



Smad binding codes broken by WW domain containing proteins

Desxifrant els codis d'unió a Smad de les proteïnes amb dominis WW

Eric Aragón Altarriba

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



Smad binding codes broken by WW domain containing proteins

Smad binding codes
broken by WW domain
containing proteins

Eric Aragón Altarriba
2013



Universitat de Barcelona

Tesi Doctoral

Smad binding codes broken by WW domain containing proteins

Desxifrant els codis d'unió a Smad de les proteïnes amb dominis WW

Memòria presentada per

Eric Aragón Altarriba

Per obtenir el grau de doctor per la Universitat de Barcelona

Revisat per:

Director de Tesi:

Dra. Maria J. Macias Hernández

Tutor de Tesi:

Dr. Albert Tauler Girona

Prof. de investigació ICREA* i
investigadora principal al grup RMN
proteïnes del IRB Barcelona
*(Institució Catalana de Recerca i Estudis
Avançats)

Prof. de Bioquímica i Biologia
Molecular
Facultat de Farmàcia
Universitat de Barcelona

Programa de Doctorat en Biomedicina

Realitzada al Institut de Recerca Biomèdica Barcelona

Parc Científic Barcelona

Barcelona, 2013



INSTITUTE
FOR RESEARCH
IN BIOMEDICINE

“The grand essentials of happiness are: something to do,
something to love and something to hope for.”

(Allan Chalmers)

Acknowledgements.

The work described in this Thesis would not have been possible without the help and support of many people and institutions. I would therefore like to express my gratitude to:

-Dr. Maria J. Macias, my PhD supervisor, for introducing me in the fields of Molecular and Structural Biology, for her unconditional and permanent support and for her guidance and confidence in the success of the projects,

-Dr. Joan Massagué and co-workers in the Memorial Sloan-Kettering Cancer Center NY, for this fruitful and pleasant collaboration,

- Dr. Joan Guinovart for his encouragement during the writing of this Thesis, Dr. Jorge Lloberas for his advice regarding the registration at the Barcelona University and Dr. Miquel Pons for my initial training as a PhD student in his lab,

-Dr. Pau Martin for his friendship, for reading the manuscript of the Thesis at various stages and for his help with the cover figures and with the final lay out of this Thesis,

-Dr. Miriam Royo (Parc Científic de Barcelona) for her advice on protein chemical ligation and Dr. Marta Vilaseca (Mass Spectrometry Core Facility) for introducing me to the IM-MS experiments,

-All my Lab mates and many colleagues from the IRB Barcelona, for all their help, and support, for giving me valuable hints and stimulating discussions and above all for many memories and experiences,

-My family, for their continuous encouragement and support.

I have been financed by the IRB Barcelona and also in part by a Consolider RNAREG (CSD2009-00080) grant. The experimental work described in this Thesis was supported by SMSI-BFU2008-02795 and SAF2011-25119 (M.J.M.) grants.

Declarations:

1. This work has been carried out under the supervision of Maria J. Macias Ph.D. (ICREA research professor and PI of the Protein NMR group at the IRB Barcelona) at the department of Structural and Computational Biology of the Institute for Research in Biomedicine of Barcelona.
2. No portion of the work referred to in this thesis has been submitted in support of an application for another degree qualification of this or any other university.
3. The work described in this Thesis has been published in two peer-reviewed papers. Reprints and references are given in the Appendix section.

Financial support was obtained from the IRB Barcelona PhD contract and by a Consolider RNAREG (CSD2009-00080) grant. This work was supported in part by BFU2008-02795 and by a SAF2011-25119 grant.



Abbreviations:

BMP	Bone morphogenetic protein
HSQC	Heteronuclear Single-Quantum Coherence
ITC	Isothermal Titration Calorimetry
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PCR	Polymerase chain reaction
SDS-PAGE	Sodium dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SMAD	Small mothers against decapentaplegic
SMURF	Smad ubiquitin regulatory factor
SPPS	Solid phase peptide synthesis
TEAD	Transcriptional enhancer factor
TEV	Tobacco etch virus
TGF- β	Transforming growth factor beta
TOCSY	Total Correlation Spectroscopy
YAP	Yes associated protein kinase

Table of Contents

LIST OF FIGURES 17

1.	<u>GENERAL INTRODUCTION</u>	21
2.	<u>AIMS OF THE PRESENT THESIS</u>	31
3.	<u>MATERIAL AND METHODS</u>	37
3.1	CLONING	37
3.2	TRANSFORMATION	37
3.3	PROTEINS	37
3.3.1	DETERMINATION OF PROTEIN CONCENTRATIONS	38
3.4	NMR SPECTROSCOPY	39
3.5	NMR TITRATION EXPERIMENTS	39
3.6	STRUCTURE CALCULATION/DETERMINATION AND REFINEMENT	39
3.7	MOLECULAR DYNAMIC SIMULATIONS	40
3.8	SEQUENCE ALIGNMENTS	40
3.9	PEPTIDE SYNTHESIS	40
3.10	MASS SPECTROMETRY	41
3.11	ISOTHERMAL TITRATION CALORIMETRY	41
3.12	ION MOBILITY-MASS SPECTROMETRY (IM-MS)	42
3.13	BUFFERS AND SOLUTIONS	43
4.	<u>PHOSPHORYLATION BY GSK3 SWITCHES SMAD1 FROM BINDING TRANSCRIPTIONAL COFACTORS TO BINDING UBIQUITIN LIGASES THAT TRIGGER SMAD1 DEGRADATION</u>	49
4.1	CHAPTER SUMMARY:	51
4.2	NOTES:	52
4.3	DIFFERENT PHOSPHORYLATION SITES IN SMAD1 ARE RECOGNIZED BY YAP AND SMURF1	53

4.3.1	INTRODUCTION	53
4.3.2	BINDING AFFINITIES DETERMINED USING ISOTHERMAL TITRATION CALORIMETRY	56
4.3.3	STRUCTURE OF THE SMURF1 WW1-WW2 PAIR BOUND TO THE DOUBLY PHOSPHORYLATED SMAD1 LINKER	59
4.3.4	STRUCTURE OF THE SMURF1 WW1-WW2 PAIR BOUND TO THE MONO- PHOSPHORYLATED SMAD1 LINKER	64
4.3.5	STRUCTURE OF THE YAP WW PAIR BOUND TO THE SMAD1 LINKER	67
4.3.6	UNDERSTANDING THE PREFERENCES OF YAP-SMAD1 INTERACTIONS: THE CDK8/9 PHOSPHORYLATED SMAD1 LINKER AVOIDS YAP BINDING WHEN IS ALSO PHOSPHORYLATED BY GSK3 AT POSITIONS	70
5.	<u>STRUCTURAL BASIS FOR THE VERSATILE INTERACTIONS OF SMAD7 WITH REGULATOR WW DOMAINS IN TGF-B PATHWAYS</u>	85
5.1	SUMMARY	87
5.2	NOTES	88
5.3	INTRODUCTION	89
5.4	RESULTS	90
5.4.1	SMURF1, SMURF2 AND YAP USE ONE SINGLE WW DOMAIN TO BIND SMAD7	90
5.4.2	STRUCTURE OF THE SMAD7 PY MOTIF BOUND TO DIFFERENT E3 UBIQUITIN LIGASE WW DOMAINS	98
5.4.3	STRUCTURE OF THE COMPLEXES BETWEEN YAP AND SMAD7 PY MOTIF: SMAD7 SELECTS YAP WW1 AND NOT WW2	103
5.5	ASSIGNMENTS AND COORDINATES	117
6.	<u>DISCUSSION</u>	121
7.	<u>CONCLUSIONS</u>	133
7.1	REFERENCES	134
8.	<u>APPENDIX</u>	143

8.1	TABLE 1	143
8.2	TABLE 2	145
8.3	<i>CURRICULUM VITAE</i>	147
8.4	REPRINTS	151
9.	<u>RESUMEN EN CASTELLANO:</u>	183
9.1	INTRODUCCIÓN GENERAL	183
9.2	OBJETIVOS DE LA PRESENTE TESIS	190
9.3	RESULTADOS	193
9.4	DISCUSIÓN	201
9.5	CONCLUSIONES	210

List of Figures

FIGURE 1: MODULAR COMPOSITION OF SMAD PROTEINS	22
FIGURE 2: MODULAR COMPOSITION OF SMAD BINDING PROTEINS.....	24
FIGURE 3: GRAPHICAL HIGHLIGHT OF THE WORK.....	32
FIGURE 4: SCHEMATIC REPRESENTATION OF THE SMAD PROTEIN DOMAINS AND THEIR MAIN FUNCTIONS	54
FIGURE 6: ITC VALUES AND PEPTIDES.....	56
FIGURE 7: IP EXPERIMENTS WITH SMAD1 MUTANTS	57
FIGURE 8: SCHEMATIC SUMMARY OF THE SMAD ACTION-TURNOVER SWITCH OPERATED BY CDK8/9 AND GSK3 IN COMBINATION WITH YAP AND SMURF1	58
FIGURE 9: NMR-MODEL OF THE COMPLEX BETWEEN THE HUMAN SMURF1 WW1-WW2 PAIR (RESIDUES 232-314) AND THE 208-233 SEGMENT OF THE SMAD1 LINKER DI-PHOSPHORYLATED AT S210 AND S214	61
FIGURE 10: SMURF1WW1 BOUND TO THE SMAD1 PS210PS214 SITE	63
FIGURE 11: SMURF1 WW2 DOMAIN BOUND TO THE PY BOX.....	63
FIGURE 12: A COMPARISON OF SMURF1 WW1 DOMAIN BOUND TO EITHER SMAD1 PS214 OR TO PS210 AND PS214 SITES	65
FIGURE 13: SCHEMATIC REPRESENTATION OF THE MODE OF BINDING OF SMURF1 TO THE SMAD1 LINKER REGION	66
FIGURE 14: COMPLEX BETWEEN THE HUMAN YAP WW1-WW2 PAIR AND THE 199-233 SEGMENT OF THE SMAD1 LINKER	68
FIGURE 15: VIEW OF YAP WW1 BOUND TO PS206 SMAD1 SITE	69
FIGURE 16: VIEW OF YAP WW2 BOUND TO PY SMAD1 SITE.....	69
FIGURE 17: VIEW OF YAP WW1 BOUND TO PS206 AND PT202/PS206 SMAD1 SITE	71
FIGURE 18: CHARGE DISTRIBUTION OF SMURF1 WW1 BOUND TO PS214 SMAD1 SITE.....	72
FIGURE 19: CHARGE DISTRIBUTION OF SMURF1 WW1 BOUND TO PS210-PS214 SMAD1 SITE	72
FIGURE 20: CHARGE DISTRIBUTION OF YAP WW1 BOUND TO PT202-PS206 SMAD1 SITE	73
FIGURE 21: CHARGE DISTRIBUTION OF YAP WW1 BOUND TO PT202-PS206 SMAD1 SITE	73
FIGURE 22: MOLECULAR SIMULATIONS PERFORMED ON SMAD1 PEPTIDES.....	74
FIGURE 23: SCHEMATIC REPRESENTATION OF THE MODE OF BINDING OF YAP TO THE SMAD1 LINKER REGION	75
FIGURE 24: DOMAIN COMPOSITION OF THREE E3 UBIQUITIN LIGASES, SMAD7 AND YAP.....	90
FIGURE 25: ITC VALUES OBTAINED FOR THE WW DOMAINS PRESENT IN SMURF1, SMURF2 AND YAP AND THE PY MOTIF OF SMAD7	91
FIGURE 26: ION MOBILITY-MASS SPECTROMETRY DATA OBTAINED FOR THE COMPLEXES OF SMURF2WW2-WW3 WITH THE SMAD7 PEPTIDE.....	93
FIGURE 27: ION MOBILITY-MASS SPECTROMETRY DATA OBTAINED FOR THE COMPLEXES OF SMURF2WW2-WW3 WITH THE SMAD7 PY SITE	95

FIGURE 28: SCHEMATIC REPRESENTATION OF THE PAIR OF WW DOMAINS AND THE LINKER PRESENT IN SMURF1	96
FIGURE 29: AFFINITY VALUES OBTAINED BY ITC TITRATING SMURF1 WW PAIRS WITH SEVERAL POINT AND DOUBLE MUTATIONS	97
FIGURE 30: UNAMBIGUOUSLY ASSIGNED NOES BETWEEN MONOMERS.....	99
FIGURE 31: MODEL OF THE WW1-WW1 HOMODIMER.....	99
FIGURE 32: REFINED STRUCTURE OF THE SMURF1 WW2 BOUND TO THE SMAD7 PEPTIDE	101
FIGURE 33 REFINED STRUCTURE OF THE SMURF2 WW3 BOUND TO THE SMAD7 PEPTIDE	102
FIGURE 34: SMURF1 WW2 AND SMURF2 WW3 ARE THE MAIN PLAYERS FOR THE RECOGNITION OF SMAD7 PY SITE	103
FIGURE 35: DETAILED VIEW OF THE INTERACTION OF THE YAP WW1 DOMAIN WITH THE SMAD7 [PY] PEPTIDE	104
FIGURE 36: SEMITRANSSPARENT SURFACE REPRESENTATION SHOWING THE CHARGE DISTRIBUTION OF THE YAP WW2 DOMAIN BOUND TO THE SMAD7 PEPTIDE	105
FIGURE 37: AFFINITY CHANGES IN SMAD7 BINDING RESULTING FROM POINT MUTATIONS IN THE YAP WW1 AND WW2 DOMAINS	106
FIGURE 38: IP EXPERIMENTS SHOW THAT PHOSPHORYLATION HAS A MINOR EFFECT IN THE INTERACTION.....	107
FIGURE 39: SCHEMATIC REPRESENTATION OF YAP BINDING TO SMAD1 AND SMAD7 SITES	107
FIGURE 40: SCHEMATIC REPRESENTATION OF THE RESULTS PRESENTED IN THIS WORK.	128
FIGURE 41: SCHEMATIC OF THE SMAD LINKER PHOSPHO-AMINO ACID CODES (INSETS) AND WW DOMAIN CODE READERS.	129

General Introduction

1. General Introduction

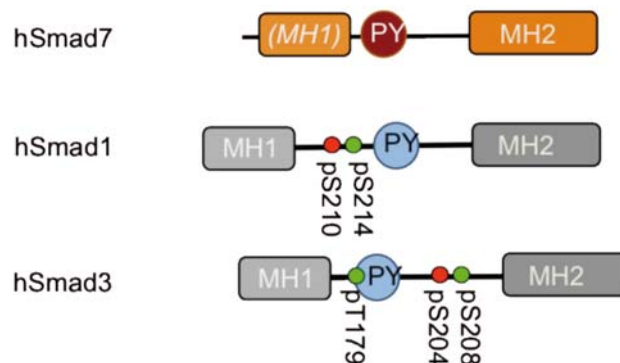
Cell fate is controlled by a multitude of signals and loss of this control has devastating consequences for living organisms. One of the key players in this network of signals is the TGF- β family of cytokines. These hormones trigger an immense amount of responses by sending activated Smad transcription factors (Sma and Mad related proteins) to the nucleus where participate in the control of stem cell pluripotency and differentiation, embryo development, tissue regeneration, and differentiated tissue homeostasis (Massagué, 1998).

According to their function, Smad proteins are classified as receptor regulated Smads (R-Smads), which include Smads 1, 5 and 8 in the BMP-driven version of the SMAD pathway, and Smads 2 and 3 in the TGF- β /Nodal/Activin pathways. R-Smads form complexes with the common co-activator Smad (Co-Smad) Smad4. The SMAD family also contains the two inhibitory Smads (I-Smads), Smad6 and Smad7, which provide critical negative regulation to these powerful and ubiquitous pathways.

All Smad proteins are modular (Shi and Massagué, 2003). R-Smads and the Co-Smad consist of two Mad Homology MH1 and MH2 domains connected by a linker that functions as a scaffold upon which other proteins can interact and modulate the functional outcome. This linker contains a conserved cluster of phosphorylation sites adjacent to a PY motif. MH1 domains of R-Smads and Smad4 bind to DNA, whereas the MH2 domain and the linker function as scaffolds for receptors, regulator proteins, and transcription cofactors to interact and determine the outcome of the signal (Shi and Massagué, 2003). Compared to R-Smads and Co-Smads, the I-Smads have low sequence similarity in the MH1 domain but conserve an MH2 domain and a linker region with a characteristic PY motif. The presence of the

common regions facilitates the competition of R-Smads and I-Smads for the receptor and ligands and it facilitates the inhibitory role of I-Smads (Figure 1). I-Smads are expressed in response to TGF- β or BMPs to provide negative feedback in the pathway (Bai and Cao, 2002; Hata et al., 1998; He et al., 2002; Kavsak et al., 2000; Nakao et al., 1997; Yan and Chen, 2011) and in response to other pathways such as STAT to oppose TGF- β signaling (Ulloa et al., 1999). Smad6 interferes with the formation of Smad1-Smad4 complexes (Hata et al., 1998) whereas Smad7 interferes with the formation of R-Smads-Smad4 complexes and inhibits TGF- β and BMP receptors (Hayashi et al., 1997; Topper et al., 1998).

Figure 1: Modular composition of Smad proteins



The MH1 sequence present in Smad7 is shown in italics only to remark its high degree of divergency.

Several key phosphorylations drive the Smad signaling process. The ligand cytokines activate receptor serine/threonine protein kinases that phosphorylate Smad proteins at the C-terminus. The BMP receptors act on Smads 1, 5 and 8 and the receptors for the TGF- β /nodal/activin/myostatin group of ligands act mainly on Smads 2 and 3 (Shi and Massagué, 2003). The phosphorylated C-terminus provides a binding site for Smad4, which is an essential component in the assembly of target-specific transcriptional complexes. These phosphorylations are reversed by protein phosphatases that

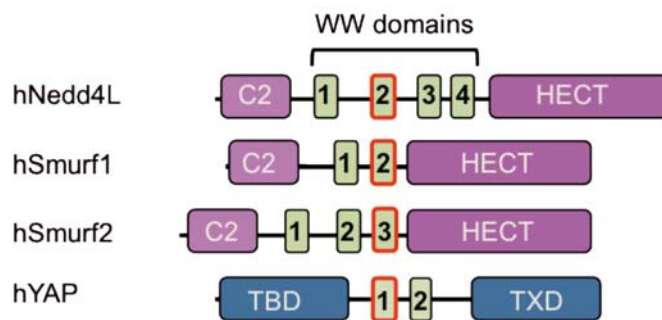
limit the general pool of activated Smad molecules (Inman et al., 2002; Lin et al., 2006; Schmierer et al., 2008; Xu et al., 2002).

Receptor-activated Smad proteins that associate with Smad4 and bind to target genes undergo a second set of phosphorylations, these catalyzed by the transcriptional cyclin-dependent kinases CDK8 and CDK9 (Alarcón et al., 2009; Gao et al., 2009) and by glycogen synthase kinase-3 (GSK3) (Alarcón et al., 2009; Fuentealba et al., 2007; Sapkota et al., 2007). CDKs 8 and 9 are part of transcriptional Mediator and Elongation complexes, respectively (Durand et al., 2005; Komarnitsky et al., 2000; Malik and Roeder, 2000). GSK3 is a Wnt- and PI3K-regulated kinase (Cohen and Frame, 2001; Wu and Pan). CDK8/9 phosphorylation of Smad serves as the priming event for phosphorylation by GSK3. These phosphorylations are clustered in an interdomain linker region and enable peak activation of Smads but also mark the proteins for poly-ubiquitination and proteasome-mediated degradation (Alarcón et al., 2009; Gao et al., 2009). Degradation of agonist-activated Smads (Alarcón et al., 2009; Lo and Massagué, 1999) occurs alongside dephosphorylation of the linker (Sapkota et al., 2007; Wrighton et al., 2006). Whereas dephosphorylation recycles the Smad proteins for repeated rounds of signaling, action-coupled destruction of Smad depletes the pool of signal transducer. In a different context, the Smad linker region is phosphorylated by MAP kinases and cell division CDKs in response to mitogens and stresses to constrain TGF- β and BMP signaling (Kretzschmar et al., 1997; Kretzschmar et al., 1999; Matsuura et al., 2009).

Four proteins are known to specifically bind to linker-phosphorylated Smads during BMP and TGF- β signal transduction. The ubiquitin ligase Smurf1 (Sapkota et al., 2007) and the transcriptional effector of the Hippo pathway YAP bind to linker-phosphorylated Smad1/5 (Alarcón et al., 2009), whereas the Smurf1-related protein Nedd4L (Gao et al., 2009) and the

peptidyl-prolyl cis/trans isomerase Pin1 (Matsuura et al., 2009) bind to linker-phosphorylated Smad2/3. YAP cooperates with Smad1 to activate genes that suppress neural differentiation in mouse embryonic stem cells in response to BMP signals (Alarcón et al., 2009). Pin1 cooperates with Smad2/3 to stimulate cancer cell migration in response to TGF- β (Matsuura et al., 2009). Smurf1 and Nedd4L target activated Smad1/5 and Smad2/3, respectively for polyubiquitination and proteasome-dependent degradation (Figure 2).

Figure 2: Modular composition of Smad binding proteins



Common to this set of Smad binding proteins is the presence of WW domains, one in Pin1, two in Smurf1 and YAP, and four in Nedd4L. WW domains are 38-40 amino acid residue units, characterized by two highly conserved tryptophans and folded as an antiparallel triple-stranded beta-sheet that typically binds proline-rich sequences (e.g. PPxY or “PY box”) or in the case of Pin1, phosphorylated Ser or Thr residues followed by Proline (pS/pT-P motifs) (Macias et al., 2002). R-Smad proteins contain a PY motif located near the CDK/GSK3 phosphorylation sites in the linker connecting the MH1 and MH2 domains. Binding of WW domains and PY motifs have been extensively reported in the literature (Macias et al. 2002, (Macias et al., 1996; Pires et al., 2001; Toepert et al., 2001) suggesting that the PY site in Smads and the WW domains present in YAP, Pin1 and in the ubiquitin ligases may be responsible

for the interactions. These lines of evidence present a scenario in which different nuclear protein kinases phosphorylate agonist-activated Smads to create docking sites for competing transcriptional cofactors and ubiquitin ligases. The outcome of these interactions governs Smad function, and is therefore important in BMP and TGF- β signal transduction. However, the convergence of activation and turnover functions on a clustered set of Smad modifications raises questions about how Smads get to act before undergoing disposal.

