds multiple ubiquitin ligases and implicates p97 in HIF1 α turnover. *Cell* **134**, 804–16 (2008).
101. Li, J.-M., Wu, H., Zhang, W., Blackburn, M. R. & Jin, J. The p97-UFD1L-NPL4 protein complex mediates cytokine-induced I κ B α proteolysis. *Mol. Cell. Biol.* **34**, 335–47 (2014).
102. Radhakrishnan, S. K., den Besten, W. & Deshaies, R. J. p97-dependent retrotranslocation and proteolytic processing govern formation of active Nrfl upon proteasome inhibition. *Elife* **3**, e01856 (2014).

103. Riemer, A. *et al.* The p97-Ufd1-Npl4 ATPase complex ensures robustness of the G₂/M checkpoint by facilitating CDC25A degradation. *Cell Cycle* **13**, 919–927 (2014).
104. Raman, M., Havens, C. G., Walter, J. C. & Harper, J. W. A Genome-wide Screen Identifies p97 as an Essential Regulator of DNA Damage-Dependent CDT1 Destruction. *Mol. Cell* **44**, 72–84 (2011).
105. Puumalainen, M.-R. *et al.* Chromatin retention of DNA damage sensors DDB2 and XPC through loss of p97 segregase causes genotoxicity. *Nat. Commun.* **5**, 3695 (2014).
106. Lavoie, C. *et al.* Tyrosine phosphorylation of p97 regulates transitional endoplasmic reticulum assembly in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13637–42 (2000).
107. Fortner, K. A. T Lymphocytes. *Kelley Firestein's Textb. Rheumatol.* 189–206 (2017). doi:10.1016/B978-0-323-31696-5.00012-7
108. Han, S., Williams, S. & Mustelin, T. Cytoskeletal protein tyrosine phosphatase PTPH1 reduces T cell antigen receptor signaling. *Eur. J. Immunol.* **30**, 1318–1325 (2000).
109. Sozio, M. S. *et al.* PTPH1 Is a Predominant Protein-tyrosine Phosphatase Capable of Interacting with and Dephosphorylating the T Cell Receptor ζ Subunit. *J. Biol. Chem.* **279**, 7760–7769 (2004).
110. Bauler, T. J. *et al.* Normal TCR Signal Transduction in Mice That Lack Catalytically Active PTPN3 Protein Tyrosine Phosphatase. *J. Immunol.* **178**, 3680–3687 (2007).
111. Bauler, T. J., Hendriks, W. J. A. J. & King, P. D. The FERM and PDZ domain-containing protein tyrosine phosphatases, PTPN4 and PTPN3, are both dispensable for T cell receptor signal transduction. *PLoS One* **3**, (2008).
112. Lichtenthaler, S. F., Lemberg, M. K. & Fluhrer, R. Proteolytic ectodomain shedding of membrane proteins in mammals—hardware, concepts, and recent developments. *EMBO J.* **37**, e99456 (2018).
113. Black, R. A. *et al.* A metalloproteinase disintegrin that releases tumour-necrosis factor- α from cells. *Nature* **385**, 729–733 (1997).
114. Zunke, F. The shedding protease ADAM17: Physiology and pathophysiology. *Biochim. Biophys. Acta - Mol. Cell Res.* **1864**, 2059–2070 (2017).
115. Zheng, Y., Schlöndorff, J. & Blobel, C. P. Evidence for Regulation of the Tumor Necrosis Factor α -Convertase (TACE) by Protein-tyrosine Phosphatase PTPH1. *J. Biol. Chem.* **277**, 42463–42470 (2002).
116. Xu, P. & Derynck, R. Direct Activation of TACE-Mediated Ectodomain Shedding by p38 MAP Kinase Regulates EGF Receptor-Dependent Cell Proliferation. *Mol. Cell* **37**, 551–566 (2010).
117. Schwarz, J. *et al.* Polo-like kinase 2, a novel ADAM17 signaling component, regulates tumor necrosis factor α ectodomain shedding. *J. Biol. Chem.* **289**, 3080–93 (2014).
118. Müllberg, J. *et al.* The soluble interleukin-6 receptor is generated by shedding. *Eur. J. Immunol.* **23**, 473–480 (1993).
119. Müllberg, J., Schooltink, H., Stoyan, T., Heinrich, P. C. & Rose-John, S. Protein kinase C activity is rate limiting for shedding of the interleukin-6 receptor. *Biochem. Biophys. Res. Commun.* **189**, 794–800 (1992).
120. DeMarco, K. R. & Clancy, C. E. Cardiac Na Channels: Structure to Function. *Curr. Top. Membr.* **78**, 287–311 (2016).
121. Gavillet, B. *et al.* Cardiac Sodium Channel Na_v 1.5 Is Regulated by a Multiprotein Complex Composed of Syntrophins and Dystrophin. *Circ. Res.* **99**, 407–414 (2006).
122. Jespersen, T. *et al.* Cardiac sodium channel Nav1.5 interacts with and is regulated by the protein tyrosine phosphatase PTPH1. *Biochem. Biophys. Res. Commun.* **348**, 1455–1462 (2006).

123. Ahern, C. A., Zhang, J.-F., Wookalis, M. J. & Horn, R. Modulation of the Cardiac Sodium Channel Na_v 1.5 by Fyn, a Src Family Tyrosine Kinase. *Circ. Res.* **96**, 991–998 (2005).
124. Ono, K. & Han, J. The p38 signal transduction pathway: activation and function. *Cell. Signal.* **12**, 1–13 (2000).
125. Sabio, G. *et al.* p38 γ regulates the localisation of SAP97 in the cytoskeleton by modulating its interaction with GKAP. *EMBO J.* **24**, 1134–1145 (2005).
126. Sabio, G. *et al.* Stress- and mitogen-induced phosphorylation of the synapse-associated protein SAP90/PSD-95 by activation of SAPK3/p38gamma and ERK1/ERK2. *Biochem. J.* **380**, 19–30 (2004).
127. Hasegawa, M. *et al.* Stress-activated Protein Kinase-3 Interacts with the PDZ Domain of α 1-Syntrophin. *J. Biol. Chem.* **274**, 12626–12631 (1999).
128. Cuenda, A. & Cohen, P. Stress-activated Protein Kinase-2/p38 and a Rapamycin-sensitive Pathway Are Required for C2C12 Myogenesis. *J. Biol. Chem.* **274**, 4341–4346 (1999).
129. Tang, J., Qi, X., Mercola, D., Han, J. & Chen, G. Essential role of p38gamma in K-Ras transformation independent of phosphorylation. *J. Biol. Chem.* **280**, 23910–7 (2005).
130. Patrignani, C. *et al.* Characterization of protein tyrosine phosphatase H1 knockout mice in animal models of local and systemic inflammation. *J. Inflamm. (Lond)*. **7**, 16 (2010).
131. Le Grevès, M., Le Grevès, P. & Nyberg, F. Age-related effects of IGF-1 on the NMDA-, GH- and IGF-1-receptor mRNA transcripts in the rat hippocampus. *Brain Res. Bull.* **65**, 369–374 (2005).
132. Whited, J. L., Robichaux, M. B., Yang, J. C. & Garrity, P. A. Ptpmeg is required for the proper establishment and maintenance of axon projections in the central brain of *Drosophila*. *Development* **126**, 4065–4076 (2007).
133. Li, S. *et al.* Protein tyrosine phosphatase PTPN3 promotes drug resistance and stem cell-like characteristics in ovarian cancer. *Sci. Rep.* **6**, 36873 (2016).
134. Zhi, H. *et al.* PTPH1 cooperates with vitamin D receptor to stimulate breast cancer growth through their mutual stabilization. *Oncogene* **30**, 1706–1715 (2011).
135. Wu, C.-W. *et al.* PTPN3 and PTPN4 tyrosine phosphatase expression in human gastric adenocarcinoma. *Anticancer Res* **26**, 1643–9 (2006).
136. Warabi, M., Nemoto, T., Ohashi, K., Kitagawa, M. & Hirokawa, K. Expression of protein tyrosine phosphatases and its significance in esophageal cancer. *Exp. Mol. Pathol.* **68**, 187–95 (2000).
137. Wang, Y., Su, Y., Ji, Z. & Lv, Z. High Expression of PTPN3 Predicts Progression and Unfavorable Prognosis of Glioblastoma. *Med. Sci. Monit.* **24**, 7556–7562 (2018).
138. Shi, Z. *et al.* PTPH1 promotes tumor growth and metastasis in human glioma. *Eur. Rev. Med. Pharmacol. Sci.* 3777–3787 (2016).
139. Tang, J., Qi, X., Mercola, D., Han, J. & Chen, G. Essential Role of p38 γ in K-Ras Transformation Independent of Phosphorylation. *J. Biol. Chem.* **280**, 23910–23917 (2005).
140. Gao, Q. *et al.* Activating mutations in PTPN3 promote cholangiocarcinoma cell proliferation and migration and are associated with tumor recurrence in patients. *Gastroenterology* **146**, 1397–1407 (2014).
141. Zhichao Wang. Protein tyrosine phosphatase PTPN3 inhibits hepatocellular carcinoma growth and metastasis by dephosphorylation of EGFR. *Ann. Oncol.* **26**, 42–70 (2015).
142. Jung, Y. *et al.* Discovery of ALK-PTPN3 gene fusion from human non-small cell lung carcinoma cell line using next generation RNA sequencing. *Genes, Chromosom. Cancer* **51**, 590–597 (2012).
143. Yuan, B. *et al.* PTPN3 acts as a tumor suppressor and boosts TGF- β signaling independent of its phosphatase activity. *EMBO J.* e99945 (2019). doi:10.15252/embj.201899945

144. Li, M.-Y. *et al.* PTPN3 suppresses lung cancer cell invasiveness by counteracting Src-mediated DAAM1 activation and actin polymerization. *Oncogene* 1–15 (2019). doi:10.1038/s41388-019-0948-6
145. Kennedy, M. B. Origin of PDZ (DHR, GLGF) domains. *Trends Biochem. Sci.* **20**, 350 (1995).
146. Cho, K.-O., Hunt, C. A. & Kennedy, M. B. The rat brain postsynaptic density fraction contains a homolog of the drosophila discs-large tumor suppressor protein. *Neuron* **9**, 929–942 (1992).
147. Liu, X. & Fuentes, E. J. *Emerging Themes in PDZ Domain Signaling: Structure, Function, and Inhibition. International Review of Cell and Molecular Biology* **343**, (Elsevier Ltd, 2019).
148. Luck, K., Charbonnier, S. & Travé, G. The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett.* **586**, 2648–2661 (2012).
149. Vincentelli, R. *et al.* Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat. Methods* **12**, 787–793 (2015).
150. Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* **114**, 3219–31 (2001).
151. Manjunath, G. P., Ramanujam, P. L. & Galande, S. Structure function relations in PDZ-domain-containing proteins: Implications for protein networks in cellular signalling. *J. Biosci.* **43**, 155–171 (2018).
152. Sheng, M. & Sala, C. PDZ Domains and the organization of supramolecular complexes. *Annu. Rev. Neurosci.* 1–29 (2001).
153. Ye, F. & Zhang, M. Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. *Biochem. J.* **455**, 1–14 (2013).
154. Venkatachalam, K., Luo, J. & Montell, C. Evolutionarily conserved, multitasking TRP channels: lessons from worms and flies. *Handb. Exp. Pharmacol.* **223**, 937–62 (2014).
155. Katz, B. & Minke, B. The Drosophila light-activated TRP and TRPL channels - Targets of the phosphoinositide signaling cascade. *Prog. Retin. Eye Res.* **66**, 200–219 (2018).
156. Tsunoda, S. *et al.* A multivalent PDZ-domain protein assembles signalling complexes in a G-protein-coupled cascade. *Nature* **388**, 243–249 (1997).
157. Tsunoda, S., Sun, Y., Suzuki, E. & Zuker, C. Independent anchoring and assembly mechanisms of INAD signaling complexes in Drosophila photoreceptors. *J. Neurosci.* **21**, 150–8 (2001).
158. Feng, W. & Zhang, M. Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. *Nat. Rev. Neurosci.* **10**, 87–99 (2009).
159. Doyle, D. A. *et al.* Crystal Structures of a Complexed and Peptide-Free Membrane Protein–Binding Domain: Molecular Basis of Peptide Recognition by PDZ. *Cell* **85**, 1067–1076 (1996).
160. Cabral, J. H. M. *et al.* Crystal structure of a PDZ domain. *Nature* **382**, 649–652 (1996).
161. Songyang, Z. *et al.* Recognition of Unique Carboxyl-Terminal Motifs by Distinct PDZ Domains. *Science (80-.).* **275**, 73–77 (1997).
162. Fuh, G. *et al.* Analysis of PDZ Domain-Ligand Interactions Using Carboxyl-terminal Phage Display. *J. Biol. Chem.* **275**, 21486–21491 (2000).
163. Stiffler, M. A. *et al.* PDZ domain binding selectivity is optimized across the mouse proteome. *Science* **317**, 364–9 (2007).
164. Niethammer, M. *et al.* CRIPT, a Novel Postsynaptic Protein that Binds to the Third PDZ Domain of PSD-95/SAP90. *Neuron* **20**, 693–707 (1998).
165. Doyle, D. A. *et al.* Crystal Structures of a Complexed and Peptide-Free Membrane Protein–Binding Domain: Molecular Basis of Peptide Recognition by PDZ. *Cell* **85**, 1067–1076 (1996).
166. Appleton, B. A. *et al.* 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**, 22312–20 (2006).