In the case of the ubiquitin ligase Nedd4L protein, the interaction with Smad2/3 requires that the Threonine preceding the PY site be phosphorylated, suggesting that in addition to the characteristic contacts between WW domains and the PY motif, a potential interaction involving the phosphorylated threonine might also occur.

The necessity of recognizing the pSP motif of Smad1 upstream of the PY site by Smurf1 and Yap1 proteins raises the question of how this long binding site is specifically recognized. An attractive possibility is that the pair of WW domains present in each protein participates in the interaction using both WW domains, one responsible for the PY recognition site and the other acting as the pSP-binding site. Although an arrangement where both domains bind a target in a synchronized manner could explain the experimental data, such an arrangement has never been reported before.

In addition, Pin1WW domain is the unique WW domain found so far that recognizes phosphorylated Ser/Thr residues followed by a proline (Verdecia et al 2000). However, the Pin1WW residues described as being responsible for the phosphate recognition are not strictly present in Yap, Smurf1 or Nedd4L WW domains. Thus, it seems that other WW domains different from Pin1 have found new solutions to the recognition of

phosphorylated residues, a feature probably critical for its specificity *in vivo*. Furthermore, Smurf1/2 and Nedd4L do not equally bind to all R-Smads since the three-ubiquitin ligases cannot be functionally interchanged both *in vitro* and *in vivo* (Gao S et al 2009) raising the question of how this specificity is achieved.

In addition to the issues of how YAP, Pin1 and the E3 HECT ubiquitin ligases read the phosphorylated code written in R-Smads, work in recent years has revealed the inhibitory Smad7 as a central hub for negative regulation of activated TGF- β and BMP receptors (Yan and Chen, 2011). Smad7 recruits ubiquitin ligases Nedd4L, Smurf1 and Smurf2 to mediate receptor polyubiquitination and route the receptor to degradative endocytosis (Ebisawa et al., 2001; Kavsak et al., 2000; Kuratomi et al., 2005). Moreover, Smad7 can simultaneously bind Smurf2 and the protein deubiquitinase USP15, recruiting both enzymes to the TGF- β receptor complex for an integrated control of receptor polyubiquitination as a function of ligand concentration (Eichhorn et al., 2012). Smad7 also binds YAP (Ferrigno et al., 2002) providing a mechanism for sequestration of this mediator of Hippo and BMP signaling (Alarcón et al., 2009). As with Smad1/2 and 3, all these interactions involve the linker region of Smad7 that contains a PY motif and the WW domain region of Nedd4L, Smurf1/2, and YAP.

Prompted by these observations we set to characterize how the interactions between Smad proteins and their targets occur and particularly how these similar WW domains can discern between Smad sequences and chose a particular one in time. To achieve such ambitious aims, the laboratory where I have done the present work established a fruitful collaboration with

that of Dr. Joan Massagué*. In our lab we set to investigate the interactions at an atomic detail between Yap1WW1/2 and Smurf1WW1-2 pair with a pS210/pS214-PY 24 peptide of Smad1, and also of Pin1 and Nedd4L with Smad3, while our collaborators set to discover the biological scenarios where these interactions take place.

*A collaboration with the group of Dr. J. Massagué, Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA.

Aims of the present Thesis

2. Aims of the present Thesis

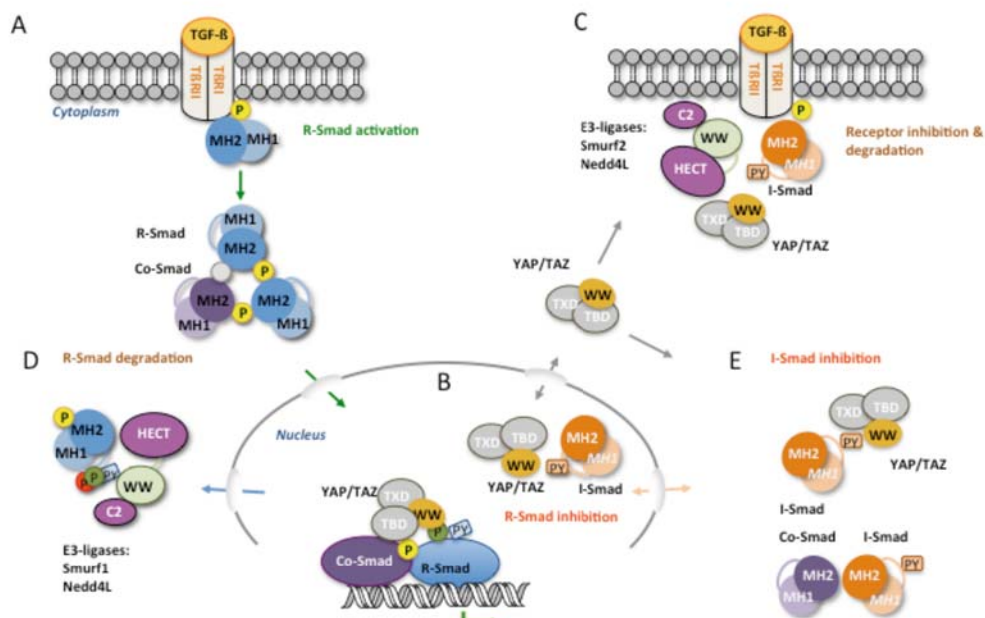
Aim 1. As it is mentioned in the General Introduction, fundamental aspects of metazoan embryo development and tissue homeostasis are controlled by TGF- β and BMP through Smad-mediated transcription of master regulator genes. In the course of this action in the nucleus, Smad proteins undergo phosphorylation events that enable peak transcriptional activity but also mark the proteins for destruction. These findings presented a paradox but also an opportunity to define how the delivery of TGF- β and BMP signals is coupled to the turnover of the Smad signal transducers. We postulated that a mechanism must exist that ensures the orderly sequence of events in this process by somehow switching Smad proteins from binding transcriptional cofactors to binding ubiquitin ligases.

Prompted by these hypotheses and using the power of a combined functional and structural approach, we set to characterize the role of the Smad1 phosphorylation sites in the interaction with YAP and Smurf1 and to describe at an atomic detail how these similar WW domain containing proteins discern between targets and choose a particular one. Since the sequential action of CDK8/9 and GSK3 kinases can alter the phosphorylation pattern of the pS/pT box, it could be that the recognition of the different combinations of pS/pT phosphorylations might act as a *switch* to favor the binding of a given pair of proteins at a given time controlling the outcome of the signal, from activation to degradation.

Aim 2. Smad7 is an inhibitory Smad that acts as a negative regulator of signaling by (TGF- β) superfamily of proteins. Smad7 is induced by TGF- β , stably interacts with activated TGF- β type I receptor (T β R-I), and interferes

with the phosphorylation of receptor-regulated Smads. I-Smads compete with R-Smads for activation using the proline rich motif of Smad7, a region similar to that of Smad 1 and Smad 3, and the region spanning the WW domains of the ligases. It has been described that Smurf1 and Smurf2 participate in the degradation of Smad 7 and also in the degradation of the TGF- β receptor. In order to define the structural bases that define the interactions we plan to identify the WW domains of YAP, Smurf1 and Smurf2, that participate in the recognition of Smad7 and thus describe the structural basis that control the competition between activation and inhibition.

Figure 3: Graphical highlight of the work



- A.** Upon binding of the TGF- β hormone to the receptor, the receptor forms a heterodimer and is activated. The activation is a two-step mechanism that starts with the activation of the T β receptor II, which is a kinase that in turns phosphorylates the receptor I. Then, the R-Smad is brought to the proximity of the membrane and is twice phosphorylated at the protein C-terminus. The presence of both phosphates creates a binding site for the common Smad4. The result of this process is the formation of a heterotrimeric complex that is ready to enter into the nucleus.
- B.** Once in the nucleus the complex is further phosphorylated, this time in the linker that connects the MH domains of the R-Smads. After phosphorylation the hetero-trimer is ready to bind to cofactors (as YAP/TAZ) than enhance the Smad specificity for DNA sequences and start to transcribe the specific genes. Among these genes is that of Smad7, one of the inhibitors of the signaling cascade.
- C.** Smad7 regulates TGF- β signaling through several processes. It can occupy the R-Smad binding site in the receptor, precluding the activation of R-Smads and its transfer to the nucleus.
- D.** After being participated in transcription, a second round of phosphorylations label Smad proteins for degradation.
- E.** Smad7 can also interact with Smad4, sequestering this protein in the cytoplasm. Smad7 can also interact with YAP/TAZ and reduce the amount of these proteins, affecting the formation of the complex with co-transcription factors.

Material and Methods

3. Material and Methods

3.1 Cloning

All constructs corresponding to Smurf1 (WW1 233-270, WW2 277-314 and WW1-2 277-314 (NP_851994.1), Smurf2 (WW2 249-287, WW3 297-333 and WW2-3 249-333 (NP_073576.1) and hYAP1 WW domains (WW1 163-206, WW2 227-266 and WW1-2 163-266) (NP_001123617.1) were amplified using the appropriated DNA templates or human cDNA and subsequently cloned into the pETM30/pETM11 vectors (a gift from the EMBL-Heidelberg Protein Expression Facility) using NcoI and HindIII sites or using recombinant strategies.

Point mutations were introduced using the QuickChange™ site directed mutagenesis Kit (Stratagene®) with the appropriate complementary mutagenic primers. All wild type and variants were confirmed by DNA sequencing and Mass Spectrometry.

3.2 Transformation

The transformation of the expression vector into the host cell's plasmid DNA was done with the heat shock method using competent BL21(DE3) *E. coli* cells (Novagen, Darmstadt, Germany). Cells were grown for 45 min at 37°C, centrifuged and dispersed on an agar plate with antibiotic for selection. The agar plate remained at 37°C overnight in order to obtain bacterial colonies.

3.3 Proteins

Unlabeled, ¹⁵N-labeled, double ¹³C-, ¹⁵N- labeled and triple ²H-, ¹³C-, ¹⁵N-labeled proteins were expressed in *E. coli* BL21 (DE3) in (L-Broth or Luria Bertani medium or in minimal medium (M9), using either H₂O or D₂O (99.89%, CortecNet) enriched with ¹⁵NH₄Cl and/or D-[¹³C] glucose as the sole

sources of carbon and nitrogen respectively (Marley et al., 2001). *E. coli* extracts were lysed using an EmulsiFlex-C5 (Avestin) cell disrupter equipped with an in-house developed Peltier temperature controller system. Soluble fusion proteins were purified by nickel-affinity chromatography (HiTrap Chelating HP column, GE Healthcare Life Science, Uppsala Sweden) and samples were eluted using buffer A with EDTA. Smurf1 WW1-WW2 protein was mostly in the insoluble fraction after centrifugation of *E. coli* lysates and was solubilized with 6M Guanidine hydrochloride and then purified using the HiTrap HP column. After buffer exchange, fusion tags were removed by overnight TEV protease digestion at 4 °C followed by a second nickel-affinity binding step. All proteins were further purified with an additional gel filtration chromatography step, using HiLoad™ Superdex 30, 75 or 200 16/60 pregrade columns (GE Healthcare), depending on the protein size. Fractions containing the purified proteins were concentrated to 1-2 mM for NMR experiments. To ensure the presence of a 1:1 protein: peptide ratio and to avoid formation of aggregates or misfolded samples, Smurf1, Smurf2 and YAP proteins were concentrated in the presence of the Smad1 peptides prior to NMR experiments. The NMR buffer was 20 mM deuterated Tris-HCl (pH 7.2-7.4), 100 mM NaCl, 0.01% NaN₃, and 2.5 mM deuterated dithiothreitol in the presence of 10% D₂O.

3.3.1 Determination of Protein Concentrations

Protein concentrations were determined by measuring the UV absorption of the protein sample using a Shimadzu UVmini-1240 UV-VIS spectrophotometer. Radiation absorption by proteins in the near UV depends on the number of aromatic amino acids with absorption maximum at 275 nm (Tyr) and at 280 nm (Trp). Since each WW domain contains several aromatic residues, the protein concentrations can be calculated using the Lambert-Beer law (3.1).

$$C = \frac{A_{280}}{\varepsilon \cdot d} \quad (3.1)$$

C: concentration, A_{280} : absorption at 280 nm, ε : extinction coefficient

[$M^{-1} \text{ cm}^{-1}$], d: length of the cuvette (1 cm).

3.4 NMR spectroscopy

NMR data were acquired at 285 K/295 K on a Bruker Avance III 600-MHz spectrometer equipped with a z-pulse field gradient unit. Backbone ^1H , ^{15}N and ^{13}C resonance assignments were obtained by analyzing 3D CBCA(CO)NH and HNCBCA experiments. Side chain resonance assignments were obtained by analyzing HCCC(CO)NH, ^{15}N -TOCSY, HCCH-TOCSY and ^{15}N -, ^{13}C NOESY spectra (Sattler et al., 1999). Inter- and intra molecular proton distance restraints were obtained from peaks assigned in 2D-NOESY experiments. All spectra were processed with NMRPipe/NMRDraw software (Delaglio et al., 1995) and were analyzed with CARA (Bartels et al., 1995). Spectra used for the calculation were integrated using the batch integration method of the XEASY package.

3.5 NMR titration experiments

^{15}N -HSQC spectra were acquired using 300- μM ^{15}N -labeled protein samples to which the unlabelled peptide was added stepwise until either saturation or 5-fold excess was achieved. Measurements were performed at 285 or at 295 K.

3.6 Structure calculation/determination and refinement

Structures were calculated with CNS 1.1 (Brünger et al., 1998), using only unambiguously assigned restraints derived from NOESY experiments, coupling constants $^3J(\text{H}^{\text{N}}, \text{H}^{\text{A}})$ from HNHA spectra and hydrogen bonds measured from D_2O exchange experiments. The protocol for the calculation consists of two iterations of 1 and 200 structures respectively, using 100,000

cooling steps. All calculated structures were water refined and ranked based on minimum values of energy and violations. The water refinement protocol is a modification of the original protocol provided with Aria (Nilges et al., 1997), which uses all experimental restraints during the refinement process. In this way, the obtained structures are in better agreement with the experimental data, while retaining good Ramachandran values. Analysis of the quality of the lowest energy structures was performed using PROCHECK-NMR (Laskowski et al., 1996). Molecular images were generated using PyMOL (DeLano, 2002). The statistics obtained from the analysis are shown in Tables S1 and S2, in the results section.

3.7 Molecular dynamic simulations

Molecular dynamic simulations were performed with the Gromacs package (Hess et al., 2008). Prior the simulations we generated an extended model of each molecule with CNS, surrounded by a charged-equilibrated, periodic cubic water box. Then the system was energy-minimized and a short position-restrained molecular dynamics was performed to equilibrate the water molecules. Finally, a 40ns molecular dynamics in explicit solvent with Particle Mesh Ewald electrostatics was carried out. Calculated structures and the results of the molecular dynamic simulations were analyzed with PyMOL (DeLano, 2002).

3.8 Sequence alignments

Sequence alignments were performed using ClustalX (Thompson et al., 1997) and BoxShade 3.21. (http://www.ch.embnet.org/software/BOX_form.html).

3.9 Peptide synthesis

In brief, all peptides were synthesized using Fmoc-solid phase peptide synthesis using the 0.1 mmol scale, with a rink amide resin (Merck Chemicals), either manually or in a CEM Liberty1 microwave synthesizer.

After completion of the entire synthesis the N-terminal amino group was acetylated. The crude peptide was purified by RP-HPLC using a SunFire™ C18 Sephasil preparative column (Waters) with an ÄKTApurifier 10 (GE Healthcare Life Sciences), using a linear gradient of 10-40% acetonitrile and 0.05% TFA and an elution time of 20 min. Each peptide was analyzed by MALDI-TOF mass spectrometry and 2D homonuclear NMR spectroscopy.

3.10 Mass spectrometry

Protein and peptide masses were determined on a 4800 Plus MALDI TOF/TOF mass spectrometer and processed and analyzed with the Data Explorer 4.8 software. The samples were co-crystallized with CHCA (α -cyano-4-hydroxycinnamic acid) and spectra were acquired using the positive reflector method.

3.11 Isothermal titration calorimetry

ITC experiments were performed using a VP-ITC MicroCalorimeter (MicroCal) at 10 and 25 °C and the ITC isotherms were fit to the simplest model with MicroCal's ORIGIN software and in a low volume nano ITC calorimeter (TA instruments) and five different temperatures 5, 15, 20, 25 and 30 °C. Protein and peptide samples were dissolved in the same buffer and centrifuged, and degassed prior the ITC experiments. For the complexes with Smurf1, Smurf2 and YAP domains, mutants and pairs we used 20 mM Tris-HCl buffer (pH 7.2-7.4), 100 mM NaCl and 0.01% NaN₃. Control experiments were performed in the buffer used for the IM-MS experiments, which was 50 mM NH₄OAc pH 7.2. All protein concentrations were determined in a Shimadzu UVmini-1240 spectrophotometer and in a NanoDrop™ 2000 measuring the UV absorption of a denatured protein solution, using the predicted extinction coefficient. Peptide concentrations were determined spectrophotometrically and by amino acid analysis.

For each titration, a 20-fold concentrated peptide solution was injected into a cell containing 190 μL of protein solution. 15 injections of 3 μl were carried out per titration with a three-minute delay after each injection. Protein concentrations of 20-80 μM were used depending on the expected affinity range of each complex to yield a sigmoidal binding curve. Experiments were performed in triplicates and at different temperatures. Binding isotherms were analyzed using the software provided by TA instruments, assuming a single-binding site for the independent domains and a model of two independent binding sites. Baseline controls were acquired with buffer and pure peptide solutions.

3.12 Ion Mobility-Mass Spectrometry (IM-MS)

Traveling wave ion mobility mass spectrometry experiments were performed on a Synapt G1 HDMS mass spectrometer (Waters, Manchester, UK). Samples were placed on a 384-well plate refrigerated at 15°C and sprayed using a Triversa NanoMate® (Advion BioSciences) automated Chip-Base nanoelectrospray working in the positive ion mode. The instrument was calibrated over the 500-8000 Da m/z range using a cesium iodide solution. The software MassLynx 4.1 SCN 704 and Driftscope 2.1 were used for data processing. Samples containing the Smad7 peptide complexes with either Smurf1 WW1-WW2 or Smurf2 WW2-WW3 pairs or with the independent WW domains (final concentrations of 30-50 μM) were prepared in 50 mM NH_4OAc pH 7.2, immediately prior to the analysis. Spray voltage was set to 1.75 kV and delivery pressure to 0.5 psi. A reduction of the source pumping speed in the backing region (5.85 mbar) was set for optimal transmission of high mass non-covalent ions. Cone voltage, extraction cone and source temperature were set to 20 V, 1V and 20 °C respectively. Ions passed through a quadrupol mass filter to the IM-MS section of the instrument.

3.13 Buffers and Solutions

Bacterial culture reagents were purchased from Conda (Madrid, Spain) all other chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA) or Merck (Darmstadt, Germany). MiniPrep PCR-purification kits (MinElute PCR Purification Kit) were purchased from QIAGEN (Hilden, Germany).

LB medium (1l)		LB-agar plates	
Bactotryptose	10 g	Agar in autoclaved	
Yeast extract	5 g	LB medium	1.5% (w/v)
NaCl	10 g	Kanamycin	25 µg/ml
Kanamycin (25 mg/ml)	1 ml		
Autoclaved			
PBS buffer 10x (1l) pH 7.4		GST elution buffer (50 ml) pH 8	
NaCl	80 g	L-Glutathion	10 mM
KCl	2 g	Tris-HCl pH 8.0	50 mM
Na ₂ H PO ₄	14.4 g	NaCl	100 mM
KH ₂ PO ₄	2.4 g	Always prepare fresh	

SOC medium (500 ml)		Trace elements 100x (1l)	
Bactotryptose	10 g	EDTA	5 g
Yeast extract	2.5 g	FeCl ₃ × 6 H ₂ O	0.833 g
NaCl [5M]	1 ml	ZnCl ₂	84 mg
KCl [1M]	1.25 ml	CuCl ₂ × 2 H ₂ O	13 mg
MgCl ₂ [1M]	5 ml	CoCl ₂ × 6 H ₂ O	10 mg
MgSO ₄ [1M]	5 ml	H ₃ BO ₃	10 mg
Glucose [1M]	10 ml	MnCl ₂ × 6 H ₂ O	1.6 g
Autoclaved		Adjust pH to 7.5 after adding EDTA	
M9 medium 10x (1l)		M9 wash 10x (1l)	
Na ₂ H PO ₄	60 g	Na ₂ H PO ₄	60 g
KH ₂ PO ₄	30 g	KH ₂ PO ₄	30 g
NaCl	5 g	NaCl	5 g
¹⁴ NH ₄ Cl/ ¹⁵ NH ₄ Cl	5 g		

M9 medium 1x (1l)		¹³C-labelled M9 medium 1x (1l)	
M9 medium 10x	100 ml	M9 medium 10x	100 ml
Trace elements 100x	10 ml	Trace elements 100x	10 ml
Glucose 20%	20 ml	¹³ C6-glucose 20%	10 ml
MgCl or MgSO ₄ [1M]	1 ml	MgCl or MgSO ₄ [1M]	1 ml
CaCl ₂ [1M]	0.3 ml	CaCl ₂ [1M]	0.3 ml
Biotin [1mg/ml]	1 ml	Biotin [1mg/ml]	1 ml
Thiamin [1mg/ml]	1 ml	M9 medium 10x	100 ml
Kanamycin [1mg/ml]	1 ml	Trace elements 100x	10 ml

Ni²⁺ column buffer A		Ni²⁺ column washing buffer	
NaCl	150 mM	NaCl	1 M
Tris-HCl	20 mM	Tris-HCl	20 mM
Imidazole pH 8.0	10 mM	Imidazole pH 8.0	10 mM

Ni²⁺ column elution buffer		Tris-HCl buffer	
NaCl	150 mM	NaCl	100 mM
Tris-HCl	20 mM	Tris-HCl	20 mM
Imidazole pH 8.0	10 mM	NaN ₃	500 mM
EDTA	50 mM		Adjust pH

Phosphate buffer		Bis-Tris buffer	
NaCl	100 mM	NaCl	100 mM
Na H ₂ PO ₄	20 mM	Bis-Tris	20 mM
Na ₂ HPO ₄	20 mM	NaN ₃	500 mM
NaN ₃	500 mM		Adjust pH

Results Section

4. Phosphorylation by GSK3 switches Smad1 from binding transcriptional cofactors to binding ubiquitin ligases that trigger Smad1 degradation

4.1 Chapter Summary:

When directed to the nucleus by TGF- β or BMP signals, Smad proteins undergo CDK8/9 and GSK3 phosphorylations that mediate the binding of YAP and Pin1 for transcriptional action and of ubiquitin ligases Smurf1 and Nedd4L for Smad destruction. Here we demonstrate that there is an order of events – Smad activation first and destruction later—controlled by a switch in the recognition of Smad phosphoserines by WW domains in their binding partners. In the BMP pathway, Smad1 phosphorylation by CDK8/9 creates binding sites for the WW domains of YAP, and subsequent phosphorylation by GSK3 switches off YAP binding and adds binding sites for Smurf1 WW domains. Similarly, in the TGF- β pathway, Smad3 phosphorylation by CDK8/9 creates binding sites for Pin1 and GSK3 then adds sites to enhance Nedd4L binding. Thus, a Smad phosphoserine code and a set of WW domain code-readers provide an efficient solution to the problem of coupling TGF- β signal delivery to turnover of the Smad signal transducers.