167. Skelton, N. J. *et al.* Origins of PDZ Domain Ligand Specificity. *J. Biol. Chem.* **278**, 7645–7654 (2003).
168. Fournane, S. *et al.* 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**, 511–523 (2011).
169. Birrane, G., Chung, J. & Ladias, J. A. A. Novel Mode of Ligand Recognition by the Erbin PDZ Domain. *J. Biol. Chem.* **278**, 1399–1402 (2003).
170. Kozlov, G., Banville, D., Gehring, K. & Ekiel, I. Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the beta2-beta3 loop to PDZ domain-ligand interactions. *J. Mol. Biol.* **320**, 813–20 (2002).
171. Terrien, E. *et al.* Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. *Sci. Signal.* **5**, ra58 (2012).
172. Zhang, J. *et al.* Structural Basis of β -Catenin Recognition by Tax-interacting Protein-1. *J. Mol. Biol.* **384**, 255–263 (2008).
173. Mostarda, S., Gfeller, D. & Rao, F. Beyond the Binding Site: The Role of the β 2 – β 3 Loop and Extra-Domain Structures in PDZ Domains. *PLoS Comput. Biol.* **8**, e1002429 (2012).
174. Tyler, R. C., Peterson, F. C. & Volkman, B. F. Distal interactions within the par3-VE-cadherin complex. *Biochemistry* **49**, 951–7 (2010).
175. Diella, F. *et al.* Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.* **Volume**, 6580 (2008).
176. Fuentes, E. J., Gilmore, S. A., Mauldin, R. V. & Lee, A. L. Evaluation of Energetic and Dynamic Coupling Networks in a PDZ Domain Protein. *J. Mol. Biol.* **364**, 337–351 (2006).
177. Chen, Y. *et al.* Genome-wide functional annotation of dual-specificity protein- and lipid-binding modules that regulate protein interactions. *Mol. Cell* **46**, 226–37 (2012).
178. Chang, B. H. *et al.* A Systematic Family-wide Investigation Reveals that ~30% of Mammalian PDZ Domains Engage in PDZ-PDZ Interactions. *Chem. Biol.* **18**, 1143–1152 (2011).
179. Luck, K. *et al.* Putting into practice domain-linear motif interaction predictions for exploration of protein networks. *PLoS One* **6**, e25376 (2011).
180. Penkert, R. R., DiVittorio, H. M. & Prehoda, K. E. Internal recognition through PDZ domain plasticity in the Par-6-Pals1 complex. *Nat. Struct. Mol. Biol.* **11**, 1122–7 (2004).
181. Tochio, H. *et al.* Formation of nNOS/PSD-95 PDZ dimer requires a preformed β -finger structure from the nNOS PDZ domain. *J. Mol. Biol.* **303**, 359–370 (2000).
182. Wang, P., Zhang, Q., Tochio, H., Fan, J.-S. & Zhang, M. Formation of a native-like β -hairpin finger structure of a peptide from the extended PDZ domain of neuronal nitric oxide synthase in aqueous solution. *Eur. J. Biochem.* **267**, 3116–3122 (2000).
183. Brenman, J. E. *et al.* Interaction of Nitric Oxide Synthase with the Postsynaptic Density Protein PSD-95 and α 1-Syntrophin Mediated by PDZ Domains. *Cell* **84**, 757–767 (1996).
184. Lenfant, N. *et al.* A genome-wide study of PDZ-domain interactions in *C. elegans* reveals a high frequency of non-canonical binding. *BMC Genomics* **11**, 671 (2010).
185. Mu, Y., Cai, P., Hu, S., Ma, S. & Gao, Y. Characterization of Diverse Internal Binding Specificities of PDZ Domains by Yeast Two-Hybrid Screening of a Special Peptide Library. *PLoS One* **9**, e88286 (2014).
186. Hsueh, Y.-P., Kim, E. & Sheng, M. Disulfide-Linked Head-to-Head Multimerization in the Mechanism of Ion Channel Clustering by PSD-95. *Neuron* **18**, 803–814 (1997).
187. Im, Y. J. *et al.* Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class II PDZ target recognition and PDZ domain-mediated multimerization. *J. Biol. Chem.* **278**, 8501–7 (2003).

188. Fanning, A. S., Lye, M. F., Anderson, J. M. & Lavie, A. Domain swapping within PDZ2 is responsible for dimerization of ZO proteins. *J. Biol. Chem.* **282**, 37710–6 (2007).
189. Chen, J., Pan, L., Wei, Z., Zhao, Y. & Zhang, M. Domain-swapped dimerization of ZO-1 PDZ2 generates specific and regulatory connexin43-binding sites. *EMBO J.* **27**, 2113–23 (2008).
190. Long, J.-F. *et al.* Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95. *J. Mol. Biol.* **327**, 203–214 (2003).
191. Dong, H. *et al.* Characterization of the glutamate receptor-interacting proteins GRIP1 and GRIP2. *J. Neurosci.* **19**, 6930–41 (1999).
192. Osten, P. *et al.* Mutagenesis reveals a role for ABP/GRIP binding to GluR2 in synaptic surface accumulation of the AMPA receptor. *Neuron* **27**, 313–25 (2000).
193. Feng, W., Shi, Y., Li, M. & Zhang, M. Tandem PDZ repeats in glutamate receptor–interacting proteins have a novel mode of PDZ domain–mediated target binding. *Nat. Struct. Mol. Biol.* **10**, 972–978 (2003).
194. Zhang, Q., Fan, J.-S. & Zhang, M. Interdomain Chaperoning between PSD-95, Dlg, and Zo-1 (PDZ) Domains of Glutamate Receptor-interacting Proteins. *J. Biol. Chem.* **276**, 43216–43220 (2001).
195. Zhu, J., Shang, Y., Chen, J. & Zhang, M. Structure and function of the guanylate kinase-like domain of the MAGUK family scaffold proteins. *Front. Biol. (Beijing)*. **7**, 379–396 (2012).
196. Pan, L., Chen, J., Yu, J., Yu, H. & Zhang, M. The Structure of the PDZ3-SH3-GuK Tandem of ZO-1 Protein Suggests a Supramodular Organization of the Membrane-associated Guanylate Kinase (MAGUK) Family Scaffold Protein Core. *J. Biol. Chem.* **286**, 40069–40074 (2011).
197. Nomme, J. *et al.* The Src Homology 3 Domain Is Required for Junctional Adhesion Molecule Binding to the Third PDZ Domain of the Scaffolding Protein ZO-1. *J. Biol. Chem.* **286**, 43352–43360 (2011).
198. Maisonneuve, P. *et al.* Regulation of the catalytic activity of the human phosphatase PTPN4 by its PDZ domain. *FEBS J.* **281**, 4852–4865 (2014).
199. Maisonneuve, P. *et al.* Molecular Basis of the Interaction of the Human Protein Tyrosine Phosphatase Non-receptor Type 4 (PTPN4) with the Mitogen-activated Protein Kinase p38 γ . *J. Biol. Chem.* **291**, 16699–16708 (2016).
200. Di Paolo, G. & De Camilli, P. Phosphoinositides in cell regulation and membrane dynamics. *Nature* **443**, 651–657 (2006).
201. Ivarsson, Y. Plasticity of PDZ domains in ligand recognition and signaling. *FEBS Lett.* **586**, 2638–2647 (2012).
202. Meerschaert, K. *et al.* The PDZ2 domain of zonula occludens-1 and -2 is a phosphoinositide binding domain. *Cell. Mol. Life Sci.* **66**, 3951–3966 (2009).
203. Wu, H. *et al.* PDZ Domains of Par-3 as Potential Phosphoinositide Signaling Integrators. *Mol. Cell* **28**, 886–898 (2007).
204. Pan, L. *et al.* Clustering and synaptic targeting of PICK1 requires direct interaction between the PDZ domain and lipid membranes. *EMBO J.* **26**, 4576–4587 (2007).
205. Ivarsson, Y. *et al.* Prevalence, Specificity and Determinants of Lipid-Interacting PDZ Domains from an In-Cell Screen and In Vitro Binding Experiments. *PLoS One* **8**, e54581 (2013).
206. Sheng, R. *et al.* Cholesterol modulates cell signaling and protein networking by specifically interacting with PDZ domain-containing scaffold proteins. *Nat. Commun.* **3**, 1249 (2012).
207. Fanning, A. S. & Anderson, J. M. PDZ domains: fundamental building blocks in the organization of protein complexes at the plasma membrane. *J. Clin. Invest.* **103**, 767–72 (1999).
208. Balla, T. Inositol-lipid binding motifs: signal integrators through protein-lipid and protein-

- protein interactions. *J. Cell Sci.* **118**, 2093–2104 (2005).
209. Xia, J., Zhang, X., Staudinger, J. & Haganir, R. L. Clustering of AMPA receptors by the synaptic PDZ domain-containing protein PICK1. *Neuron* **22**, 179–87 (1999).
 210. Chung, H. J., Xia, J., Scannevin, R. H., Zhang, X. & Haganir, R. L. Phosphorylation of the AMPA receptor subunit GluR2 differentially regulates its interaction with PDZ domain-containing proteins. *J. Neurosci.* **20**, 7258–67 (2000).
 211. Pagon, L. *et al.* The PDZ-binding motif of MCC is phosphorylated at position –1 and controls lamellipodia formation in colon epithelial cells. *Biochim. Biophys. Acta - Mol. Cell Res.* **1823**, 1058–1067 (2012).
 212. Boisguerin, P. *et al.* Characterization of a Putative Phosphorylation Switch: Adaptation of SPOT Synthesis to Analyze PDZ Domain Regulation Mechanisms. *ChemBioChem* **8**, 2302–2307 (2007).
 213. Pedersen, S. W. *et al.* Site-Specific Phosphorylation of PSD-95 PDZ Domains Reveals Fine-Tuned Regulation of Protein–Protein Interactions. *ACS Chem. Biol.* **12**, 2313–2323 (2017).
 214. Long, J.-F. *et al.* Autoinhibition of X11/Mint scaffold proteins revealed by the closed conformation of the PDZ tandem. *Nat. Struct. Mol. Biol.* **12**, 722–728 (2005).
 215. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18249–54 (2009).
 216. Bhattacharya, S. *et al.* A Conformational Switch in the Scaffolding Protein NHERF1 Controls Autoinhibition and Complex Formation. *J. Biol. Chem.* **285**, 9981–9994 (2010).
 217. Jemth, P. & Gianni, S. PDZ Domains: Folding and Binding †. *Biochemistry* **46**, 8701–8708 (2007).
 218. Gautier, C., Laursen, L., Jemth, P. & Gianni, S. Seeking allosteric networks in PDZ domains. *Protein Eng. Des. Sel.* **31**, 367–373 (2018).
 219. Adato, A. *et al.* Usherin, the defective protein in Usher syndrome type IIA, is likely to be a component of interstereocilia ankle links in the inner ear sensory cells. *Hum. Mol. Genet.* **14**, 3921–3932 (2005).
 220. Mburu, P. *et al.* Defects in whirlin, a PDZ domain molecule involved in stereocilia elongation, cause deafness in the whirler mouse and families with DFNB31. *Nat. Genet.* **34**, 421–428 (2003).
 221. Yang, J. *et al.* Ablation of Whirlin Long Isoform Disrupts the USH2 Protein Complex and Causes Vision and Hearing Loss. *PLoS Genet.* **6**, e1000955 (2010).
 222. Via, A., Uyar, B., Brun, C. & Zanzoni, A. How pathogens use linear motifs to perturb host cell networks. *Trends Biochem. Sci.* **40**, 36–48 (2015).
 223. Davey, N. E. *et al.* Attributes of short linear motifs. *Mol. BioSyst.* **8**, 268–281 (2012).
 224. James, C. & Roberts, S. Viral Interactions with PDZ Domain-Containing Proteins—An Oncogenic Trait? *Pathogens* **5**, 8 (2016).
 225. Kiyono, T. *et al.* Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11612–6 (1997).
 226. Lee, S. S., Weiss, R. S. & Javier, R. T. Binding of human virus oncoproteins to hDlg/SAP97, a mammalian homolog of the Drosophila discs large tumor suppressor protein. *Proc. Natl. Acad. Sci.* **94**, 6670–6675 (1997).
 227. Rousset, R., Fabre, S., Desbois, C., Bantignies, F. & Jalinot, P. The C-terminus of the HTLV-1 Tax oncoprotein mediates interaction with the PDZ domain of cellular proteins. *Oncogene* **16**, 643–654 (1998).
 228. Weiss, R. S. & Javier, R. T. A carboxy-terminal region required by the adenovirus type 9 E4 ORF1 oncoprotein for transformation mediates direct binding to cellular polypeptides. *J. Virol.*

- 71, 7873–80 (1997).
229. Javier, R. T. & Rice, A. P. Emerging Theme: Cellular PDZ Proteins as Common Targets of Pathogenic Viruses. *J. Virol.* **85**, 11544–11556 (2011).
 230. Ganti, K. *et al.* The human papillomavirus E6 PDZ binding motif: From life cycle to malignancy. *Viruses* **7**, 3530–3551 (2015).
 231. Margolis, B. Apicobasal polarity complexes. *J. Cell Sci.* **118**, 5157–5159 (2005).
 232. Martin-Belmonte, F. & Perez-Moreno, M. Epithelial cell polarity, stem cells and cancer. *Nat. Rev. Cancer* **12**, 23–38 (2012).
 233. Javier, R. T. & Rice, A. P. Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. *J. Virol.* **85**, 11544–11556 (2011).
 234. Thomas, M. & Banks, L. Upsetting the Balance: When Viruses Manipulate Cell Polarity Control. *J. Mol. Biol.* (2018). doi:10.1016/j.jmb.2018.04.016
 235. Gutiérrez-González, L. H. & Santos-Mendoza, T. Viral targeting of PDZ polarity proteins in the immune system as a potential evasion mechanism. *FASEB J.* fj.201900518R (2019). doi:10.1096/fj.201900518R
 236. Huse, M. Lymphocyte polarity, the immunological synapse and the scope of biological analogy. *Bioarchitecture* **1**, 180–185 (2011).
 237. Goldenring, J. R. A central role for vesicle trafficking in epithelial neoplasia: intracellular highways to carcinogenesis. *Nat. Rev. Cancer* **13**, 813–820 (2013).
 238. Javier, R., Raska, K., Macdonald, G. J. & Shenk, T. Human adenovirus type 9-induced rat mammary tumors. *J. Virol.* **65**, 3192–202 (1991).
 239. Latorre, I. J. *et al.* Viral oncoprotein-induced mislocalization of select PDZ proteins disrupts tight junctions and causes polarity defects in epithelial cells. *J. Cell Sci.* **118**, 4283–93 (2005).
 240. Kong, K., Kumar, M., Taruishi, M. & Javier, R. T. The human adenovirus E4-ORF1 protein subverts discs large 1 to mediate membrane recruitment and dysregulation of phosphatidylinositol 3-kinase. *PLoS Pathog.* **10**, e1004102 (2014).
 241. Glaunsinger, B. A., Weiss, R. S., Lee, S. S. & Javier, R. Link of the unique oncogenic properties of adenovirus type 9 E4-ORF1 to a select interaction with the candidate tumor suppressor protein ZO-2. *EMBO J.* **20**, 5578–5586 (2001).
 242. Chlenski, A. *et al.* Organization and expression of the human zo-2 gene (tjp-2) in normal and neoplastic tissues. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1493**, 319–324 (2000).
 243. Xie, L., Yamamoto, B., Haoudi, A., Semmes, O. J. & Green, P. L. PDZ binding motif of HTLV-1 Tax promotes virus-mediated T-cell proliferation in vitro and persistence in vivo. *Blood* **107**, 1980 (2006).
 244. Hirata, A. *et al.* PDZ domain-binding motif of human T-cell leukemia virus type 1 Tax oncoprotein augments the transforming activity in a rat fibroblast cell line. *Virology* **318**, 327–336 (2004).
 245. Higuchi, M. *et al.* Cooperation of NF- κ B/p100 Activation and the PDZ Domain Binding Motif Signal in Human T-Cell Leukemia Virus Type 1 (HTLV-1) Tax1 but Not HTLV-2 Tax2 Is Crucial for Interleukin-2-Independent Growth Transformation of a T-Cell Line. *J. Virol.* **81**, 11900–11907 (2007).
 246. Yoshida, S. *et al.* Knockdown of synapse-associated protein Dlg1 reduces syncytium formation induced by human T-cell leukemia virus type 1. *Virus Genes* **37**, 9–15 (2008).
 247. Nejmeddine, M. *et al.* HTLV-1-Tax and ICAM-1 act on T-cell signal pathways to polarize the microtubule-organizing center at the virological synapse. *Blood* **114**, 1016–1025 (2009).
 248. Igakura, T. *et al.* Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. *Science* **299**, 1713–6 (2003).
 249. Préhaud, C., Lay, S., Dietzschold, B. & Lafon, M. Glycoprotein of Nonpathogenic Rabies