4.2 Notes:

1. This work has been published in *Genes and Development* 2011, under the title: "A Smad action-turnover switch operated by WW domain readers of a phosphoserine code" by **Eric Aragón, Nina Goerner**, Alexia-Ileana Zaromytidou, Qiaoran Xi, Albert Escobedo, Joan Massagué and Maria J. Macias. Eric Aragón and Nina Goerner are joint first authors. A reprint of this work is attached in the appendix section. The work was accepted without the need of additional corrections and earned the cover page of the journal.

2. From the above mentioned paper, only the work corresponding to YAP and Smurf1 interactions with different fragments of Smad1, -which I have carried out with the supervision of Dra. M. J. Macias- is included in the present chapter. The sections that correspond to Pin1 and Nedd4L complexes with Smad3 fragments were included in the thesis defended by Nina Goerner.

3. Binding experiments using full-length proteins and mammalian cells were performed by Drs S. Gao, A.-I. Zaromytidou and Q. Xi, under the supervision of Dr. J. Massagué. They are included in the text to provide the grounds for the structural work presented here.

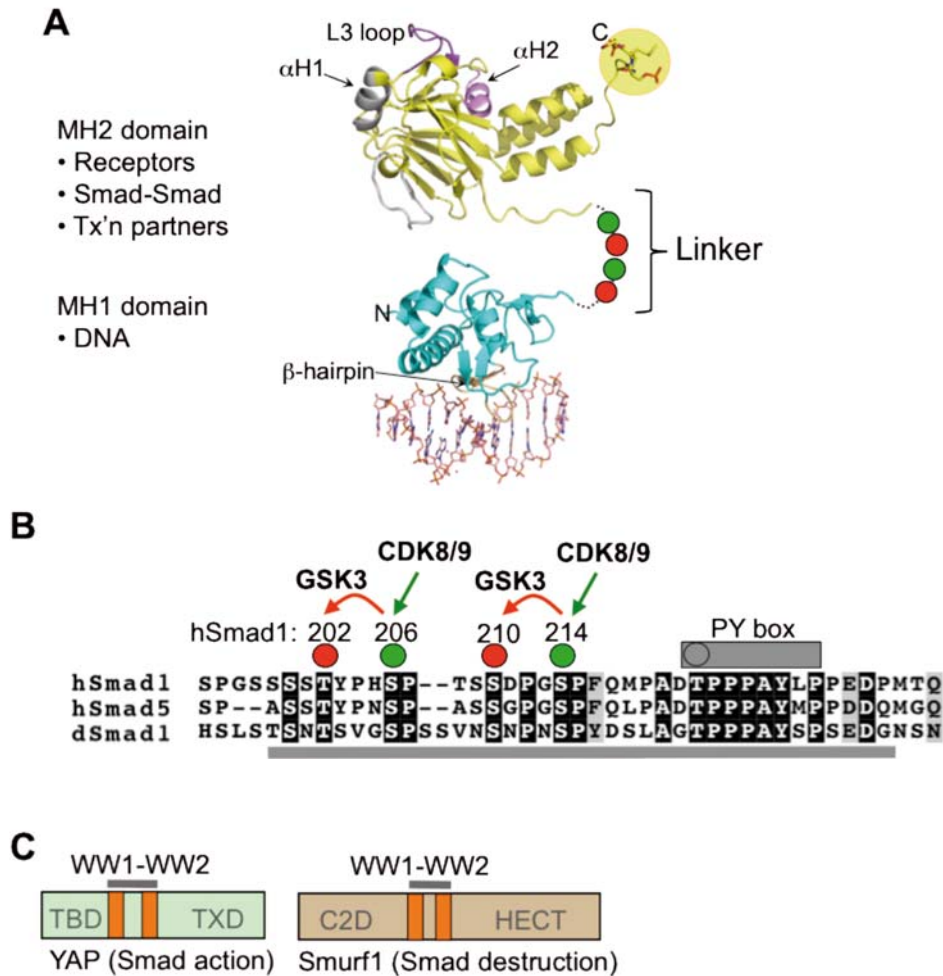
4.3 Different phosphorylation sites in Smad1 are recognized by YAP and Smurf1

4.3.1 Introduction

Smad proteins consist of a globular amino-terminal MH1 (Mad Homology 1) domain with DNA-binding activity, a carboxy-terminal MH2 domain that mediates key protein-protein interactions, and an inter-domain linker region with a conserved cluster of phosphorylation sites adjacent to a PY motif (Figure 4A, B) (Shi and Massagué, 2003). The presence of the PY site in the linker and the observation that a few partners of Smad proteins contain WW domains prompted us to investigate whether the PY site and motifs located in its proximity were responsible for the interactions (Figure 4C). Furthermore, since the proteins that interact with the linker were involved in a wide range of processes including transcription activation and protein degradation, we set up to investigate whether some phosphorylation sites could define different patterns that are recognized by some proteins but not by others. If these hypotheses were true, then the presence or absence of these phosphorylated residues could establish a *code* only interpreted by its specialized readers, thus defining a switch that enhances activation or protein degradation according to the cellular needs.

It has been previously described that phosphorylation of these sites in the linker follows BMP- and TGF- β -driven C-terminal phosphorylation and nuclear translocation of Smads (Alarcón et al., 2009; Fuentealba et al., 2007; Sapkota et al., 2007). In Smad1, CDK8/9 phosphorylate S206 and S214, which prime T202 and S210, respectively, for phosphorylation by GSK3.

Figure 4: Schematic representation of the Smad protein domains and their main functions



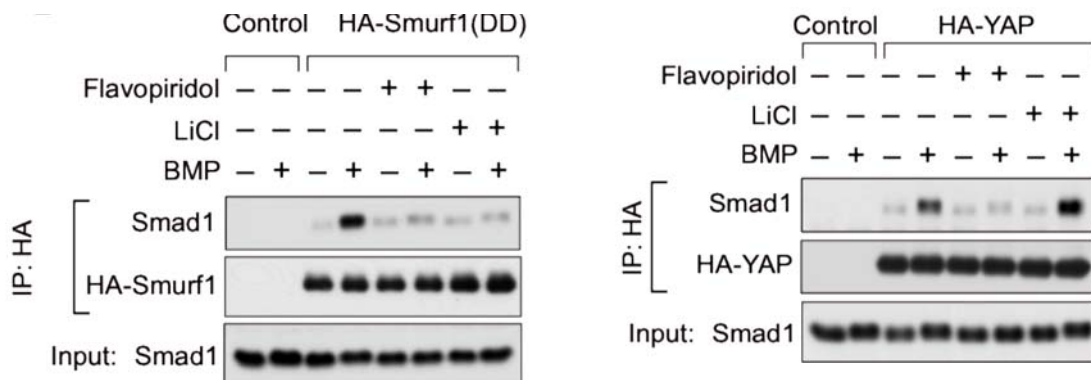
(A) The MH1 domain (cyan) contains a β -hairpin that mediates binding to dsDNA (orange) (PDB code: 1MHD) (Shi et al., 1998). The MH2 domain (yellow) binds to the type I TGF- β receptor, which involves the L3 loop (magenta), to Smad4 via the phosphorylated C-terminus (highlighted) and the α -helix 1 (grey), and to various DNA-binding cofactors and histone-modifying enzymes (PDB code: 1KHX) (Wu et al., 2001). The inter-domain linker region (dotted line) contains CDK8/9 and GSK3 phosphorylation sites represented by green and red circles, respectively.

(B) Sequence alignment of the linker region of human Smads 1 and 5 and Drosophila MAD (dSmad1) proteins, with conserved residues highlighted. The conserved CDK8/9 sites (green) and CDK8/9-primed GSK3 sites (red), and the PY box are shown. The Smad1 (199-232) segment used in this work is underlined.

(C) The domain composition of Smurf1 and YAP proteins and the regions that mediate binding to linker-phosphorylated Smad1 are indicated.

To dissect this process our collaborators tested the effect of pharmacological inhibitors of CDK8/9 and GSK3 in human embryonic kidney HEK293 cells expressing epitope-tagged Smurf1 or YAP constructs. The results of these experiments were that incubation of the cells with BMP rapidly induced the formation of Smad1-YAP and Smad1-Smurf1 complexes. The CDK8/9 inhibitor flavopiridol, which inhibits all BMP-induced linker phosphorylations (Alarcón et al., 2009), prevented the formation of both complexes. Addition of LiCl, which inhibits GSK3 site phosphorylation (Fuentealba et al., 2007), also prevented the Smad1-Smurf1 interaction. Interestingly, LiCl did not inhibit but rather increased the level of Smad1-YAP complex. In brief, these results suggested that the formation of YAP-Smad1 complex in response to BMP requires CDK8/9 but not GSK3, whereas the formation of Smurf1-Smad1 complex requires both kinase activities (Figure 5).

Figure 5: IP experiments



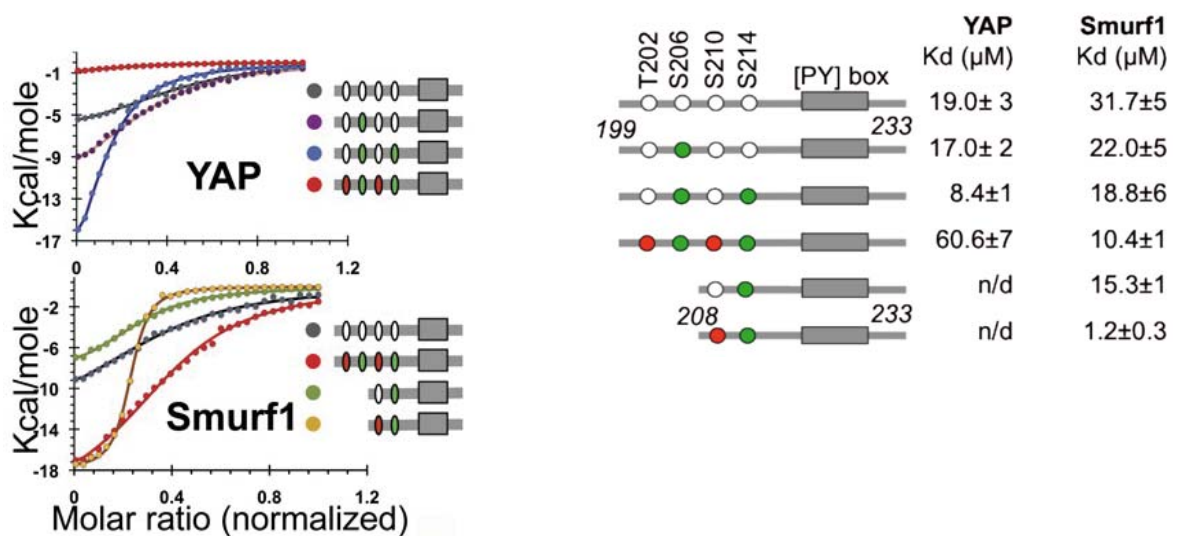
Left: BMP-dependent formation of a complex between HA-Smurf1(DD) and endogenous Smad1 in HEK293 cells, and effects of flavopiridol and LiCl on the formation of this complex.

Right: BMP-dependent formation of a complex between HA-YAP and endogenous Smad1 in HEK293 cells, and effects of flavopiridol and LiCl on the formation of this complex.

4.3.2 Binding affinities determined using isothermal titration calorimetry

To rationalize these binding preferences, we have characterized and quantified the effect of the possible phosphorylations of Smad1 in the interactions with Smurf1 and YAP proteins. To obtain this information we have combined the use of NMR and isothermal titration calorimetry (ITC) binding assays to investigate the interaction between recombinant WW1, WW2 and WW1-WW2 segments of Smurf1 and YAP proteins and Smad1 linker phosphopeptides. We tested versions of the Smad1 199-233 linker region with no phosphorylation, with phosphorylation at the CDK8/9 sites S206 and S214, or with additional phosphorylation at the GSK3 sites T202 and S210 (Figure 6).

Figure 6: ITC values and peptides



Left: ITC curves for the binding of YAP and Smurf1 WW1-WW2 segments to Smad1 synthetic peptides.

Right: Synthetic Smad1 peptides and their affinity for recombinant WW1-WW2 segments of YAP and Smurf1. Colored circles denote phosphorylation of the residues; white circles denote unphosphorylated residues and the gray rectangle represents the PY motif.

YAP WW₁-WW₂ segment bound unphosphorylated Smad1 peptide with $K_D=19.0\pm 3 \mu\text{M}$ and CDK8/9-phosphorylated peptide with $K_D=8.4\pm 1 \mu\text{M}$. Notably, this gain in affinity was fully erased by phosphorylation at the GSK3 sites ($K_D=60.6\pm 7 \mu\text{M}$) (Figure 6). In contrast, the affinity of the Smurf1 WW₁-WW₂ segment for the Smad1 peptide was increased by phosphorylation at the CDK8/9 sites, and further increased by phosphorylation at the GSK3 sites. Further refinement of the Smurf1-Smad1 interaction revealed a strong preference of Smurf1 for pS214 over pS206, achieving the highest affinity ($K_D=1.2\pm 0.3 \mu\text{M}$) with a Smad1 208-233 peptide containing pS210 and pS214. To confirm these observations, mutations of S210 or S214 or of both residues to alanine were introduced in full length Smad1 protein and overexpressed in HEK293 cells. Under these conditions our collaborators observed that individual or double mutations reduced the Smad1-YAP interaction in cells, but mutation of both residues inhibits the Smad1-Smurf1 binding (Figure 7).

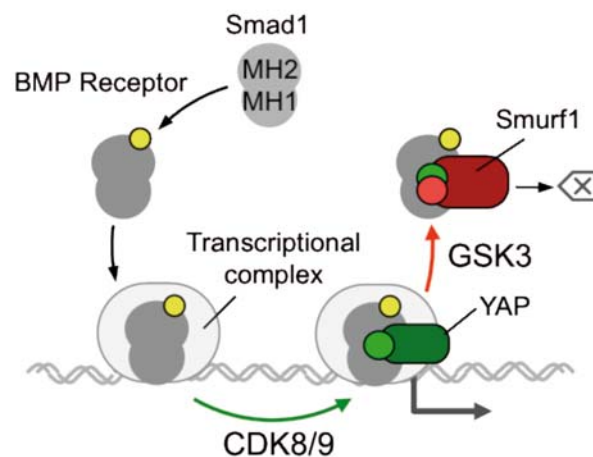
Figure 7: IP experiments with Smad1 mutants



Effect of alanine mutations in the PY box and the indicated phosphorylation sites on the ability of Flag-tagged Smad1 constructs to bind HA-Smurf1(DD) in HEK293 cells.

Collectively these results suggest that CDK8/9-mediated phosphorylation of the Smad1 linker creates binding sites for competing YAP and Smurf1 WW1-WW2 domains, and GSK3 switches this balance in favor of Smurf1 binding and at the expense of the YAP interaction (**Figure 8**). Guided by these observations we set to characterize the interactions of Smurf1 and YAP WW1-WW2 pairs with a CDK8/9 phosphorylated Smad1 peptide fragment and to compare it with the doubly phosphorylated CDK8/9 and GSK3 Smad1 using multidimensional NMR in solution.

Figure 8: Schematic summary of the Smad action-turnover switch operated by CDK8/9 and GSK3 in combination with YAP and Smurf1



After activation by the BMP receptor R-Smads are phosphorylated at the C-terminus. Once they are phosphorylated they can form a transcriptional complex and bind to the target DNA. Additional phosphorylation by CDK8/9 creates binding sites for transcriptional cofactors, like YAP. Phosphorylation by GSK3 creates binding sites for ubiquitin ligases like Smurf1.

4.3.3 Structure of the Smurf1 WW1-WW2 pair bound to the doubly phosphorylated Smad1 linker

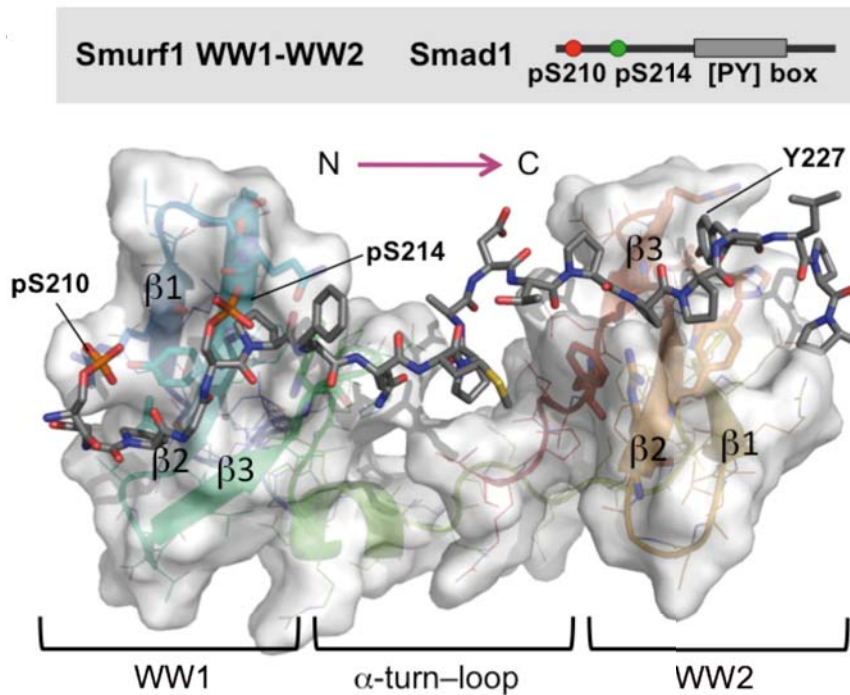
We used NMR spectroscopy to calculate the structure of the Smurf1 WW1-WW2 segment bound to the Smad1 linker peptide (208-233) di-phosphorylated at S210 and S214 in solution. Triple resonance NMR spectroscopy was applied to assign the WW1-WW2 pair in this complex whereas homonuclear and half-filter spectra were used to assign the Smad1 peptide and its contacts with Smurf1 (Figure 9 and S1). To facilitate the assignment of the pair, samples of the independent domains were also prepared and assigned. Since the structural analysis of complexes requires the availability of protein samples in high quantities and purity, I had to optimize the production of the proteins and also the purification protocols until I could obtain reproducible conditions.

In brief, individual domains were prepared as GST fusion proteins and obtained in the soluble fractions of the cultures, when either using LB or in minimal medium. Moreover, after the cleavage with the Tobacco Etch Virus (TEV) protease, samples containing the WW domains remained soluble and were concentrated up to 1 mMolar for the structural studies. However, the construct with the WW1-WW2 pair was more difficult to obtain, due to its tendency to form aggregates and precipitates. To solve the problem I first introduced a point mutation (Cys to Met) in the linker connecting both WW domains. This change avoids the presence of dimers via the formation of a disulfide bridge. To test whether or not such a change could affect the structure and binding properties of the WW1-WW2 pair we performed binding controls using wild type and mutant protein. Indeed, titrations experiments followed by

NMR and ITC, revealed that the mutation did not affect the binding. Unfortunately, even in the presence of the mutation, the protein solubility was not improved. I was finally able to avoid the formation of insoluble aggregates by preparing the complex at low protein concentration and concentrate the samples for the structural studies afterwards. Thanks to the stabilization of the protein by the formation of the complex, we could finally acquire the full set of NMR experiments required for the assignment of the complex and the determination of its structure in solution.

In the complex, each of the WW domains adopts the typical triple-stranded anti-parallel β -sheet fold, even in the case of WW1 that lacks the first highly conserved tryptophan. The WW domains do not contact each other but each contacts a portion of the linker. This arrangement provides enough freedom for the WW domains to adopt an anti-parallel orientation forming a continuous binding surface that smoothly cradles the phosphorylated Smad1 linker. The Smad1 linker adopts an extended conformation with the di-phosphorylated T208-P215 segment bound to the WW1 domain and the PY motif bound to the WW2 domain (Figure 9). The segment between P215 and the PY motif forms a turn defined by interactions between F217 and A220 backbone atoms. The ten-residue segment connecting the two Smurf1 WW domains adopts an alpha turn helical structure in its first half (Figure 9). This configuration allows access to CKIP1 (casein kinase 2-interacting protein-1), a protein that binds to this region to enhance the Smurf1-Smad1 interaction (Lu et al., 2008).