- Viruses Is a Key Determinant of Human Cell Apoptosis. *J. Virol.* **77**, 10537–10547 (2003).
250. Préhaud, C. *et al.* Attenuation of rabies virulence: takeover by the cytoplasmic domain of its envelope protein. *Sci. Signal.* **3**, ra5 (2010).
 251. Loh, S. H. Y., Francescut, L., Lingor, P., Bähr, M. & Nicotera, P. Identification of new kinase clusters required for neurite outgrowth and retraction by a loss-of-function RNA interference screen. *Cell Death Differ.* **15**, 283–298 (2008).
 252. Leslie, N. R., Maccario, H., Spinelli, L. & Davidson, L. The significance of PTEN's protein phosphatase activity. *Adv. Enzyme Regul.* **49**, 190–196 (2009).
 253. Babault, N. *et al.* Peptides Targeting the PDZ Domain of PTPN4 Are Efficient Inducers of Glioblastoma Cell Death. *Structure* **19**, 1518–1524 (2011).
 254. Caillet-Saguy, C. *et al.* Strategies to interfere with PDZ-mediated interactions in neurons: What we can learn from the rabies virus. *Prog. Biophys. Mol. Biol.* **119**, 53–59 (2015).
 255. Khan, Z. *et al.* Structure-based optimization of a PDZ binding motif within a viral peptide stimulates neurite outgrowth. *J. Biol. Chem.* jbc.RA119.008238 (2019). doi:10.1074/jbc.RA119.008238
 256. Jing, M., Bohl, J., Brimer, N., Kinter, M. & Vande Pol, S. B. Degradation of tyrosine phosphatase PTPN3 (PTPH1) by association with oncogenic human papillomavirus E6 proteins. *J. Virol.* **81**, 2231–9 (2007).
 257. Bruni L *et al.* ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in the World. Summary Report 17 June 2019. (2019). Available at: <https://www.hpvcentre.net/statistics/reports/XWX.pdf>. (Accessed: 1st August 2019)
 258. Graham, S. V. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. *Clin. Sci. (Lond)*. **131**, 2201–2221 (2017).
 259. Graham, S. V. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. *Clin. Sci. (Lond)*. **131**, 2201–2221 (2017).
 260. DiGiuseppe, S., Bienkowska-Haba, M., Guion, L. G. & Sapp, M. Cruising the cellular highways: How human papillomavirus travels from the surface to the nucleus. *Virus Res.* **231**, 1–9 (2017).
 261. Doorbar, J. *et al.* The Biology and Life-Cycle of Human Papillomaviruses. *Vaccine* **30**, F55–F70 (2012).
 262. Roden, R. B. S. & Stern, P. L. Opportunities and challenges for human papillomavirus vaccination in cancer. *Nat. Rev. Cancer* **18**, 240–254 (2018).
 263. de Villiers, E.-M., Fauquet, C., Broker, T. R., Bernard, H.-U. & zur Hausen, H. Classification of papillomaviruses. *Virology* **324**, 17–27 (2004).
 264. Vande Pol, S. B. & Klingelutz, A. J. Papillomavirus E6 oncoproteins. *Virology* **445**, 115–37 (2013).
 265. Maufort, J. P., Shai, A., Pitot, H. C. & Lambert, P. F. A role for HPV16 E5 in cervical carcinogenesis. *Cancer Res.* **70**, 2924–31 (2010).
 266. de Freitas, A. C., de Oliveira, T. H. A., Barros, M. R., Venuti, A. & Venuti, A. hrHPV E5 oncoprotein: immune evasion and related immunotherapies. *J. Exp. Clin. Cancer Res.* **36**, 71 (2017).
 267. Estêvão, D., Costa, N. R., Gil da Costa, R. M. & Medeiros, R. Hallmarks of HPV carcinogenesis: The role of E6, E7 and E5 oncoproteins in cellular malignancy. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1862**, 153–162 (2019).
 268. Mittal, S. & Banks, L. Molecular mechanisms underlying human papillomavirus E6 and E7 oncoprotein-induced cell transformation. *Mutat. Res. Mutat. Res.* **772**, 23–35 (2017).
 269. Huibregtse, J. M., Scheffner, M. & Howley, P. M. Localization of the E6-AP regions that

- direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins. *Mol. Cell. Biol.* **13**, 4918–27 (1993).
270. Zanier, K. *et al.* Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. *Science* **339**, 694–8 (2013).
271. Zanier, K. *et al.* Solution structure analysis of the HPV16 E6 oncoprotein reveals a self-association mechanism required for E6-mediated degradation of p53. *Structure* **20**, 604–617 (2012).
272. Nominé, Y. *et al.* Structural and Functional Analysis of E6 Oncoprotein: Insights in the Molecular Pathways of Human Papillomavirus-Mediated Pathogenesis. *Mol. Cell* **21**, 665–678 (2006).
273. Zanier, K. *et al.* Formation of well-defined soluble aggregates upon fusion to MBP is a generic property of E6 proteins from various human papillomavirus species. *Protein Expr. Purif.* **51**, 59–70 (2007).
274. Lee, C. & Laimins, L. A. Role of the PDZ domain-binding motif of the oncoprotein E6 in the pathogenesis of human papillomavirus type 31. *J. Virol.* **78**, 12366–77 (2004).
275. Delury, C. P. *et al.* The Role of Protein Kinase A Regulation of the E6 PDZ-Binding Domain during the Differentiation-Dependent Life Cycle of Human Papillomavirus Type 18. *J. Virol.* **87**, 9463–9472 (2013).
276. Nicolaides, L. *et al.* Stabilization of HPV16 E6 protein by PDZ proteins, and potential implications for genome maintenance. *Virology* **414**, 137–145 (2011).
277. Watson, R. A., Thomas, M., Banks, L. & Roberts, S. Activity of the human papillomavirus E6 PDZ-binding motif correlates with an enhanced morphological transformation of immortalized human keratinocytes. *J. Cell Sci.* **116**, 4925–4934 (2003).
278. Spanos, W. C. *et al.* Deletion of the PDZ motif of HPV16 E6 preventing immortalization and anchorage-independent growth in human tonsil epithelial cells. *Head Neck* **30**, 139–147 (2008).
279. James, M. A., Lee, J. H. & Klingelutz, A. J. Human papillomavirus type 16 E6 activates NF- κ B, induces cIAP-2 expression, and protects against apoptosis in a PDZ binding motif-dependent manner. *J. Virol.* **80**, 5301–7 (2006).
280. Nguyen, M. L., Nguyen, M. M., Lee, D., Griep, A. E. & Lambert, P. F. The PDZ Ligand Domain of the Human Papillomavirus Type 16 E6 Protein Is Required for E6 's Induction of Epithelial Hyperplasia In Vivo. **77**, 6957–6964 (2003).
281. Gardiol, D. *et al.* Oncogenic human papillomavirus E6 proteins target the discs large tumour suppressor for proteasome-mediated degradation. *Oncogene* **18**, 5487–5496 (1999).
282. Nakagawa, S. & Huibregtse, J. M. Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. *Mol. Cell. Biol.* **20**, 8244–53 (2000).
283. Kranjec, C. & Banks, L. A systematic analysis of human papillomavirus (HPV) E6 PDZ substrates identifies MAGI-1 as a major target of HPV type 16 (HPV-16) and HPV-18 whose loss accompanies disruption of tight junctions. *J. Virol.* **85**, 1757–64 (2011).
284. Massimi, P., Gammoh, N., Thomas, M. & Banks, L. HPV E6 specifically targets different cellular pools of its PDZ domain-containing tumour suppressor substrates for proteasome-mediated degradation. *Oncogene* **23**, 8033–8039 (2004).
285. Feigin, M. E. *et al.* Mislocalization of the Cell Polarity Protein Scribble Promotes Mammary Tumorigenesis and Is Associated with Basal Breast Cancer. *Cancer Res.* **74**, 3180–3194 (2014).
286. Thomas, M. *et al.* Oncogenic human papillomavirus E6 proteins target the MAGI-2 and MAGI-3 proteins for degradation. *Oncogene* **21**, 5088–5096 (2002).
287. Kranjec, C., Massimi, P. & Banks, L. Restoration of MAGI-1 expression in human

- papillomavirus-positive tumor cells induces cell growth arrest and apoptosis. *J. Virol.* **88**, 7155–69 (2014).
288. Shin, K., Straight, S. & Margolis, B. PATJ regulates tight junction formation and polarity in mammalian epithelial cells. *J. Cell Biol.* **168**, 705–711 (2005).
 289. Michel, D. *et al.* PATJ connects and stabilizes apical and lateral components of tight junctions in human intestinal cells. *J. Cell Sci.* **118**, 4049–4057 (2005).
 290. Storrs, C. H. & Silverstein, S. J. PATJ, a tight junction-associated PDZ protein, is a novel degradation target of high-risk human papillomavirus E6 and the alternatively spliced isoform 18 E6. *J. Virol.* **81**, 4080–90 (2007).
 291. Hamazaki, Y., Itoh, M., Sasaki, H., Furuse, M. & Tsukita, S. Multi-PDZ domain protein 1 (MUPP1) is concentrated at tight junctions through its possible interaction with claudin-1 and junctional adhesion molecule. *J. Biol. Chem.* **277**, 455–61 (2002).
 292. Lee, S. S., Glaunsinger, B., Mantovani, F., Banks, L. & Javier, R. T. Multi-PDZ domain protein MUPP1 is a cellular target for both adenovirus E4-ORF1 and high-risk papillomavirus type 18 E6 oncoproteins. *J. Virol.* **74**, 9680–93 (2000).
 293. Hampson, L., Li, C., Oliver, A. W., Kitchener, H. C. & Hampson, I. N. The PDZ protein Tip-1 is a gain of function target of the HPV16 E6 oncoprotein. *Int. J. Oncol.* **25**, 1249–56 (2004).
 294. Razanskas, R. & Sasnauskas, K. Interaction of hepatitis B virus core protein with human GIPC1. *Arch. Virol.* **155**, 247–250 (2010).
 295. Handa, K. *et al.* E6AP-dependent degradation of DLG4/PSD95 by high-risk human papillomavirus type 18 E6 protein. *J. Virol.* **81**, 1379–89 (2007).
 296. Jeong, K. W., Kim, H.-Z., Kim, S., Kim, Y. S. & Choe, J. Human papillomavirus type 16 E6 protein interacts with cystic fibrosis transmembrane regulator-associated ligand and promotes E6-associated protein-mediated ubiquitination and proteasomal degradation. *Oncogene* **26**, 487–499 (2007).
 297. Chang, J. T.-C. *et al.* Highly potent and specific siRNAs against E6 or E7 genes of HPV16- or HPV18-infected cervical cancers. *Cancer Gene Ther.* **17**, 827–36 (2010).
 298. Hoppe-Seyler, K., Bossler, F., Braun, J. A., Herrmann, A. L. & Hoppe-Seyler, F. The HPV E6/E7 Oncogenes: Key Factors for Viral Carcinogenesis and Therapeutic Targets. *Trends Microbiol.* **26**, 158–168 (2018).
 299. Martinez-Zapien, D. *et al.* Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature* **529**, 541–545 (2016).
 300. Stutz, C. *et al.* Intracellular Analysis of the Interaction between the Human Papillomavirus Type 16 E6 Oncoprotein and Inhibitory Peptides. *PLoS One* **10**, e0132339 (2015).
 301. Zanier, K. *et al.* The E6AP Binding Pocket of the HPV16 E6 Oncoprotein Provides a Docking Site for a Small Inhibitory Peptide Unrelated to E6AP, Indicating Druggability of E6. *PLoS One* **9**, e112514 (2014).
 302. Guo, C. *et al.* Potent Anti-Tumor Effect Generated by a Novel Human Papillomavirus (HPV) Antagonist Peptide Reactivating the pRb/E2F Pathway. *PLoS One* **6**, e17734 (2011).
 303. Nizard, M. *et al.* Immunotherapy of HPV-associated head and neck cancer: Critical parameters. *Oncoimmunology* **2**, e24534 (2013).
 304. Stern, P. L. *et al.* Therapy of Human Papillomavirus-Related Disease. *Vaccine* **30**, F71–F82 (2012).
 305. Selzer, L. & Zlotnick, A. Assembly and Release of Hepatitis B Virus. 1–18 (2017).
 306. Venkatakrisnan, B. & Zlotnick, A. The Structural Biology of Hepatitis B Virus: Form and Function. *Annu. Rev. Virol.* **3**, 429–451 (2016).
 307. Hu, J. & Liu, K. Complete and Incomplete Hepatitis B Virus Particles: Formation, Function, and Application. *Viruses* **9**, (2017).