Figure 9: NMR-model of the complex between the human Smurf1 WW1-WW2 pair (residues 232-314) and the 208-233 segment of the Smad1 linker di-phosphorylated at S210 and S214



Smurf1 is shown as a semitransparent surface with all elements of secondary structure represented. The Smad1 peptide is shown with a stick representation, with the backbone colored in gray. There are several relative orientations of the WW domains that satisfy all experimental NMR restraints (shown in Supplementary Figure S1) and due to this, we call this complex NMR-model.

To facilitate the presentation of the different complexes in the text, I will use the one-letter amino acid notation for Smad1 residues and the three-letter notation for residues in Smurf1. The Smurf1 WW1 domain binds the Smad1 pS210 residue through contacts with Tyr251, Arg243 and Leu253 side-chains. The Tyr251 hydroxyl and the Arg243 guanidinium groups jointly coordinate the phosphate group of pS210. pS214 also contacts Tyr251, and the phosphate group is coordinated by the hydroxyl of Thr245 and the side-chains of Gln247 and Gln249. P215 is packed parallel to the aromatic ring of Trp262 and P212 is

sandwiched between the Leu253 and Ser260 side-chains in a cavity perpendicular to the β -sheet (Figure 10). The pS210-D211-P212-G213 segment forms a turn, favored by a D211-P212 *cis* bond, whereas pS214-P215 is in *trans*. Single alanine mutations of Arg243, Gln247 or Gln249 decreased the affinity to K_D values of approximately 30 μ M, confirming the importance of these residues in the interaction of WW1 with the pS210 and pS214 phosphate groups.

The Smurf1 WW2 domain binds to the PY motif in a manner similar to canonical Group 1 WW complexes (Macias et al., 2002). P224 and P225 contact Tyr297 and Phe308, respectively, and Y227 binds between His301 and Arg304 (Figures 11, S2). The six residues after the tyrosine in the PY motif fold over the first strand of WW2. Abundant contacts are observed between P229 and P230 and His301 and Glu287 respectively. The side chain of E231 points towards the Tyr297 hydroxyl and shows contacts with the Arg289 side chain.

Figure 10: Smurf1 WW1 bound to the Smad1 pS210/pS214 site

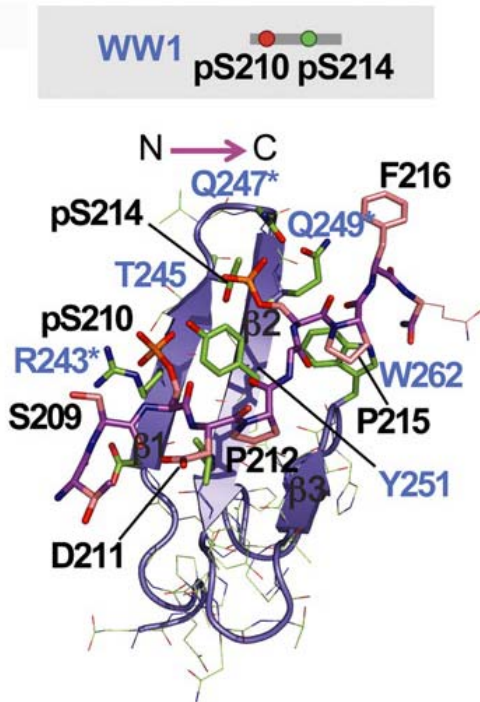


Figure 11: Smurf1 WW2 domain bound to the PY box

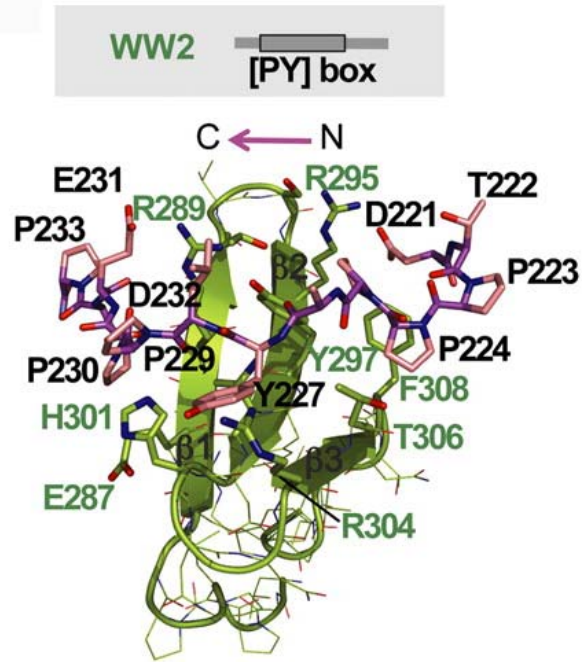


Figure 10 shows a detailed view of the refined structure of Smurf1 WW1 domain (slate) bound to the di-phosphorylated pS210/pS214 region of the Smad1 linker. Key residues in Smad1 (black) and Smurf1 (blue) are indicated. Asterisks, three residues that when jointly mutated to alanine decreased the binding affinity of the complex by approximately 25-fold. A superimposition of the calculated structures is shown as Supplementary Figure S2 left panel.

Figure 11 shows a detailed view of the refined structure of the Smurf1 WW2 domain (green) bound to the PY motif of Smad1. Key residues in Smad1 (black) and Smurf1 (green) are indicated. A superimposition of the calculated structures is shown as Supplementary Figure S2 right panel.

4.3.4 Structure of the Smurf1 WW1-WW2 pair bound to the mono- phosphorylated Smad1 linker

According to the affinity values we have obtained by ITC measurements and also by the experiments performed by our collaborators using full-length proteins in cells, Smurf1 WW1-WW2 pair binds stronger to Smad1 sequence when it is phosphorylated at Ser 210 and Ser 214 than when is only phosphorylated at position 214. With the aim of finding out the rules that specify Smurf1 preferences towards the di-phosphorylated Smad1 sequence we have also solved the structure of the Smurf1 WW1-WW2 segment bound to the Smad1 linker mono-phosphorylated at pS214 (Figure S2 middle panel). As we have expected most of the contacts between the two molecules are like those in Smurf1 WW1-WW2 bound to the pS210/pS214 di-phosphorylated peptide. However, in the mono-phosphorylated complex we observe that the D21 -P215 fragment is bound differently (compare Figure 11 with 12). S210 is less ordered than in the phosphorylated state and only weak contacts are observed between D211 and Arg243 side-chains. We attribute to these weaker contacts the intermediate affinity of the WW1-WW2 domain for the Smad1 linker mono-phosphorylated at S214.

Figure 12: A comparison of Smurf1 WW1 domain bound to either Smad1 pS214 or to pS210 and pS214 sites

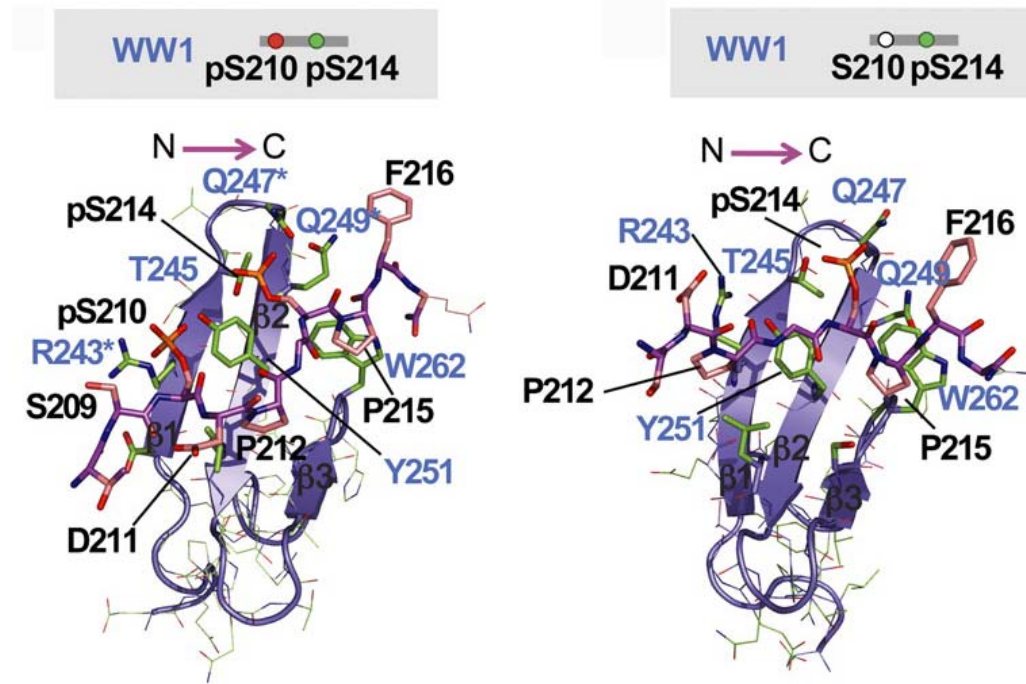
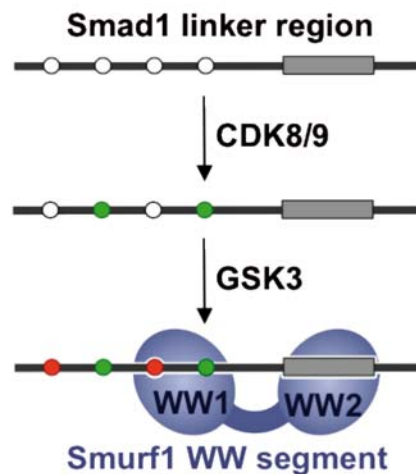


Figure 12 shows a detailed view of the refined structure of the Smurf1 WW1 domain (slate) bound to the mono-phosphorylated pS214 region of Smad1 linker. A superimposition of the calculated structures is shown as S2 middle panel. For comparison, Figure 11 is also represented next to it.

In conclusion, formation of the Smurf1-Smad1 complex involves recognition of the Smad1 PY motif by the Smurf1 WW2 domain and of the Smad1 GSK3 and CDK phosphorylated sites by the WW1 domain (Figure 13).

Figure 13: Schematic representation of the mode of binding of Smurf1 to the Smad1 linker region



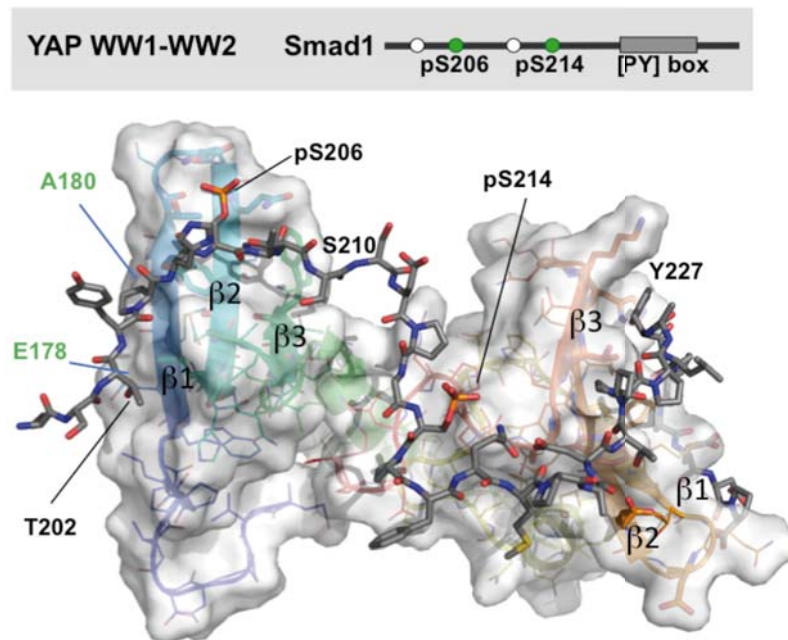
Smad proteins are phosphorylated at CDK8/9 sites (*green circle*) in the MH1-MH2 interdomain linker region represented in grey. This phosphorylation creates high-affinity binding sites for transcriptional partners, such as YAP and probably others, thus achieving peak transcriptional action. Phosphorylation by CDK8/9 also primes the Smads for GSK3-mediated phosphorylation (*red symbol*) at the -4 position, which favors the binding of ubiquitin ligases Smurf1 (BMP pathway).

4.3.5 Structure of the YAP WW pair bound to the Smad1 linker

Alanine mutations in either the first or in the second WW domain in YAP WW1-WW2 pair almost completely abolished the interaction of overexpressed YAP and Smad1 in transiently transfected human cells, suggesting that both domains are essential for this interaction (Figure S3). Given the high affinity of the YAP WW1-WW2 module for a Smad1 (199-233) linker peptide phosphorylated at pS206 and pS214 (the CDK8/9 sites, refer to Figure 6), we solved the structure of this complex first. We used double and triple labeled WW1-WW2 samples (YAP 163-266 segment) to assign the protein resonances in combination with filtered and homonuclear experiments to obtain the chemical shifts of the bound peptide and the contacts between both molecules. We also used independent domains to assist in the assignment. In contrast to the problems we had when preparing the pair of WW1-WW2 domains from Smurf1, all proteins corresponding to the YAP WW domains (isolated domains and also the domain pair) were highly soluble and behave well in the presence or absence of the peptides. However, to avoid the presence of potential aggregates we decided to use the same approach as that used with Smurf1 complexes and prepare the YAP-Smad1 complex using a diluted protein sample and concentrate the mixture for the structural studies. In the complex, the 25-residue connector between the two WW domains adopts a helix-loop-helix structure, as determined on the basis of the detected nuclear Overhauser effects (NOEs) and carbon chemical shift analysis. Several contacts are present between each WW domain and this connector, but these contacts do not prevent the WW domains from adopting an optimal orientation for interactions with the Smad1 linker. No contacts were observed between the WW domains. In the complex between the WW1-WW2 pair and the Smad1 34 residue-peptide both WW domains adopt the canonical fold and participate in the interaction with Smad1 (Figures 14 and Supplementary Figure S4).

The WW1 domain contacts the pS206 region and the WW2 domain contacts the PY motif. The Smad1 pS206 and P207 side-chains are accommodated in the aromatic cavity formed by Tyr188 and Trp199 in the YAP WW1 domain (Figure 15). The pS206 phosphate group is at a hydrogen bond distance from the hydroxyl groups of Thr182 and Tyr 188 and the Gln186 side chain. Trp199 is also involved in a network of contacts with residues comprised between P207 and S209. The structure of the WW2 domain bound to the Smad1 PY motif resembles that of Smurf1 WW2 bound to this region. The interaction is well defined, involving eight Smad1 residues between D221 and D232 and 9 out of 13 residues on the WW2 domain surface (Figure 16).

Figure 14: Complex between the human YAP WW1-WW2 pair and the 199-233 segment of the Smad1 linker



NMR-model of the complex between the human YAP WW1-WW2 pair (residues 163-266) and the 199-233 segment of the Smad1 linker di-phosphorylated at S206 and S214. YAP is shown as a semitransparent surface and Smad1 as gray sticks.

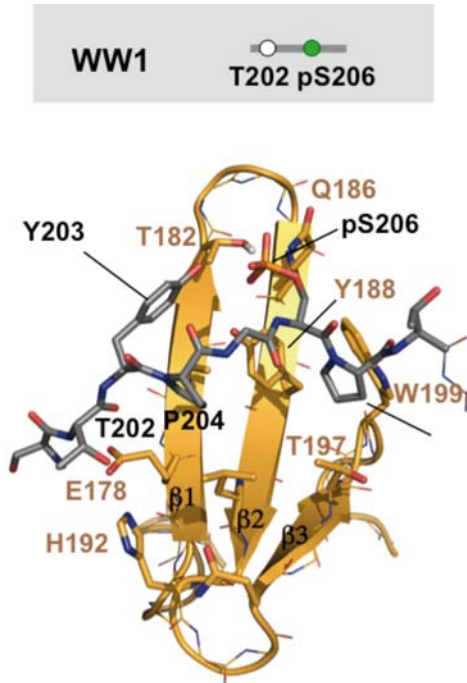
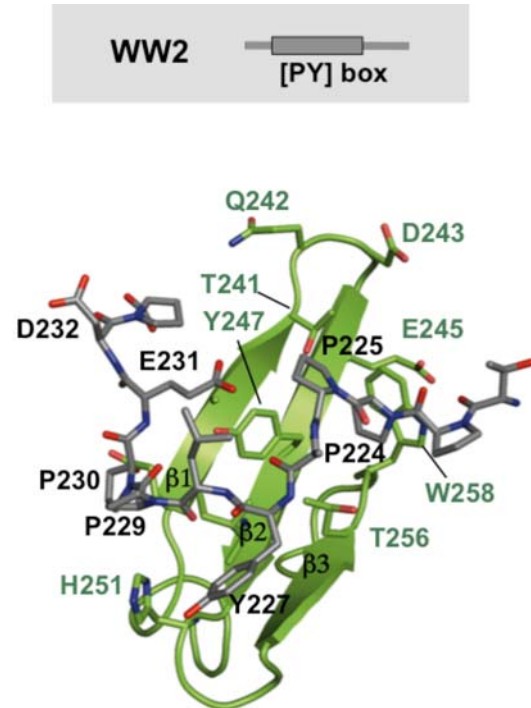
Figure 15: View of YAP WW1 bound to pS206 Smad1 site**Figure 16: View of YAP WW2 bound to PY Smad1 site**

Figure 15 shows a detailed view of the refined structure of the YAP WW1 domain (gold) bound to the mono-pS206 phosphorylation site of Smad1 (gray). Key residues in Smad1 (black) and YAP brown are indicated. A superimposition of the calculated structures is shown as Supplementary Figure S4 left panel.

Figure 16 shows a detailed view of the refined structure of the YAP WW2 domain (green) bound to the PY motif region of Smad1 (gray), with the key residues indicated. A superimposition of the calculated structures is shown as Supplementary Figure S4 right panel.

We also observed NOEs from the E231 side-chain with T241 and Tyr247 and from D232 to Gln242, suggesting the presence of inter-molecular salt-bridges between these residues. Thus, formation of the YAP-Smad1 complex involves recognition of the Smad1 PY and the CDK phosphorylated site pS206 by the WW2 and WW1 domains, respectively.

4.3.6 Understanding the preferences of YAP-Smad1 interactions: The CDK8/9 phosphorylated Smad1 linker avoids YAP binding when is also phosphorylated by GSK3 at positions

We analyzed the interactions between the YAP WW1-WW2 pair and Smad1 linker peptides containing GSK3 phosphorylation sites with the aim of understanding why the presence of these sites reduces the affinity of the interaction. However, NMR-based titrations with a peptide containing pT202, pS206, pS210 and pS214 requires a 4-5 fold peptide excess to induce chemical shift changes in YAP WW1-WW2 pair (Figure S5), corroborating the weak interaction measured by ITC and precluding the determination of the complex structure. In a complex of YAP WW1-WW2 domains with a peptide containing pT202, pS206 pS214 and the PY site, but no phosphorylation at S210, the NOEs detected from the N-terminal end of the peptide to the WW1 domain of YAP were weak and the structure is only defined for the P204-pS206-P207 site (Figure 17, S5B). In particular the Glu178 side-chain and the His192 ring in the WW1 domain fail to contact the methyl of T202. Y203 and P204 are only partially ordered.

Figure 17: View of YAP WW1 bound to pS206 and pT202/pS206 Smad1 site

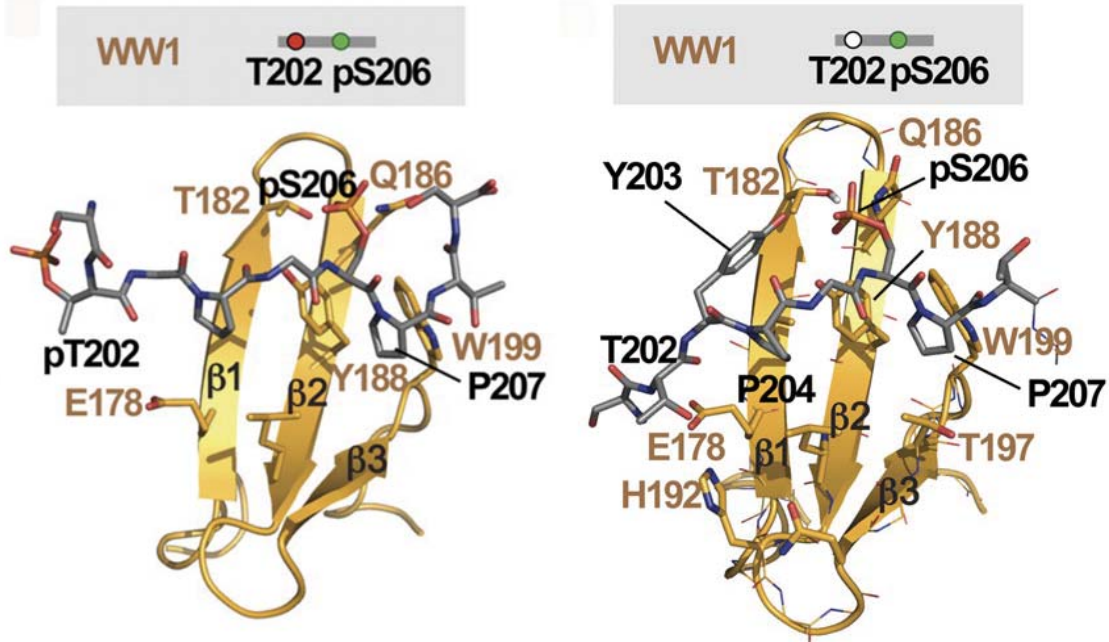


Figure 17 shows a detailed view of the refined structure of the YAP WW1 domain (gold) bound to the di-phosphorylated pT202, pS206 region of Smad1 linker. The monophosphorylated complex (Figure 15) is also represented to facilitate the comparison of both binding modes.

Thus according to this data, the presence of a phosphate group in T202 destabilizes the interaction of Smad1 with the YAP WW1 domain and the presence of phosphates at both T202 and at S210 drastically reduces the interaction between YAP and Smad1 linker.

To rationalize the different binding preferences of the Smurf1 and YAP complexes we compared the charge distribution on the surfaces of Smurf1 and YAP WW1 domains. Both WW1 domains contain Gln residues in the surroundings of pS214 and pS206 respectively but the positively charged patch of Smurf1 WW1 that interacts with pS210 (Figures 18, 19) is absent in YAP.