308. Li, H.-C. *et al.* Nuclear Export and Import of Human Hepatitis B Virus Capsid Protein and Particles. *PLoS Pathog.* **6**, e1001162 (2010).
309. Diab, A., Foca, A., Zoulim, F., Durantel, D. & Andrisani, O. The diverse functions of the hepatitis B core/capsid protein (HBc) in the viral life cycle: Implications for the development of HBc-targeting antivirals. *Antiviral Res.* **149**, 211–220 (2018).
310. Kim, D. H., Kang, H. S. & Kim, K.-H. Roles of hepatocyte nuclear factors in hepatitis B virus infection. *World J. Gastroenterol.* **22**, 7017 (2016).
311. Nassal, M. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J. Virol.* **66**, 4107–16 (1992).
312. Blondot, M., Bruss, V. & Kann, M. Intracellular transport and egress of hepatitis B virus. *J. Hepatol.* **64**, S49–S59 (2016).
313. Jung, J. *et al.* Phosphoacceptors threonine 162 and serines 170 and 178 within the carboxyl-terminal RRRS/T motif of the hepatitis B virus core protein make multiple contributions to hepatitis B virus replication. *J. Virol.* **88**, 8754–67 (2014).
314. Chen, C., Wang, J. C.-Y. & Zlotnick, A. A Kinase Chaperones Hepatitis B Virus Capsid Assembly and Captures Capsid Dynamics in vitro. *PLoS Pathog.* **7**, e1002388 (2011).
315. Wynne, S. ., Crowther, R. . & Leslie, A. G. . The Crystal Structure of the Human Hepatitis B Virus Capsid. *Mol. Cell* **3**, 771–780 (1999).
316. Selzer, L., Katen, S. P. & Zlotnick, A. The Hepatitis B Virus Core Protein Intradimer Interface Modulates Capsid Assembly and Stability. (2014).
317. Ceres, P. & Zlotnick, A. Weak Protein–Protein Interactions Are Sufficient To Drive Assembly of Hepatitis B Virus Capsids. *Biochemistry* **41**, 11525–11531 (2002).
318. Bourne, C. R., Katen, S. P., Fulz, M. R., Packianathan, C. & Zlotnick, A. A mutant hepatitis B virus core protein mimics inhibitors of icosahedral capsid self-assembly. *Biochemistry* **48**, 1736–42 (2009).
319. Packianathan, C., Katen, S. P., Dann, C. E. & Zlotnick, A. Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly. *J. Virol.* **84**, 1607–15 (2010).
320. Alexander, C. G., Jürgens, M. C., Shepherd, D. A., Freund, S. M. V & Ashcroft, A. E. Thermodynamic origins of protein folding , allostery , and capsid formation in the human hepatitis B virus core protein. (2013). doi:10.1073/pnas.1308846110
321. Klumpp, K. *et al.* High-resolution crystal structure of a hepatitis B virus replication inhibitor bound to the viral core protein. *Proc. Natl. Acad. Sci.* **112**, 15196–15201 (2015).
322. Yu, X., Jin, L., Jih, J., Shih, C. & Hong Zhou, Z. 3.5Å cryoEM Structure of Hepatitis B Virus Core Assembled from Full-Length Core Protein. *PLoS One* **8**, e69729 (2013).
323. Nassal, M. Hepatitis B viruses: Reverse transcription a different way. *Virus Res.* **134**, 235–249 (2008).
324. Chua, P. K., Tang, F.-M., Huang, J.-Y., Suen, C.-S. & Shih, C. Testing the balanced electrostatic interaction hypothesis of hepatitis B virus DNA synthesis by using an in vivo charge rebalance approach. *J. Virol.* **84**, 2340–51 (2010).
325. Lewellyn, E. B. & Loeb, D. D. The arginine clusters of the carboxy-terminal domain of the core protein of hepatitis B virus make pleiotropic contributions to genome replication. *J. Virol.* **85**, 1298–309 (2011).
326. Le Pogam, S., Chua, P. K., Newman, M. & Shih, C. Exposure of RNA templates and encapsidation of spliced viral RNA are influenced by the arginine-rich domain of human hepatitis B virus core antigen (HBcAg 165-173). *J. Virol.* **79**, 1871–87 (2005).
327. Chu, T.-H., Liou, A.-T., Su, P.-Y., Wu, H.-N. & Shih, C. Nucleic acid chaperone activity

- associated with the arginine-rich domain of human hepatitis B virus core protein. *J. Virol.* **88**, 2530–43 (2014).
328. Tan, Z. *et al.* The interface between hepatitis B virus capsid proteins affects self-assembly, pregenomic RNA packaging, and reverse transcription. *J. Virol.* **89**, 3275–84 (2015).
329. Katoh, M. GIPC gene family. *Int. J. Mol. Med.* **9**, 585–9 (2002).
330. Rousset, R., Fabre, S., Desbois, C., Bantignies, F. & Jalinot, P. The C-terminus of the HTLV-1 Tax oncoprotein mediates interaction with the PDZ domain of cellular proteins. *Oncogene* **16**, 643–654 (1998).
331. Favre-Bonvin, A., Reynaud, C., Kretz-Remy, C. & Jalinot, P. Human papillomavirus type 18 E6 protein binds the cellular PDZ protein TIP-2/GIPC, which is involved in transforming growth factor beta signaling and triggers its degradation by the proteasome. *J. Virol.* **79**, 4229–37 (2005).
332. Deres, K. *et al.* Inhibition of hepatitis B virus replication by drug-induced depletion of nucleocapsids. *Science* **299**, 893–6 (2003).
333. Delaney, W. E. *et al.* Phenylpropenamide derivatives AT-61 and AT-130 inhibit replication of wild-type and lamivudine-resistant strains of hepatitis B virus in vitro. *Antimicrob. Agents Chemother.* **46**, 3057–60 (2002).
334. Zlotnick, A. *et al.* Core protein: A pleiotropic keystone in the HBV lifecycle. *Antiviral Res.* **121**, 82–93 (2015).
335. Julien, S. G., Dubé, N., Hardy, S. & Tremblay, M. L. Inside the human cancer tyrosine phosphatome. *Nat. Rev. Cancer* **11**, 35–49 (2011).
336. Hendriks, W. J. A. J. & Böhmer, F.-D. *Non-transmembrane PTPs in Cancer in Protein Tyrosine Phosphatases in Cancer* (ed. Neel B. G. & Tonks N. K.). (Springer, 2016).
337. Chen, K. E. *et al.* Substrate specificity and plasticity of FERM-containing protein tyrosine phosphatases. *Structure* **23**, 653–664 (2015).
338. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.* **3**, 39–53 (2003).
339. Caillet-Saguy, C. *et al.* Regulation of the Human Phosphatase PTPN4 by the inter-domain linker connecting the PDZ and the phosphatase domains. *Sci. Rep.* **7**, 7875 (2017).
340. Genera, M. *et al.* Structural and functional characterization of the PDZ domain of the human phosphatase PTPN3 and its interaction with the human papillomavirus E6 oncoprotein. *Sci. Rep.* **9**, 7438 (2019).
341. Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M. J. & Shoelson, S. E. Crystal structure of the tyrosine phosphatase SHP-2. *Cell* **92**, 441–50 (1998).
342. Gupta, S. *et al.* The C-terminal Tail of Presenilin Regulates Omi/HtrA2 Protease Activity. *J. Biol. Chem.* **279**, 45844–45854 (2004).
343. Walsh, N. P., Alba, B. M., Bose, B., Gross, C. A. & Sauer, R. T. OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* **113**, 61–71 (2003).
344. Chen, M. J., Dixon, J. E. & Manning, G. Genomics and evolution of protein phosphatases. *Sci. Signal.* **10**, eaag1796 (2017).
345. Tonikian, R. *et al.* A specificity map for the PDZ domain family. *PLoS Biol.* **6**, e239 (2008).
346. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**, 125–142 (1984).
347. Hirakawa, T., Galet, C., Kishi, M. & Ascoli, M. GIPC binds to the human lutropin receptor (hLHR) through an unusual PDZ domain binding motif, and it regulates the sorting of the internalized human choriogonadotropin and the density of cell surface hLHR. *J. Biol. Chem.* **278**, 49348–57 (2003).