Figure 18: Charge distribution of Smurf1 WW1 bound to pS214 Smad1 site

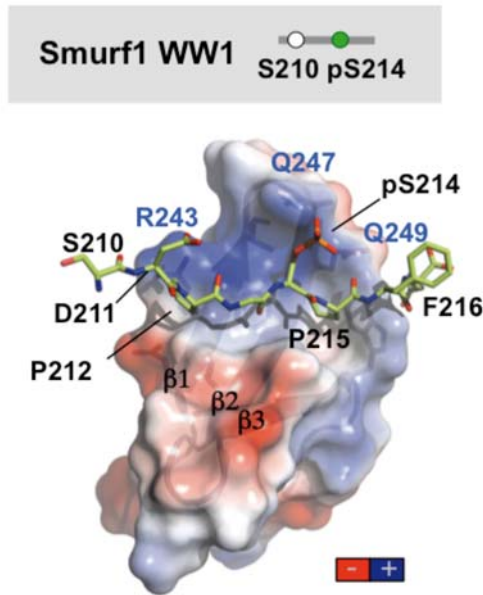
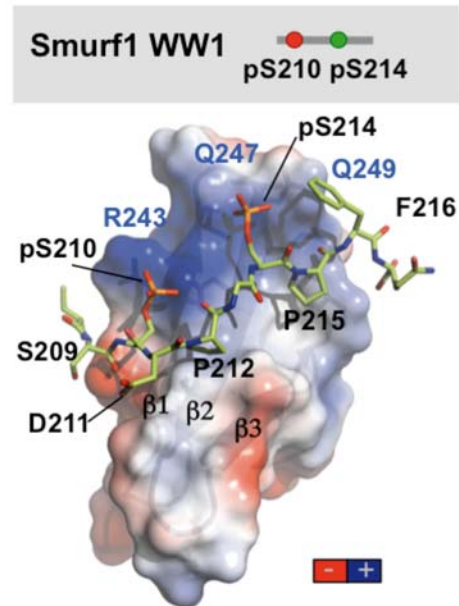


Figure 19: Charge distribution of Smurf1 WW1 bound to pS210-pS214 Smad1 site



Figures 18 and 19 show the charge distribution on the surface of Smurf1 WW1 domain in complex with the Smad1 linker mono-phosphorylated at S214 (left) or di-phosphorylated at S210 and S214 (right). Negatively charged patches are shown in red and positively charged patches in dark blue respectively. Smurf1 WW1 is shown as a semitransparent surface and Smad1 as green sticks. Key residues in Smad1 (black) and Smurf1 WW1 (blue) are shown.

Instead, YAP contains a negatively charged region suited to interact with T202 but incompatible with the presence of a phosphate group at T202 (Figure 20, 21).

Notably, in the complex of YAP WW1-WW2 the Smad1 linker segment between residues S210 and D221 runs across the inter-WW connector with three prolines (P212, P215 and P219) in *trans* (refer to Figure 14). The *trans* configuration of the D211P and pS214P bonds favors the formation of two β -turns that facilitate the interaction of the Smad1 pS206 site with YAP. This feature likely explains the higher affinity of YAP for the pS206/pS214 di-

phosphorylated peptide ($8.4 \pm 1 \mu\text{M}$) compared to its affinity for the pS206 mono-phosphorylated peptide ($17.0 \pm 2 \mu\text{M}$).

Figure 20 Charge distribution of YAP WW1 bound to pT202-pS206 Smad1 site

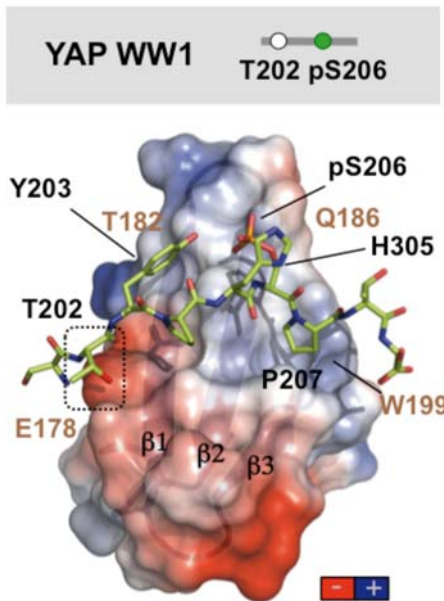
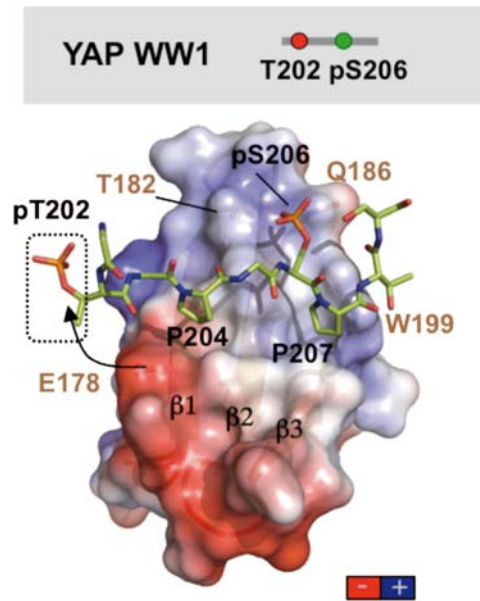


Figure 21: Charge distribution of YAP WW1 bound to pT202-pS206 Smad1 site

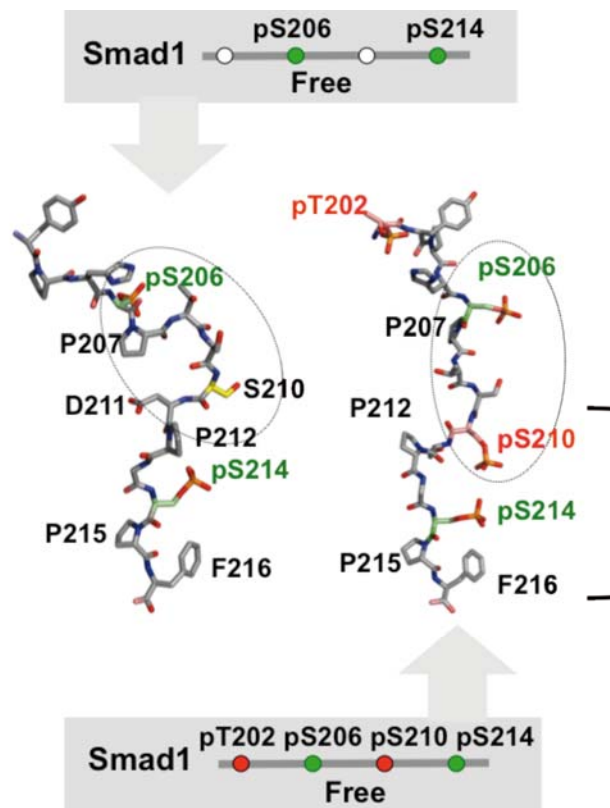


Figures 20 and 21 show the charge distribution on the surface of YAP WW1 domain in complex with the Smad1 linker mono-phosphorylated at S206 (left) or di-phosphorylated at T202 and S206 (right). The YAP WW1 domain is shown as a semitransparent surface and with the same orientation as in Figure 3. The position of T202 is shown in a box. The conformational change observed in pT202 is represented with an arrow.

The negative effect of pS210 on the YAP-Smad1 interaction observed by ITC and NMR titration experiments could arise from a conformational change in the Smad1 fragment forced by electrostatic repulsion between the pS210 phosphate group and the negatively charged D211. Since we cannot determine the structures of these complexes we decided to apply molecular dynamic simulations with the help of NMR based titrations to test the potential impact of this phosphate in peptides containing pS214, pS210 and pS206. According to

the simulations it seems that pS210 favors an extended conformation without the β -turn centered at D211-P212 (Figure 22) that would decrease the likelihood of WW1 interacting with the pS206 site.

Figure 22: Molecular simulations performed on Smad1 peptides

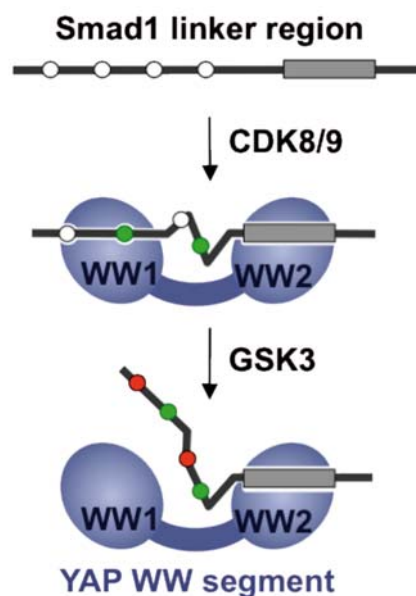


Molecular simulations performed on two peptides corresponding to Smad1 phosphorylated at S206 and S214 (left), or at T202, S206, S210 and S214 (right), Key residues are labeled.

Collectively these observations suggest that phosphorylation of Smad1 by CDK8/9 creates a binding site for the YAP WW1 domain in pS206 and the downstream all-*trans* configuration imposed by D211P and pS214P favors this binding interaction. GSK3 phosphorylation of the Smad1 linker at T202 and,

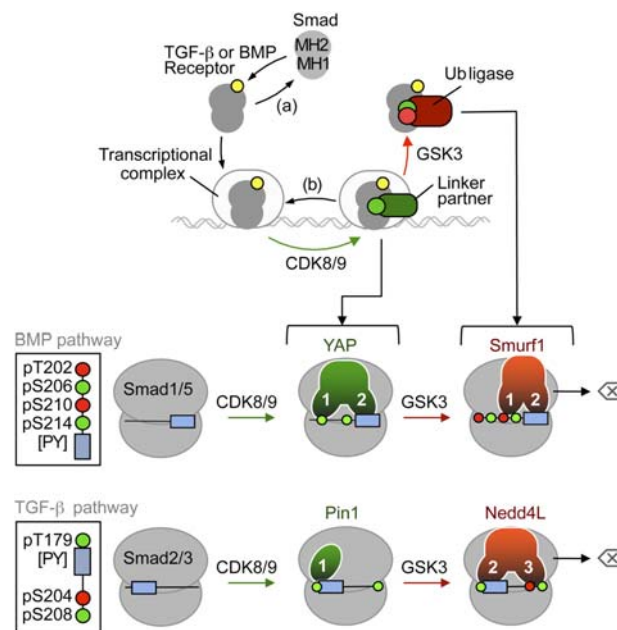
particularly, at S210 create a conformation that avoids recognition by the YAP WW1 domain (Figure 23) while favoring recognition by the Smurf1 WW1 domain. A representation of YAP and Smurf1 preferences for Smad1 recognition is shown in the general Scheme of page 76. The interaction of Nedd4L and Smad3 is also included in the Scheme, to show that the mechanisms described in this chapter apply to both BMP and TGF beta signaling cascades.

Figure 23: Schematic representation of the mode of binding of YAP to the Smad1 linker region



Smad proteins are phosphorylated at CDK8/9 sites (*green circle*) in the MH1-MH2 interdomain linker region represented in grey. This phosphorylation creates high-affinity binding sites for transcriptional partners, such as YAP and probably others, thus achieving peak transcriptional action. Phosphorylation by CDK8/9 also primes the Smads for GSK3-mediated phosphorylation (*red symbol*) at the -4 position, which favors the binding of ubiquitin ligases Smurf1 (BMP pathway) and at the same time prevent the interaction with coactivators.

**The Smad action-turnover switch in the BMP and TGF- β pathways:
pSer codes and WW-domain code readers**



Top panel, schematic summary of the Smad action-turnover switch in the BMP and TGF- β pathways. Following receptor-mediated phosphorylation (*yellow circle*) Smad proteins translocate to the nucleus and assemble transcriptional complexes, which are phosphorylated at CDK8/9 sites (*green circle*) in the MH1-MH2 interdomain linker region. This phosphorylation creates high-affinity binding sites for transcriptional partners, such as YAP in the case of the BMP mediator Smad1, and Pin1 in the case of the TGF- β mediator Smad3, and probably others, thus achieving peak transcriptional action. Phosphorylation by CDK8/9 also primes the Smads for GSK3-mediated phosphorylation (*red symbol*) at the -4 position, which favors the binding of ubiquitin ligases Smurf1 (BMP pathway) and Nedd4L (TGF- β pathway), leading to proteasome-dependent degradation of Smad molecules that participate in transcription

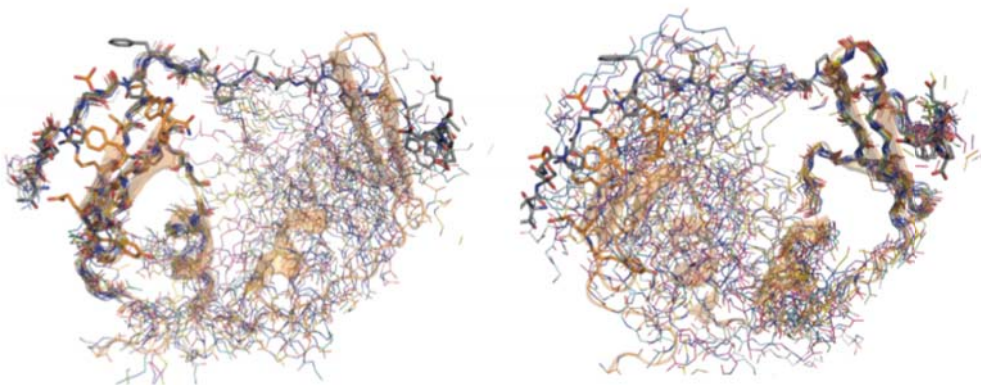
(*erase symbol*). Alternatively, C-terminal Smad phosphatases (a) and linker phosphatases (b) reverse these phosphorylation states.

Bottom panels, schematic of the Smad linker phospho-amino acid codes (*insets*) and WW-domain code readers. The conserved CDK8/9 phosphorylation sites (*green circles*) and GSK3 sites (*red circles*) are located at the indicated positions relative to the PY box (*slate box*). Amino acid positions correspond to Smad1 and Smad3. In the BMP pathway, the YAP WW1 domain binds to pS206 in Smad1, as long as p210 is not phosphorylated. The Smurf1 WW1 domain binds with higher affinity to the pS210-pS214 motif. The WW2 domains bind the [PY] motif. In the TGF- β pathway, the sole WW domain of Pin1 binds the pT179[PY] motif, as does the WW2 domain of Nedd4L. However, the Nedd4L WW3 domain increases the binding affinity by recognizing the pS204-pS208 motif. See the text for additional details and citations on the known roles of these WW-domain proteins in Smad signal transduction.

Supplementary Figures:

Figure S1 Refined structures of Smurf1 WW1-WW2 complexes

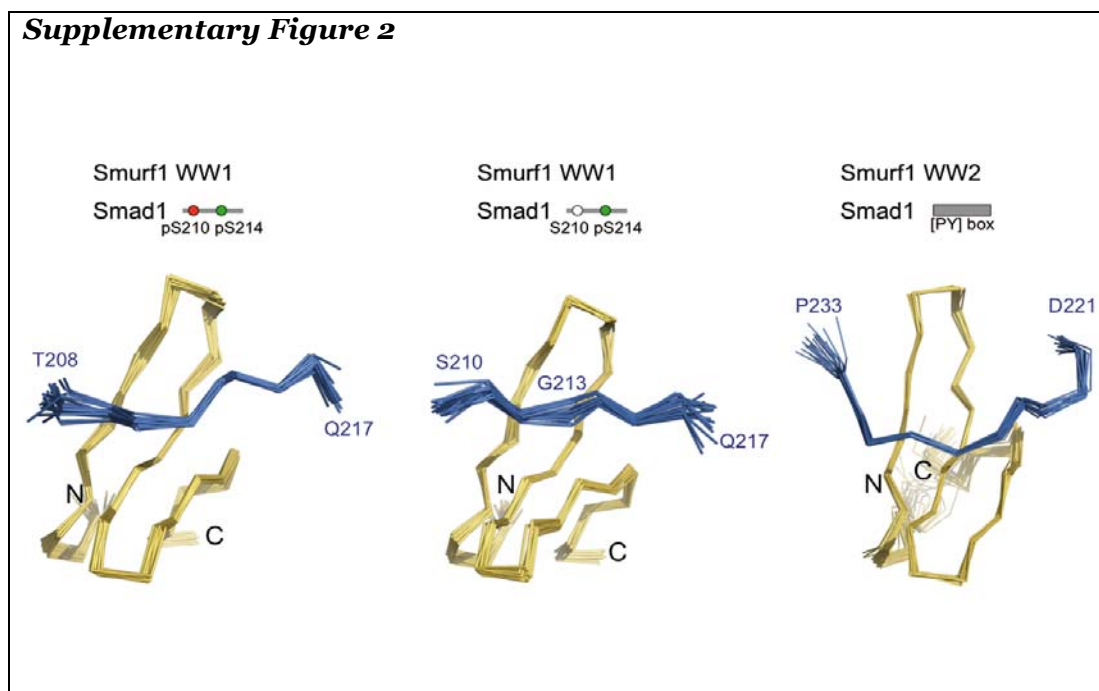
Supplementary Figure 1



15-Refined NMR structures of the Smurf1 WW1-WW2 pair (residues 232-314) bound to the 208-233 segment of the Smad1 linker di-phosphorylated at S210 and S214. The structure was fitted to either the WW1 domain (*left image*; 0.7 Å for backbone heavy atoms of the domain and the 208-216 region of Smad1) or to the WW2 domain (*right image*; 0.5 Å and the 222-232 region of Smad1). A semitransparent cartoon representation is used for the structure with lowest energy, with some side-chains highlighted. The structures were calculated using exclusively experimental dihedral and unambiguously assigned NOE restraints. Attempts to obtain RDC restraints by dissolving the complex in polyalcohol media only produced aggregated and precipitated samples. Secondary structure elements are based on NOEs and carbon chemical shift

differences with respect to random coil values. The backbone assignment of the protein in the bound state was obtained using a triple labeled sample.

Figure S2 Independently refined structures of Smurf1 WW1 and WW2 complexes with Smad1 peptides

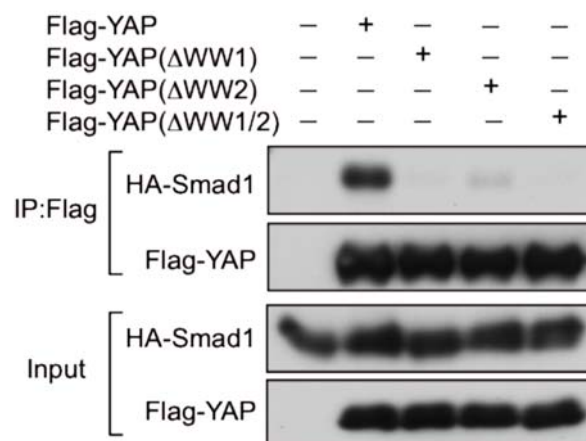


A backbone superposition (ribbon) of the WW1 domain bound to the di-phosphorylated (left, 0.25 Å R.M.S.D. for 20 best structures) and mono-phosphorylated segments (middle, 0.35 Å RMSD for the 20 best structures), or of the WW2 domain (right, 0.5 Å RMSD for the 20 best structures) bound to the extended PY site. The domains are oriented as in the representations of Figure 2. To obtain these refined structures we have used the data acquired for the complex containing the WW1-WW2 pair but focused the analysis in the characterization of either the WW1 or the WW2 interaction sites and not in the

determination of the relative orientation of one domain with respect to the other. In the two complexes investigated we did not detect contacts between the WW domains. All complexes have been obtained concentrating the protein in the presence of the peptide under study. Attempts to concentrate the protein in the absence of the ligand yield mixtures of aggregated with folded proteins.

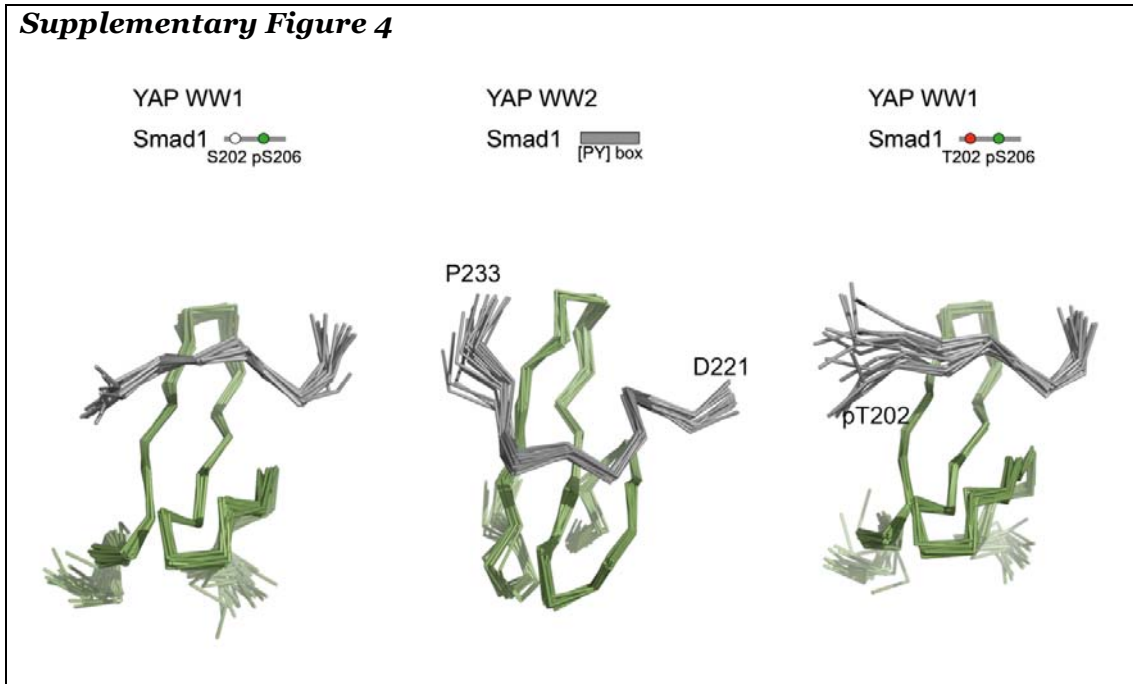
Figure S3. Both YAP WW domains contribute to bind Smad1

Supplementary Figure 3



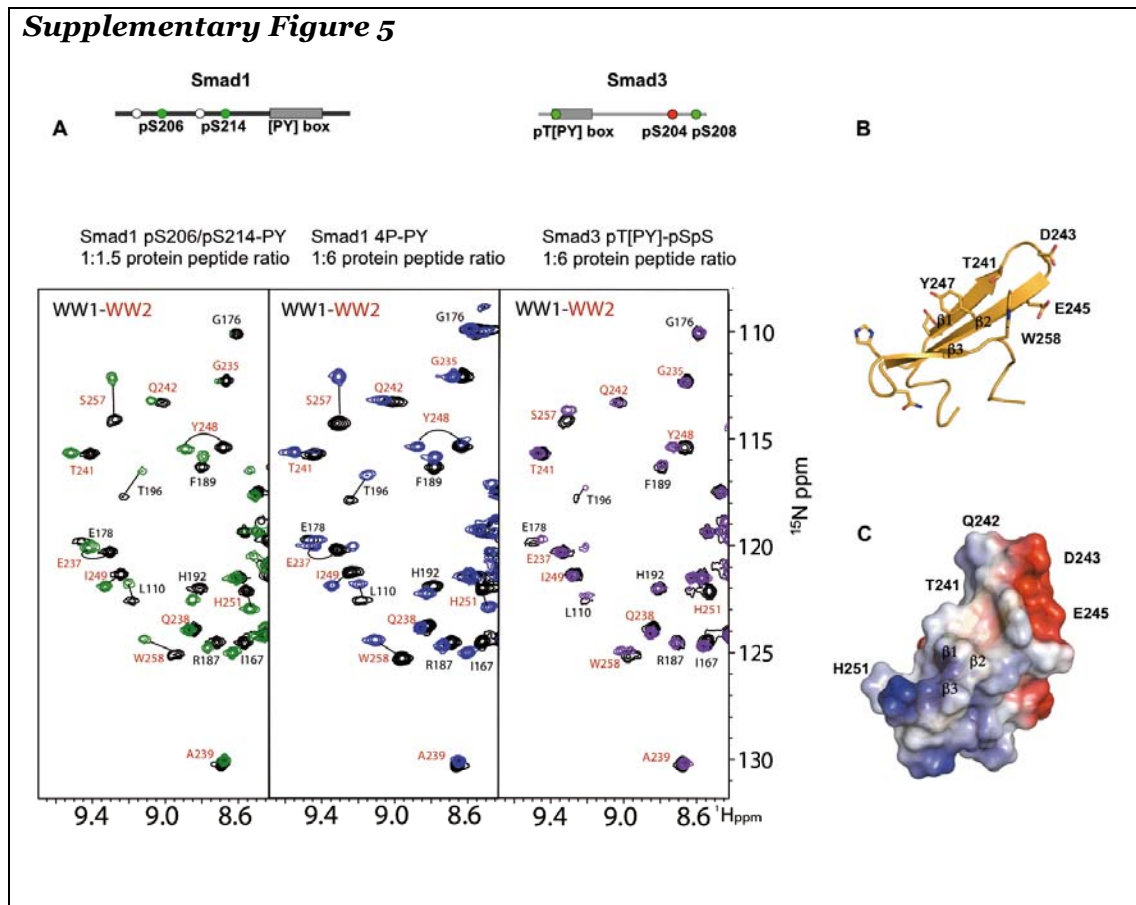
Effect of individual or combined WW domain deletions on the formation of a complex between Flag-tagged YAP and HA-tagged Smad1 in HEK293 cells.

Figure S4. YAP WW1 WW2 complexes with pS206pS214PY and pT202pS206pS214PY



A backbone superposition (ribbon) of the WW1 domain bound to the mono-phosphorylated (left, 0.46 Å R.M.S.D. for 20 best structures), or of the WW2 domain bound to the extended PY site (middle, 0.5 Å RMSD for the 20 best structures), and di-phosphorylated segments (right, 0.8 Å RMSD for the 20 best structures). The domains are oriented as in the representations of Figure 3. To obtain these refined structures we have used the data acquired for the complexes containing the WW1-WW2 pair and Smad1 (199-233) peptide but focused the analysis in the characterization of either the WW1 or the WW2 interaction sites and not in the determination of the relative orientation of one domain with respect to the other. In the two complexes investigated we did not detect contacts between the WW domains. Both complexes have been obtained concentrating the protein in the presence of the peptide under study to reduce the presence of aggregated proteins.

Figure S5. YAP titrations with Smad1 and Smad3 derived peptides



Note: Independently purified WW1 and WW2 domains interacted with a 12-residue Smad1 peptide including the PY motif and with a 17 residue pT[PY] Smad3 peptide, with K_D in the of 52-70 μ M range at 298 K. These values are similar to other values previously reported for YAP1 WW1 domain and protein ligands (Macias et al., 1996; Pires et al., 2001; Toepert et al., 2001). NMR binding experiments performed with peptides containing pS206/pS214-PY sites and the pair of WW domains induced chemical shift changes in both WW domains, as did titrations with tri-phosphorylated peptide (pT202/pS206/pS214) or with pS214 mono-phosphorylated peptide using a 1:1

peptide protein ratio. However, binding experiments using up to a 1:1 peptide: protein ratio with peptides phosphorylated at the GSK site pS210 (peptides pS210/pS214 and pT202/pS206/pS210/pS214) induced weak changes that suggested interactions only with the PY motif. To explore the interactions using peptide in excess, we have introduced a mutation in WW1 (L190Y), to avoid binding of two Smad1 molecules each to one WW domain. The HSQC experiment shown in (A) (Left 1:1.5) Smad1 pS206pS214-PY and Middle 1:6 protein:peptide ratio Smad1-4P-PY (pT202pS206pS210pS214-PY) show the expected changes in both WW domains resulting from the interaction of WW1 with the pS site and of WW2 with the PY site respectively. However titrations using Smad3 pT[PY]-pSpS peptide and a similar protein-peptide ratio as used with Smad1 4P-PY yield much smaller chemical shift changes (right). The presence of the negatively charged residues D243 and E245 in YAP WW2 (shown in the B and C panels) seems to reduce the affinity of this WW domain towards [PY] motifs preceded by a phosphorylated residue, thus explaining that YAP WW1-WW2 interacts with Smad1 with much a higher affinity than with Smad3 as observed previously (Alarcón et al., 2009).

4.3.6.1 ACCESSION NUMBERS

For each of the complexes (**short names** given below) we have deposited 20 structures in the Protein Data Bank and the list of restraints and chemical shifts in the BioMagResBank database. Structure statistics of the NMR refined complexes analyzed with Procheck are given in the Appendix section, Table 1.

- For the Smurf1-Smad1 complexes the corresponding PDB and BMRB codes are respectively: **WW1-pS214**: 2laz, 17541; **WW1-pS210pS214**: 2lbo, 17542; and **WW2-PY**: 2lb1, 17543.
- For the YAP-Smad1 complexes: **WW1-pS206**: 2lay, 17540; **WW1-pT202pS206**: 2lax, 17539 and **WW2-PY**: 2law, 17538.

5. Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF- β pathways

5.1 Summary

TGF- β and BMP signaling is mediated by Smads 1-5 (R-Smads and Co-Smads) and inhibited by Smad7, a major hub of regulation of TGF- β and BMP receptors by negative feedback and antagonistic signals. The transcription coactivator factor YAP and the E3 ubiquitin ligases Smurf1/2 and Nedd4L target R-Smads for activation or degradation, respectively. Pairs of WW domain in these regulators bind PY motifs and adjacent CDK/MAPK and GSK3 phosphorylation sites in R-Smads in a selective and regulated manner. In contrast, here we show that Smad7 binds YAP, Smurf1, Smurf2 and Nedd4L constitutively, the binding involving a PY motif in Smad7 and no phosphorylation. We also provide a structural basis for how regulators that use WW domain pairs for selective interactions with R-Smads, resort to one single versatile WW domain for binding Smad7 to centralize regulation in the TGF- β and BMP pathways.

5.2 Notes

1. This work has been published in *Structure 2012* with the title: "*Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF- β Pathways*" by Eric Aragón, Nina Goerner, Qiaoran Xi, Tiago Lopes, Sheng Gao, Joan Massagué and Maria J. Macias. A reprint is attached in the appendix section. The work was accepted with minor comments and earned the cover page of the Journal in October.

2. From the above mentioned paper, the work corresponding to YAP Smurf1 and Smurf2 interactions with a fragment of Smad7, which I have carried out with the supervision of Dra. M. J. Macias is included in the present chapter. The sections that correspond to Nedd4L complexes with Smad7 were included in the thesis defended by Nina Goerner.

3. Binding experiments using full-length proteins and mammalian cells were performed by Drs S. Gao, and Q. Xi, under the supervision of Dr. J. Massagué. They are included to provide the grounds for the structural work presented here.

5.3 Introduction

Compared to R-Smads and Co-Smads, the inhibitory Smads (I-Smads) have low sequence similarity in the MH1 domain but share an MH2 domain and the linker region including the PY motif (Fig. 1a). I-Smads are expressed in response to TGF- β or BMPs to provide negative feedback in the pathway (Hata et al., 1998), and in response to other pathways such as STAT to oppose TGF- β signaling (Ulloa et al., 1999). I-Smads negatively regulate the access of R-Smads to TGF- β receptors and also interfere with the formation of R-Smads-Smad4 complexes (Hayashi et al., 1997). Receptor-bound Smad7 recruits ubiquitin ligases Nedd4L, Smurf1 and Smurf2 to initiate receptor ubiquitination and down-regulation (Ebisawa et al., 2001; Kavsak et al., 2000; Kuratomi et al., 2005). Smad7 also binds YAP (Ferrigno et al., 2002) providing a possible mechanism for sequestration of this transcription factor that participates in Hippo and BMP signaling (Alarcón et al., 2009). These interactions involve the Smad7 linker and the WW domain region of the binding proteins.

To deepen our current understanding of the relationship between Smad7 and the R-Smads, we set to investigate the interactions between the WW domains of Smurf1/2, and YAP proteins and the Smad7 PY motif region using NMR and other complementary techniques such as Ion-Mobility Mass Spectrometry, calorimetry and cellular biology. The obtained data revealed that the interaction with Smad7 requires a single WW domain, the PY site of Smad7 and is independent from phosphorylation. This is in sharp contrast with the requirements of the recognition of R-Smads, which need the presence of WW domain pairs and both the PY site and the phosphorylated pSP motifs.

Overall, the results illustrate the versatility of different WW domain containing proteins as mediators of convergent interactions with a common

Smad7 target, in addition to their discriminating interactions with different R-Smad proteins.

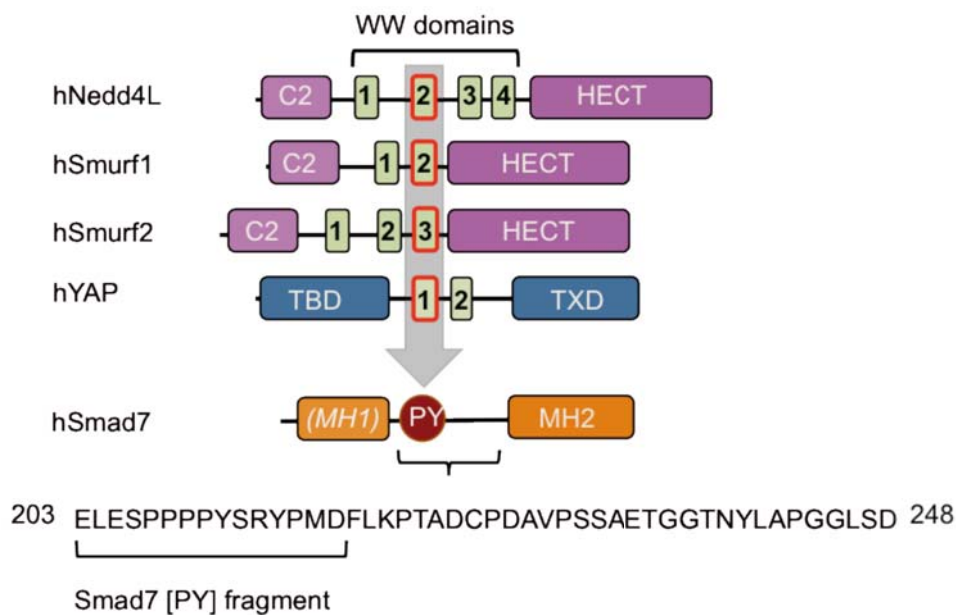
5.4 Results

5.4.1 Smurf1, Smurf2 and YAP use one single WW domain to bind Smad7

5.4.1.1 SMURF1 AND SMURF2 WW PAIRS ASSOCIATE VIA WW-WW INTERACTIONS

In order to characterize the protein regions involved in the interaction with the Smad7 linker we applied isothermal titration calorimetry (ITC) binding assays using recombinant proteins, either containing independent WW domains or all consecutive pairs and a 15-residue peptide, corresponding to amino acid residues E203-D217 of Smad7, and including the entire PY motif, (Figure 24).

Figure 24: Domain composition of three E3 ubiquitin ligases, Smad7 and YAP



The three-ubiquitin ligases contain the characteristic C2 domain, a central region with a variable number of WW domains and the catalytic HECT domain, each domain represented as a rectangle. The human Yes Associated Protein (YAP) contains a TEA

binding domain, two WW domains and a transactivator domain (TXD). Smad7 contains a canonical MH2 group and a divergent MH1. To highlight this divergence we have labeled the MH1 domain using italics and in brackets. The detailed sequence of the Smad7 linker is shown (residues 203-248). The synthesized PY peptide is underlined. The WW domains that mediate the interaction with the PY motif of Smad7 are indicated with an arrow.

The Smurf1 WW1 and Smurf2 WW2 domains show very low affinity for the Smad7 peptide whereas the Smurf1 WW2 and Smurf2 WW3 showed binding dissociation constants of 4.1 μM and a 1:1 stoichiometry with the Smad7 peptide. The affinity of the Smurf1 WW1-WW2 and Smurf2 WW2-WW3 pairs is $0.7 \pm 0.3 \mu\text{M}$ at 5 °C, $1.7 \pm 0.5 \mu\text{M}$ at 15 °C and $5.0 \pm 0.3 \mu\text{M}$ at 25 °C (Figures 24 and 25). These values are in agreement with previous reports for other WW interactions (Aragón et al., 2011; Chong et al., 2006, 2011; Gao et al., 2009; Kanelis et al., 2006; Pires et al., 2001; Ramirez-Espain et al., 2007).

Figure 25: ITC values obtained for the WW domains present in Smurf1, Smurf2 and YAP and the PY motif of Smad7

Protein	WW domain	K_D [μM]
hSmurf1	WW1	>100
	WW2	4.1 ± 0.1
	WW1/2	1.7 ± 0.5 N=0.6
hSmurf2	WW1	>100
	WW2	>100
	WW3	4.5 ± 0.2
	WW2/3	1.7 ± 0.4 N=0.6
hYAP	WW1	6.9 ± 0.3
	WW2	59.8 ± 3.4
	WW1/2	8.0 ± 0.3 N=1.1

ITC affinity values for the recombinant fragments of Smurf1, Smurf2 and YAP and the Smad7 synthesized peptide. Binding experiments have been performed at least three times, using two protein expression batches, and different buffers and temperatures. Values hereby presented were obtained at 15 °C.

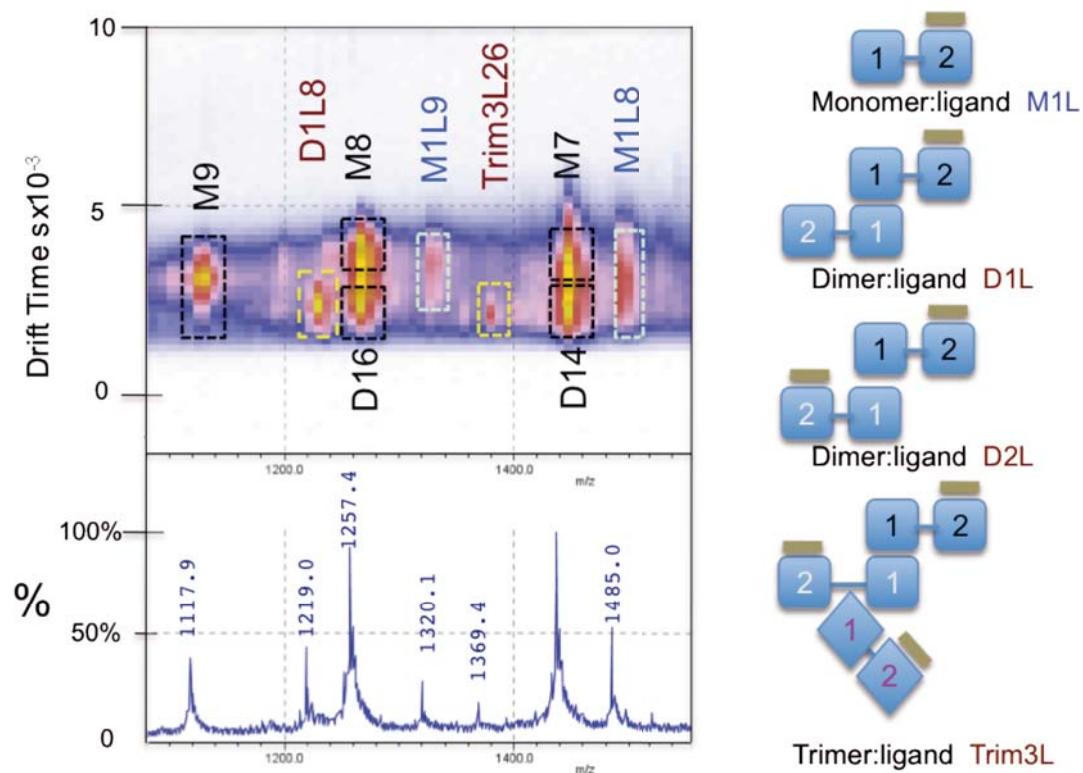
Thus, the affinity increase due to the presence of the WW domain pair is about 2-fold with respect to the values obtained with the Smurf1 WW2 or with the Smurf2 WW3 domains at a given temperature. However, with both protein pairs the affinity is calculated with a stoichiometry below 1 at all temperatures (Figure 25) (0.6-0.8 range). ITC experiments in two different buffer solutions (tris and ammonium acetate, pH 7) yielded similar values and stoichiometries. We considered the possibility that formation of protein aggregates via association of WW domains could affect the interpretation of the stoichiometry and affinity of Smad7 binding to the proteins containing WW-WW pairs. Indeed, using NMR we observed NOEs in Smurf1 samples containing either the WW1 independent domain or the WW1-WW2 pair that define a dimer *via* WW1-WW1 contacts. Since NMR and ITC experiments were carried out at different concentration ranges (milimolar *versus* micromolar), we made use of Ion Mobility-Mass Spectrometry (IM-MS) to investigate the potential presence of aggregates at the protein concentrations used in ITC experiments (30-50 μ M).

5.4.1.2 IM-MS AS AN AUXILIARY TECHNIQUE TO CHARACTERIZE PROTEIN AGGREGATES

Ion-mobility spectrometry-mass spectrometry (IMS-MS) is a method that is potentially able to rapidly separate ions according to their mobility, on a millisecond timescale using ion-mobility spectrometry and then uses mass spectrometry to identify components within a sample. Instead of responding to molecular fragments, however, IMS uses soft ionization. The ions are generated by atmospheric-pressure chemical ionization. Sample material is heated to yield a vapor that is swept into a small drift chamber where a beta radiation source ionizes the molecules. The resulting ions travel through a drift tube at distinct speeds that are related to their mass and geometry towards a detector. Using this technique to support the NMR and ITC data we have obtained, I could identify the presence of monomeric complexes, protein

dimers bound to only one ligand, dimers bound to two ligands and higher order species in the solutions corresponding to the complexes of each Smurf1 and Smurf2 WW-WW pairs and Smad7 peptide. The dimers, trimers and other higher order species were identified based on their different specific ionization masses and/or on their characteristic drift-times (Figure 26 and Supplementary Figure S6).

Figure 26: Ion Mobility-Mass Spectrometry data obtained for the complexes of Smurf2WW2-WW3 with the Smad7 peptide



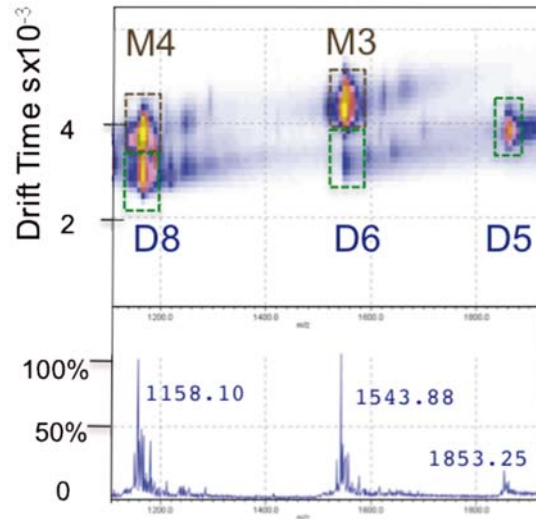
Left: An expansion of the Ion Mobility-Mass Spectrometry data obtained for the complexes of Smurf2WW2-WW3 with the Smad7 peptide displaying nine characterized species. The full spectrum and also that corresponding to the Smurf1WW1-WW2 in complex with Smad7 are shown as Supplementary Figure S6, S8, S9). The analysis of the different species identified for both complexes are collected in the Supplementary Figure S10. Each ion was assigned to a given species based on its characteristic mobility. Abbreviations used are ML (Monomer with one

ligand), D1L and D2L (Dimer with one or two ligands respectively), Trim1L, Trim2L and Trim3L (Trimers with one, two or three ligands). Numbers following the specie's name reflect the protonation state. We have unambiguously detected dimers with one or with two bound ligands and trimers in both Smurf1 and Smurf2 complexes.