348. Bohlson, S. S., Zhang, M., Ortiz, C. E. & Tenner, A. J. CD93 interacts with the PDZ domain-containing adaptor protein GIPC: implications in the modulation of phagocytosis. *J. Leukoc. Biol.* **77**, 80–89 (2005).
349. Ligensa, T. *et al.* A PDZ domain protein interacts with the C-terminal tail of the insulin-like growth factor-1 receptor but not with the insulin receptor. *J. Biol. Chem.* **276**, 33419–27 (2001).
350. Jeanneteau, F., Diaz, J., Sokoloff, P. & Griffon, N. Interactions of GIPC with Dopamine D2 , D3 but not D4 Receptors Define a Novel Mode of Regulation of G Protein-coupled Receptors. *Mol. Biol. Cell* **15**, 696–705 (2004).
351. Chastre, E. *et al.* TRIP6, a novel molecular partner of the MAGI-1 scaffolding molecule, promotes invasiveness. *FASEB J.* **23**, 916–928 (2009).
352. Cuppen, E. *et al.* The zyxin-related protein TRIP6 interacts with PDZ motifs in the adaptor protein RIL and the protein tyrosine phosphatase PTP-BL. *Eur. J. Cell Biol.* **79**, 283–93 (2000).
353. Petit, M. M. R., Crombez, K. R. M. O., Vervenne, H. B. V. K., Weyns, N. & Van de Ven, W. J. M. The tumor suppressor Scrib selectively interacts with specific members of the zyxin family of proteins. *FEBS Lett.* **579**, 5061–5068 (2005).
354. Sugi, T. *et al.* Crystal structures of autoinhibitory PDZ domain of Tamalin: implications for metabotropic glutamate receptor trafficking regulation. *EMBO J.* **26**, 2192–2205 (2007).
355. Kang, B. S. *et al.* PDZ Tandem of Human Syntenin: Crystal Structure and Functional Properties. *Structure* **11**, 459–468 (2003).
356. Elkins, J. M. *et al.* Unusual binding interactions in PDZ domain crystal structures help explain binding mechanisms. *Protein Sci.* **19**, 731–41 (2010).
357. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).
358. Franzosa, E. A. & Xia, Y. Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci.* **108**, 10538–10543 (2011).
359. Delhommel, F. *et al.* Deciphering the unconventional peptide binding to the PDZ domain of MAST2. *Biochem. J.* **469**, 159–168 (2015).
360. Hultqvist, G. *et al.* Energetic Pathway Sampling in a Protein Interaction Domain. *Structure* **21**, 1193–1202 (2013).
361. Fuentes, E. J., Der, C. J. & Lee, A. L. Ligand-dependent Dynamics and Intramolecular Signaling in a PDZ Domain. *J. Mol. Biol.* **335**, 1105–1115 (2004).
362. Christensen, N. R. *et al.* PDZ Domains as Drug Targets. *Adv. Ther.* **2**, 1800143 (2019).

Etude structurale et fonctionnelle de la phosphatase humaine PTPN3 et de ses interactions avec des virus oncogènes

Résumé:

La tyrosine phosphatase non récepteur de type 3 (PTPN3) est une phosphatase humaine contenant un domaine PDZ (PSD-95/Dlg/ZO-1) avec un rôle soit de suppresseur de tumeur soit de promoteur de tumeur dans de nombreux cancers. Cependant, sa fonction dans la signalisation cellulaire reste encore floue. Fait intéressant, les papillomavirus humains (HPV) génitaux à haut risque de types 16 et 18 et le virus de l'hépatite B (HBV) ciblent le domaine PDZ de PTPN3 (PTPN3-PDZ) par le biais de motifs de reconnaissance au domaine PDZ (PBMs) dans l'oncoprotéine E6 et la protéine de capsid Hbc de HPV et HBV respectivement. Nous avons étudié de façon détaillée les interactions entre le domaine PDZ de PTPN3 et ses ligands cellulaires et viraux. Pour étudier les propriétés structurales et fonctionnelles du domaine PDZ de PTPN3 et de son interaction avec le PBM de E6, nous avons combiné des expériences de biophysique, de RMN et de diffraction aux rayons X. Nous avons étendu notre étude structurale de PTPN3-PDZ à d'autres partenaires cellulaires et viraux. Ces données permettent de mieux comprendre les principaux déterminants structuraux de la reconnaissance des PBMs. Nous nous sommes ensuite intéressés plus particulièrement à la protéine de capsid Hbc de HBV. En criblant une bibliothèque contenant l'ensemble des domaines PDZ des protéines humaines, nous avons identifié 28 partenaires cellulaires potentiels impliqués majoritairement dans la polarité cellulaire et interagissant avec le PBM de Hbc. Nous avons confirmé que PTPN3 pouvait interagir avec le PBM de Hbc au sein d'une capsid virale et nous avons montré que les PBMs viraux interagissaient avec PTPN3-PDZ avec des affinités similaires aux ligands endogènes de PTPN3. En utilisant des hépatocytes infectés par HBV, nous avons observé que la surexpression de PTPN3 avait des effets multiples sur l'infection. Enfin, nous avons étudié l'interactome de PTPN3-PDZ afin de mieux comprendre le rôle de PTPN3 dans la signalisation cellulaire et les effets perturbateurs de HBV sur celle-ci.

Mots-clés: [PTPN3, protéine tyrosine phosphatase, virus de l'hépatite B, papillomavirus humain, cristallographie aux rayons X, RMN, biophysique]

Structural and functional study of the human phosphatase PTPN3 and its interaction with oncogenic viruses

Abstract:

The human protein tyrosine phosphatase non-receptor type 3 (PTPN3) is a PDZ (PSD-95/Dlg/ZO-1) domain-containing phosphatase with a tumor-suppressive or a tumor-promoting role in many cancers, although its role in cell signalling is still unclear. Interestingly, the high-risk genital human papillomavirus (HPV) types 16 and 18 and the hepatitis B virus (HBV) target the PDZ domain of PTPN3 through PDZ-binding motifs (PBMs) in their E6 and Hbc proteins. Here, I report a detailed study of the interactions between the PDZ domain of PTPN3 and its cellular and viral ligands. First, we combined biophysical, NMR and X-ray experiments to investigate the structural and functional properties of the PDZ domain of PTPN3 and its interaction with the E6 PBM. We then extended our structural study of PTPN3-PDZ to other cellular and viral partners, and gained insights into the main structural determinants of recognition of PBMs. We then focused on the HBV Hbc protein. We screened a library of human PDZ-containing proteins for Hbc binders and identified 28 cellular Hbc-interacting partners, most of which are involved in cell polarity. We confirmed that PTPN3 can bind the Hbc PBM in the context of the viral capsid, and we showed that viral PBMs interact with PTPN3-PDZ with similar affinities to endogenous PTPN3 ligands. Using HBV-infected hepatocytes we observed that overexpression of PTPN3 has multiple effects on HBV infection. Finally, we investigated the interactome of PTPN3-PDZ to gain insights into the role of this protein in cell signalling and the disruptive effects of HBV.

Keywords: [PTPN3, protein tyrosine phosphatase, hepatitis B virus, human papillomavirus, X-ray crystallography, NMR, biophysics]