Right: Schematic representation of the species identified by Ion Mobility-Mass Spectrometry (IM-MS) for the complexes of Smurf1 WW1-WW2 and Smurf2 WW2-WW3 pairs with the Smad7 PY site. The WW domains are represented as blue rectangles (labeled in black, white and violet to represent monomer, dimer and trimers respectively). The Smad7 peptide is represented as a green thick -line on top of the WW2 domain. Contacts involving two WW1 domains or between the WW1 domain of one molecule with the linker connecting the WW pair of a second molecule are based on experimental NOEs.

Dimers were also observed by IM-MS analysis performed with reference samples containing either the WW1 or the WW2 domains of Smurf1 in the unbound state (Figure 27). Interestingly, in the presence of the Smad7 peptide, the dimer population was reduced in the sample containing the WW2 domain, while that of the WW1 domain was unaffected (Figure S7).

Figure 27 Ion Mobility-Mass Spectrometry data obtained for the complexes of Smurf2^{WW2}-WW3 with the Smad7 PY site



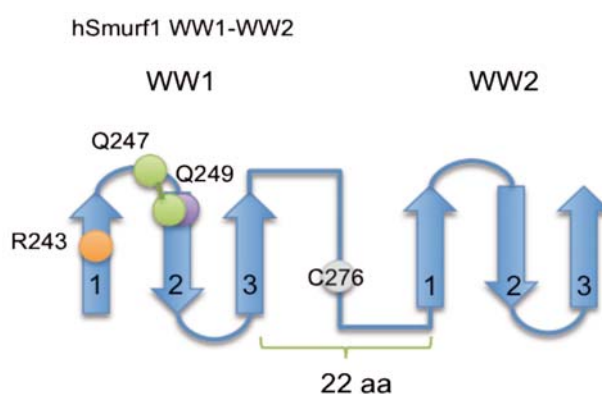
A region of the Ion Mobility-Mass Spectrometry data obtained for the Smurf1 WW1 dimer, (the full spectrum is shown as Supplementary Figure 7). As in Figure 26, each ion was assigned to a given species based on its characteristic drift-time.

Thus, using this technique that requires samples in the low micromolar concentration range, we could explain the stoichiometries observed in the calorimetry experiments. Under the experimental conditions that I used to characterize the interactions (at very diluted protein concentrations) the <1 stoichiometry may result from the co-existence of protein monomers, dimers, trimers and tetramers each binding one equivalent of Smad7 peptide, plus the presence of other species bound to two, three or four Smad7 equivalents.

Since the presence of the Smad7 ligand does not prevent the formation of the higher order species, we interpret that Smurf1 WW1 and Smurf2 WW2 domains have a minor role in binding to short PY containing sequences, but an important role in protein oligomerization and aggregation, whereas the WW2

and WW3 domains were responsible for the ligand recognition. This observation differs from a previously reported interpretation of Smurf2 WW2-WW3 bound to Smad7 (Chong et al., 2011), where the WW2 domain also participates in hydrophobic and electrostatic interactions with the Smad7 peptide.

Figure 28: Schematic representation of the pair of WW domains and the linker present in Smurf1







Representation of the WW pair present in Smurf1. The mutated positions used in the ITC binding experiments and the three strands of each WW domain are labeled.

To characterize further the role of Smurf1 WW1 domain we introduced mutations in the WW1-WW2 pair domain in equivalent positions to those that in Smurf2 were proposed to contact Smad7, and also two additional control mutations in a position that did not participate in the interaction of Smurf2 and Smad7 but in the dimer formation (Figure 28, 29). In all cases, single and double mutations (Arg243Ala, Gln249, Gln247 and Gln249 to Glu) reduced the affinity by 3-fold while the Arg243Glu mutation reduced the affinity by 10-fold. The results suggest that these residues do not play a key role in binding but that they may participate in protein homodimerization, perhaps enhanced by the use of recombinant protein fragments. Based on this interpretation of the

binding data, we conclude that Smurf1 and Smurf2 use their WW2 domain and WW3 domain, respectively, as their primary binding sites for the Smad7 PY peptide.

Figure 29: Affinity values obtained by ITC titrating Smurf1 WW pairs with several point and double mutations

Protein	WW1/2 domain	K_D [μ M]
hSmurf1 C276M	WT	1.7 ± 0.5 N=0.6
	 Q247/Q249 E	7.3 ± 1.5
	 Q249E	6.2 ± 1.2
	 R243A	6.3 ± 0.7
	 R243E	15.7 ± 3.3

List of the measured affinity values obtained for the mutations included in the Smurf1 WW1-WW2 pair.

Finally, ITC titrations performed at 15°C with the YAP WW domains revealed that the YAP WW1 domain preferentially binds to the Smad7 peptide, with a dissociation constant of $6.90 \pm 0.28 \mu$ M, while WW2 binds with a dissociation constant 9-fold weaker and the WW1-WW2 pair binds slightly worse than the isolated WW1 ($8.0 \pm 0.3 \mu$ M and N=1.1). Thus, the interaction with the Smad7 peptide mainly involves the YAP WW1 domain (Figure 25).

All together, these results suggest that in each case a specific, single WW domain is sufficient for high-affinity recognition of the Smad7 PY site (Figure 24).

5.4.2 Structure of the Smad7 PY motif bound to different E3 ubiquitin ligase WW domains

5.4.2.1 CHARACTERIZATION OF THE SMAD7 PY INTERACTION WITH THE WW-WW PAIRS BY NMR

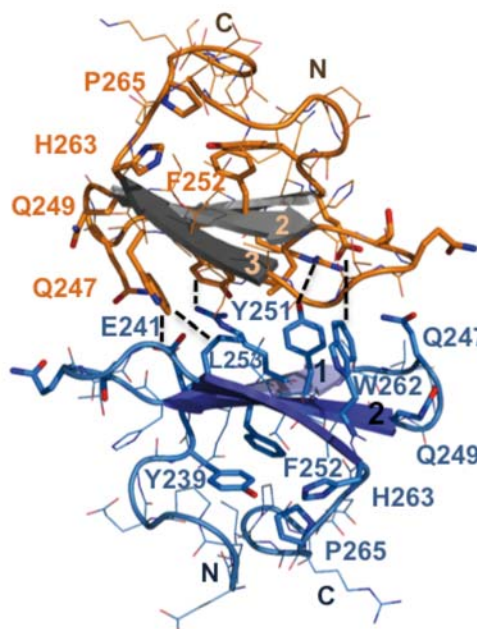
To compare the binding modes of the E3 ubiquitin ligases Smurf1 and Smurf2 and the different Smad proteins, we investigated the interactions between the Smurf1/2 WW-WW pairs and also the independent WW2/3 domain with the Smad7 peptide using triple resonance NMR spectroscopy.

Under these conditions the NMR assignment of the complexes reveals that, both Smurf1 and Smurf2 proteins interact with the Smad7 peptide using the WW2 domain, in the case of Smurf1, and the WW3 for Smurf2. As observed during the ITC titrations and IM-MS experiments, the Smurf1 WW1-WW2 pair displays a high tendency to form dimers and other higher order aggregates.

Using 2D- and 3D-NOESY experiments we characterized a population of dimers formed *via* interactions between the WW1 domains of two molecules according to intermolecular NOEs that fit as a beta-clam. To illustrate these interactions we generated a model using the structure of the Smurf1 WW1 domain (pdb entry: 2laz) and the unambiguously assigned NOEs detected between monomers (Figure 30 and 31). We have also detected a minor population of dimers formed by interactions between the WW2 domain of one molecule and the pair of prolines present in the linker connecting the WW1-WW2 pair of a second molecule, which can explain the trimeric and tetrameric species identified by IM-MS (Figure 26). We observed as well the dimerizing tendency with the WW2 domain of Smurf2 in the WW2-WW3 pair.

Figure 30: Unambiguously assigned NOEs between monomers **Figure 31: Model of the WW1-WW1 homodimer**

Smurf1WW1/2 and Smad7	
Intermolecular NOEs	
Monomer 1	Monomer2
Trp262	Glu241
Trp262	Leu253
Tyr251	Arg243



Left: Figure 30 shows a list with the unambiguously assigned NOEs that defined the dimer interface represented with dotted lines in Figure 31. Monomers are colored in orange-gray (top) and cobalt blue (bottom), with key residues highlighted.

Right: Figure 31 shows a detailed view of the binding interface of the dimer between two Smurf1 WW1 domains (residues 233-270).

The beta-clam arrangement in the dimer is similar to that described for the WW2 domain of the mouse Salvador homolog 1 protein (Ohnishi et al., 2007). With the WW domain pairs we detected a broadening of the intermolecular NOEs that defined the peptide in the bound conformation and two sets of NOEs for the Y211 with residues in the WW2/WW3 domains. We interpreted the broadening and the presence of the second set of signals for the Y211 aromatic ring as the result of the peptide bound in several complexes, for instance, the main conformations that correspond to the monomer in complex with one ligand and the symmetric dimer with two bound ligands in equilibrium with an asymmetric dimer bound to a single ligand (Figures 26). At 298K and in the presence of 2% DMSO both sets of NOEs corresponding to the

Y211 collapse to one set that we interpret it as the bound monomer. Under these experimental conditions, we did not observe however contacts between the domains in the Smurf1 WW1-WW2 pair, or between the WW1 and the Smad7 peptide as described for the Smurf2 WW2-WW3 complex with Smad7 (Chong et al., 2011), or as we previously observed in the complex of Smurf1 WW1-WW2 pair with Smad1 (Aragón et al., 2011).

Based on these observations we focused the structural work and the characterization of the complexes using the WW2 domain of Smurf1 and in the WW3 of Smurf2.

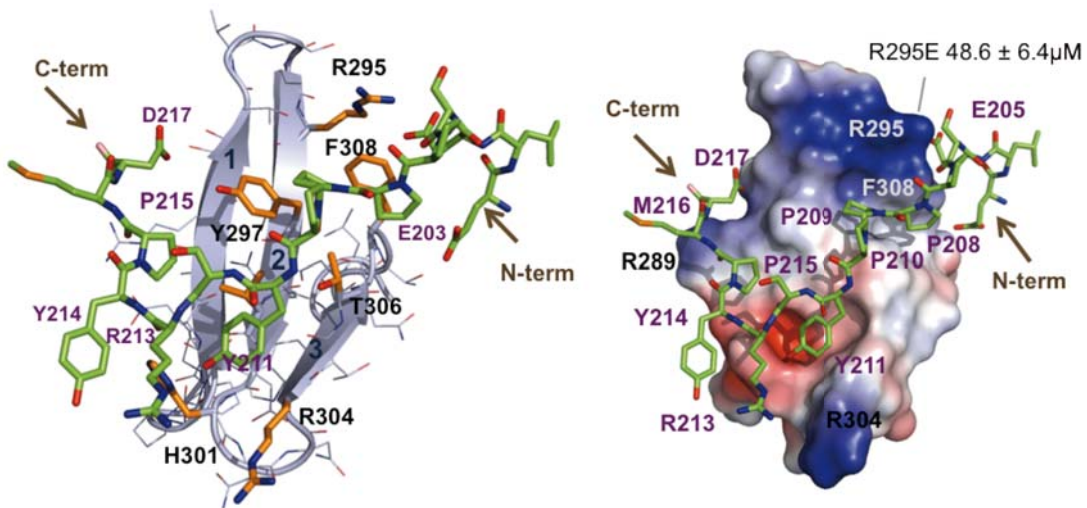
5.4.2.2 STRUCTURES OF SMURF1 WW2 AND OF SMURF2 WW3 DOMAINS IN COMPLEX WITH SMAD7 PY PEPTIDE

As both Smurf complexes are very similar, I will describe them in parallel, with the corresponding residues separated by a slash. The Smurf1WW2- and Smurf2WW3-Smad7 complexes are well defined, based on numerous contacts detected from E205-P215 Smad7 residues with the domains (Figures 32 and 33, Supplementary Figures S11 and S12 and Table 2, shown in the Appendix).

In both complexes the Smad7 fragment also forms a turn, centered at positions Y211-S212-R213, especially in the Smurf1 complex. A comparison of these two complexes with that of Nedd4L revealed some additional differences, for instance E205 is interacting with Arg295/Arg312 in the second strand, but no contacts with the peptide are observed for the residues located in loop 1 of the WW domains, which in these cases are Ser293/Thr310 and not a Lys as in the Nedd4L WW2 domain. As a consequence, E205 and D217 are less defined in the Smurf1 WW2 and Smurf2 WW3 complexes. A mutation introduced in the Smurf1 WW2 domain (Arg295Glu) reduces the affinity from $4.1 \pm 0.1 \mu\text{M}$ to

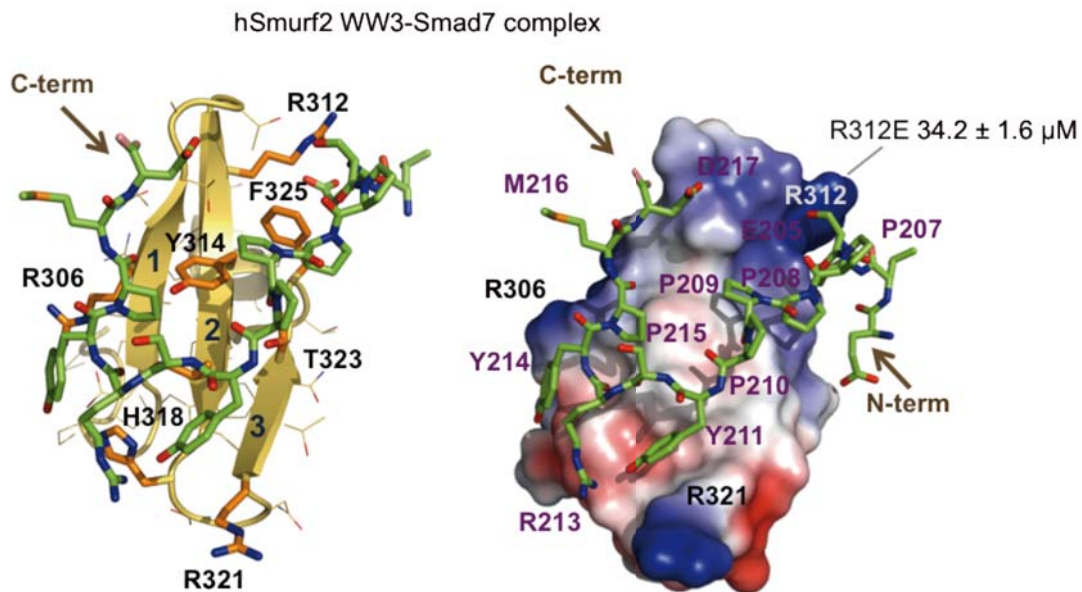
$48.6 \pm 6.4 \mu\text{M}$, suggesting an active implication of Arg295 in the peptide interaction. The equivalent mutation introduced in Smurf2WW3 also reduces the affinity from $4.5 \pm 0.2 \mu\text{M}$ to $34.2 \pm 1.6 \mu\text{M}$.

Figure 32: Refined structure of the Smurf1 WW2 bound to the Smad7 peptide



Refined structure of the Smurf1 WW2 domain (residues 233-270, light-blue) bound to the Smad7 peptide (203-217, green). Next to it is the charge distribution on the surface of the Smurf1 WW2 domain in complex with the peptide. The structure with the lowest energy was selected for both representations. The family of 25-calculated structures is shown in Supplementary Figure S11. N- and C-termini of Smad7 peptide are represented with brown chocolate arrows. Positively charged sites are colored in blue and negatively charged sites in red. Key residues that participate in the complex interaction are labeled in purple (Smad7) and in black (Smurf1). The gray line indicates the position of the R295E mutation, which decreases the binding affinity of the complex by approximately 8-fold.

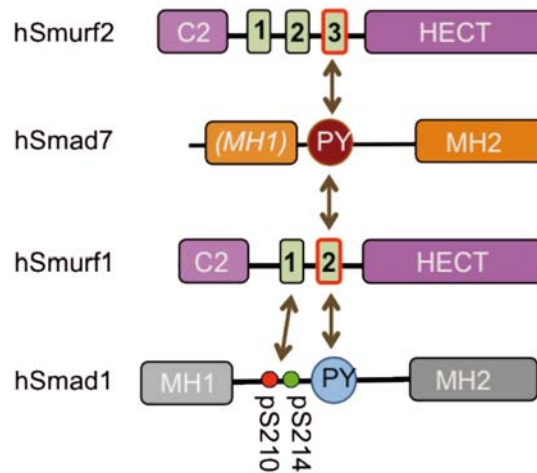
Figure 33 Refined structure of the Smurf2 WW3 bound to the Smad7 peptide



Smurf2 WW3 domain (297-313, gold) bound to the Smad7 peptide (green) represented with a similar orientation as that displayed for Smurf1 WW2, Figure 32 for the. The R312E mutation, which decreases the binding affinity of the complex by approximately 10-fold, is indicated with a gray line.

The complexes described in this work are similar to the previously characterized Smurf2WW3 and Smad7 (Chong et al., 2006), with the main differences involving the contacts with the N-terminal site of Smad7 (E205), the position of P215, and the absence of intra peptide contacts from residues M216-D217. On the other hand, they differ from the complex between the Smurf1WW1-WW2 pair and Smad1, where both WW domains have a direct role in ligand recognition (Figure 34) (Alarcón et al., 2009; Aragón et al., 2011) and from the complex between Smurf2WW2-WW3 and Smad7, where the contacts with the C-term part of Smad7 that we observe to occur with the WW3 domain were detected with the first loop of the WW2 domain (Chong et al., 2011).

Figure 34: Smurf1 WW2 and Smurf2 WW3 are the main players for the recognition of Smad7 PY site

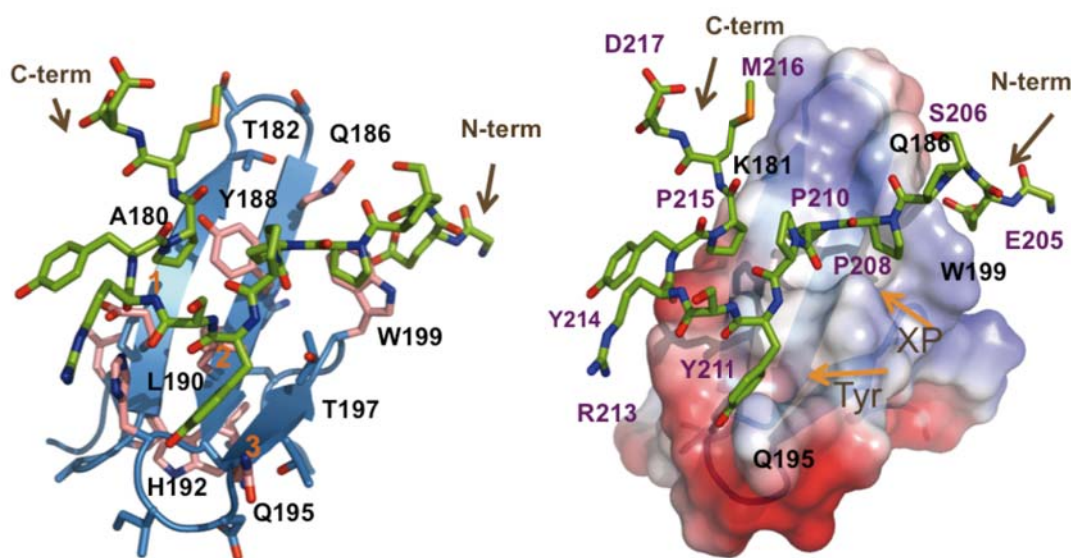


5.4.3 Structure of the complexes between YAP and Smad7 PY motif: Smad7 selects YAP WW1 and not WW2

The Smad7 linker fragment binds to the YAP WW1-WW2 pair and to its independent WW1 domain with similar affinities suggesting that the WW1 is the preferred binding site. To provide a structure-based interpretation for the different affinities observed between the YAP WW1 and WW2 domains, the structure of each domain in complex with the Smad7 fragment was determined and the similarities and differences analyzed in detail. As both YAP WW1 and WW2 sequences are very similar, I also describe them in parallel, with the corresponding residues separated by a slash. In each complex the WW1 and WW2 domains adopt the canonical WW fold and bind to the Smad7 PY core in

a similar manner (Figure 35 and Supplementary Figure S13 and Table 2, Appendix).

Figure 35: Detailed view of the interaction of the YAP WW1 domain with the Smad7 [PY] peptide



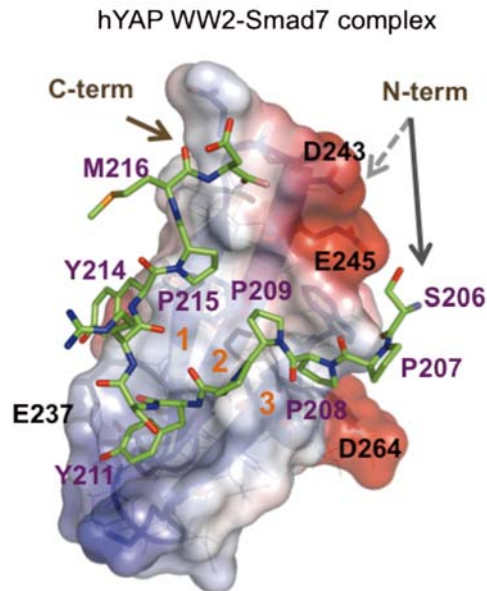
Left: Detailed view of the interaction of the YAP WW1 domain (residues 163-206, blue) with the Smad7 [PY] peptide (green) and next to it, the charge distribution of the domain in the complex, shown as a semitransparent surface representation. The structure with the lowest energy was selected for both representations. The family of 25-calculated structures is shown as Supplementary Figure S13. Critical residues involved in the interaction are labeled in purple (Smad7) and in black (YAP WW1). Both N- and C-termini of Smad7 peptide) are represented with arrows.

The Y211 ring is accommodated in its respective tyrosine binding cavities formed by Leu190/Ile249, His192/His251, Gln195/Lys254 and Thr197/Thr256 while the pyrrolidine rings of Smad7 P208 and P209 are respectively accommodated in the XP cavities formed by Tyr188/Tyr247 and Trp199/Trp258 (Figures 36 and Supplementary Figure S14 and Table 2, Appendix). The main differences between both complexes are the new contacts observed between Smad7 Y214 and Tyr247, Ile249 and Glu237 in the WW2 complex and more significantly, the absence of interactions between the residues located in loop 1 and the residues preceding or following the PY site,

which are observed in the complex with WW1, and in the complexes corresponding to the three E3 ubiquitin ligases. The absence of these interactions could be interpreted on the basis of the charge distribution of the YAP WW2 domain, which differs from that of the WW1 domain.

The presence of negatively charged residues in loop 1 (Asp243 and Glu245) repels both the accommodation of the N-term part of Smad7 as well as the negatively charged residues located in the C-term extension of the PY motif. Point mutations introduced in Asp243 and Glu245 residues to Glutamine resulted in an improved affinity by a factor of two (Figures 36 and 37).

Figure 36: Semitransparent surface representation showing the charge distribution of the YAP WW2 domain bound to the Smad7 peptide



A semitransparent surface representation showing the charge distribution of the YAP WW2 domain (residues 227-266, deep blue) bound to the Smad7 peptide represented as sticks (green). The structure with the lowest energy was selected for both representations with some side-chains and elements of secondary structure highlighted. The family of 25-calculated structures is shown as Supplementary Figure S14. The peptide's N-terminus (Ser206) is shifted from the loop1 with respect to the orientation in the complex with WW1. The different positions of the peptide in both

complexes are shown with two arrows, with a straight gray line (WW2 complex) and with a dotted gray line (the position in the WW1 complex).

Figure 37: Affinity changes in Smad7 binding resulting from point mutations in the YAP WW1 and WW2 domains

WW domain	K_D [μ M]
WW1, Q186E	14.0 ± 0.2
WW2, D243S E245Q	27.0 ± 0.8

The Q186E change in the WW1 domain reduces the affinity by two-fold, while a double change introduced in the WW2 linker -to mimic the sequence of the first WW domain- improves the affinity by a factor of two. The positions in the structure of the mutated residues are indicated in Figures A and B.

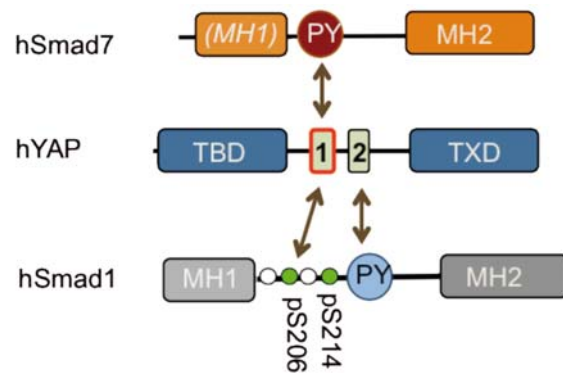
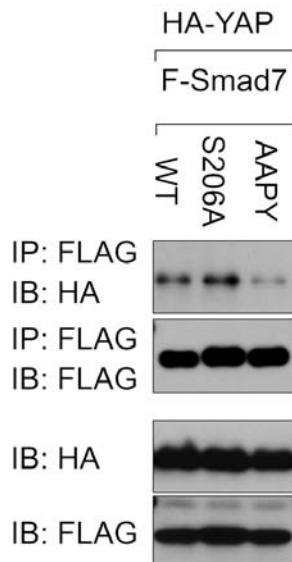
5.4.3.1 THE YAP WW1-SMAD7 STRUCTURE DOES NOT SUPPORT A ROLE (OR ROOM) FOR A PHOSPHATE GROUP ON S206

The YAP WW1-Smad7 structure does not support a role (or room) for a phosphate group on S206 (S206 is tightly bound by Trp199), suggesting that Ser206 phosphorylation would not enhance the interaction with YAP *in vivo*. To explore this possibility our collaborators made use of immunoprecipitation experiments (IP) performed using plasmids encoding flag epitope-tagged with either full-length protein Smad7 (wild type) or with the S206A and AAPY variants transfected into the HEK293T cell line and the HA-YAP construct. These experiments showed that binding is observed with both the wild type and the S206A variant, while the AAPY variant cannot precipitate HA-YAP (Figure 38) suggesting that the Ser206 or its potential modification would play a minor role in the function.

In summary, the first WW domain of YAP is the preferred binding site for the Smad7 PY motif and the discrimination between both domains is defined in terms of the more abundant protein-peptide contacts observed with the WW1

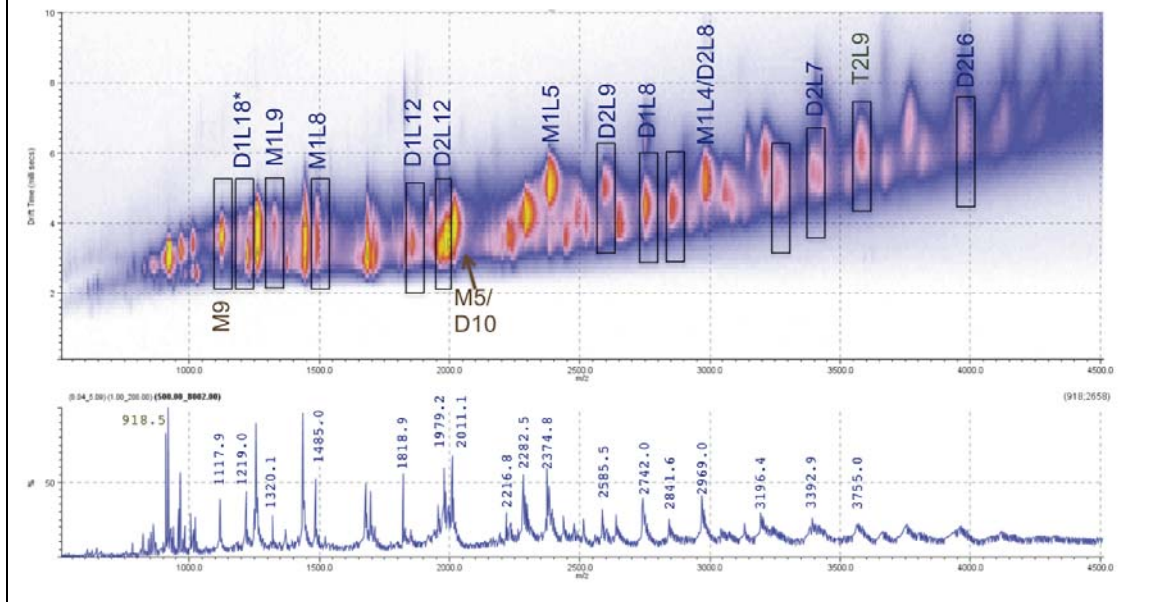
than with the WW2 domain. Again, the interaction of YAP with Smad7 is different from that with Smad1, where both WW domains participate in the complex and where WW2 is responsible for the interaction with the Smad1 PY site while the WW1 recognizes the phosphorylated pSerPro motif (Figure 39).

Figure 38: IP experiments show that phosphorylation has a minor effect in the interaction **Figure 39:** Schematic representation of YAP binding to Smad1 and Smad7 sites

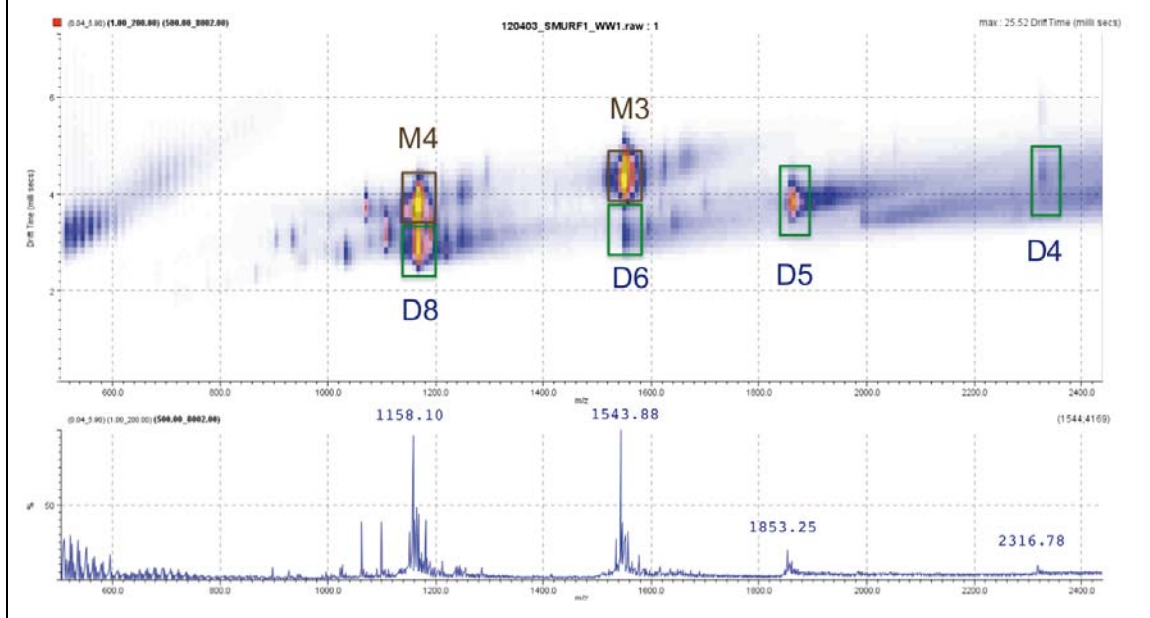


HEK293T cells expressing the indicated constructs were analyzed.

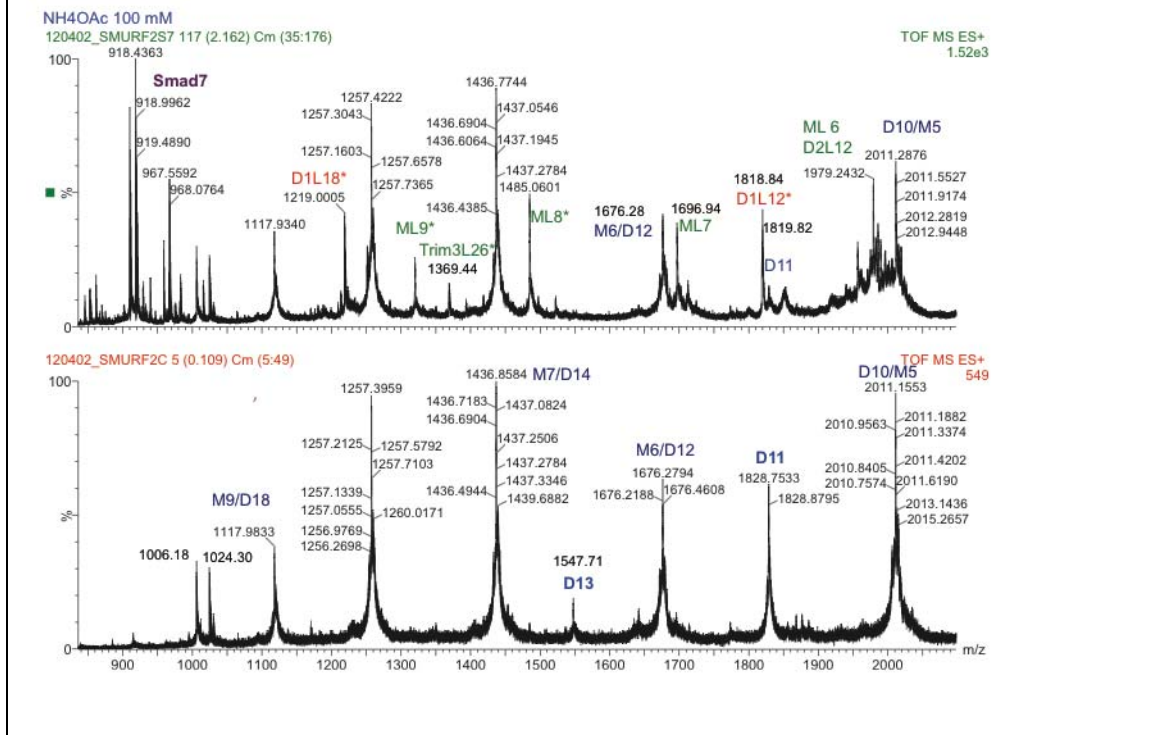
Schematic representation of the binding modes of YAP with Smad7 (top) and with Smad1 (bottom). The interaction with the PY site of Smad7 requires only the first WW domain, while binding to the pS and PY sites of Smad1 requires both WW domains.

Supplementary Figure 6

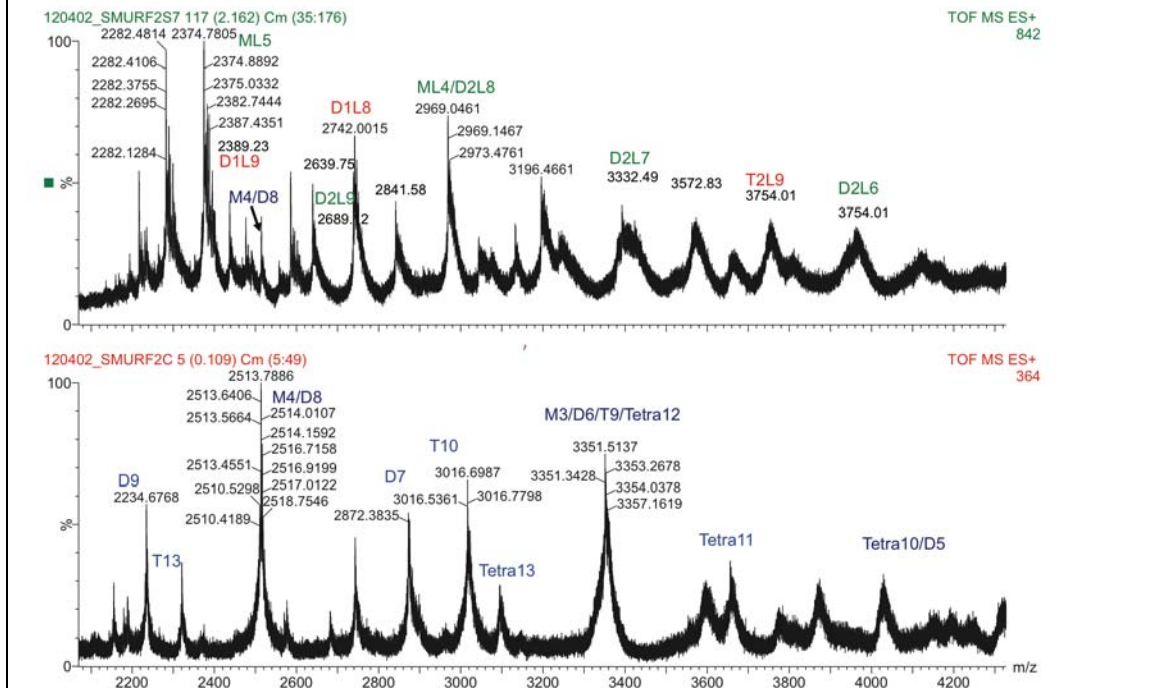
Ion Mobility-Mass Spectrum obtained for a 40 micromolar sample of Smurf2 WW2-WW3 domain in complex with Smad7 peptide dissolved in 50 mM ammonium acetate at pH 7.2. A region of this spectrum is shown in Figure 26.

Supplementary Figure 7

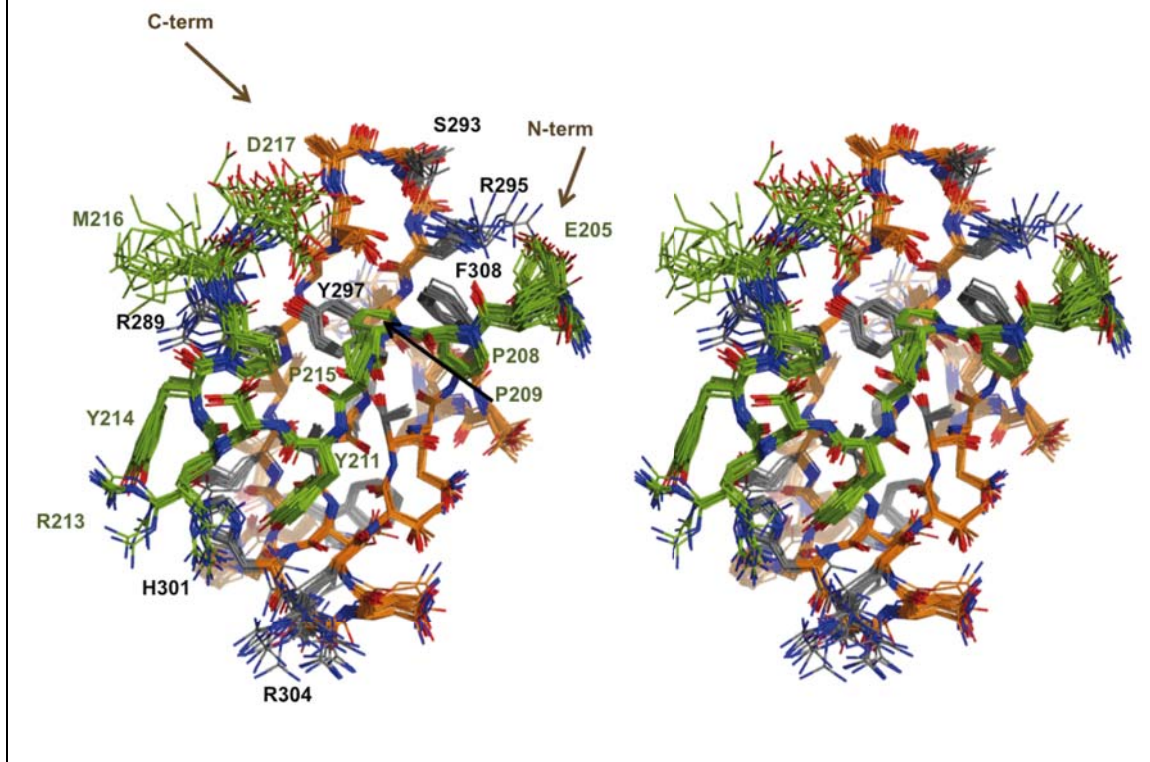
Ion Mobility-Mass Spectrum obtained for a 30 micromolar sample of Smurf1 WW1 domain dissolved in 50 mM ammonium acetate at pH 7.2. A region of this spectrum is shown in Figure 27.

Supplementary Figure 8

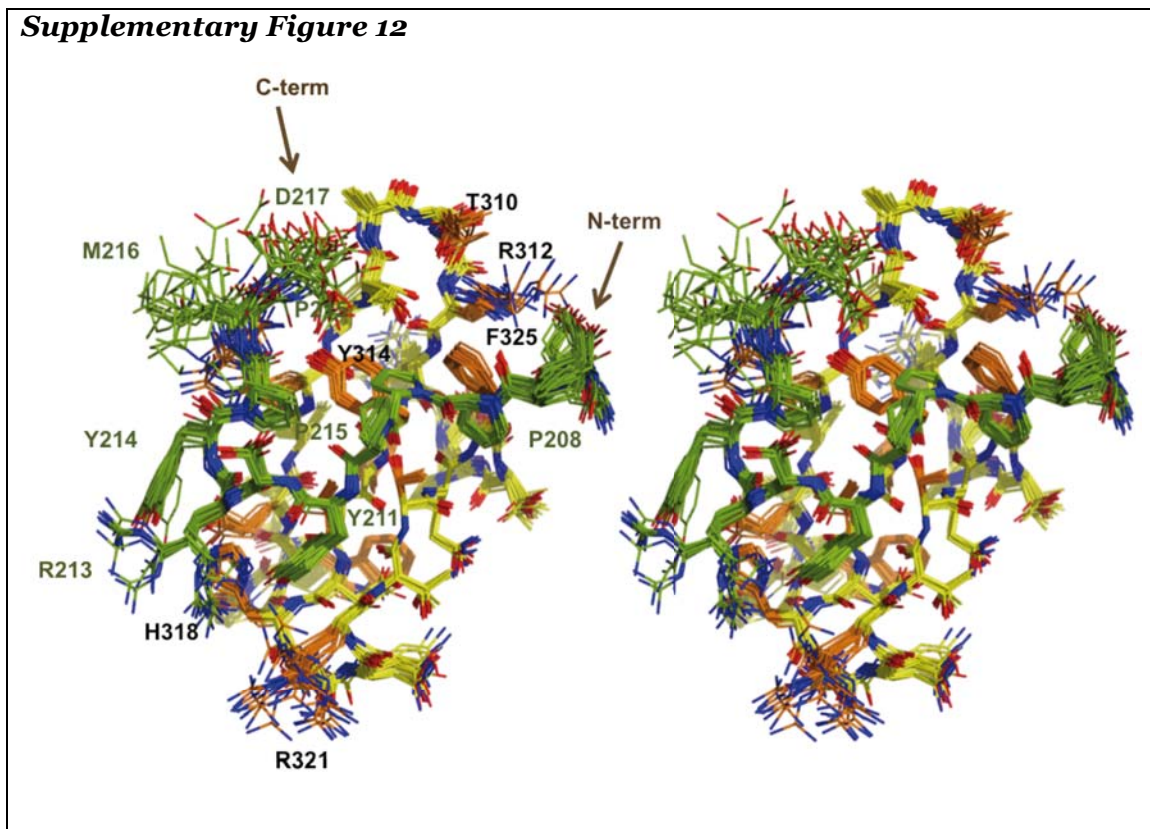
Different species detected by IM-MS for Smurf2 WW2-WW3 in complex with Smad7 peptide (Top panel) and free Smurf2 WW2-WW3 free (down panel). The data continues in Supplementary Figure 9. A list containing the species detected for both Smurf1 WW1-WW2 and Smurf2 WW2-WW3 complexes is shown as a table. Unbound peptide is labeled as Smad7. The different species are labeled based on their corresponding mass; ions corresponding to unbound protein are labeled in blue. Symmetric complexes such as dimers with two bound peptides, trimers with three peptides and so on are labeled in green. Protein dimers or trimers bound to a single peptide are labeled in red. Unique species are marked with a star.

Supplementary Figure 9**Supplementary Figure 10**

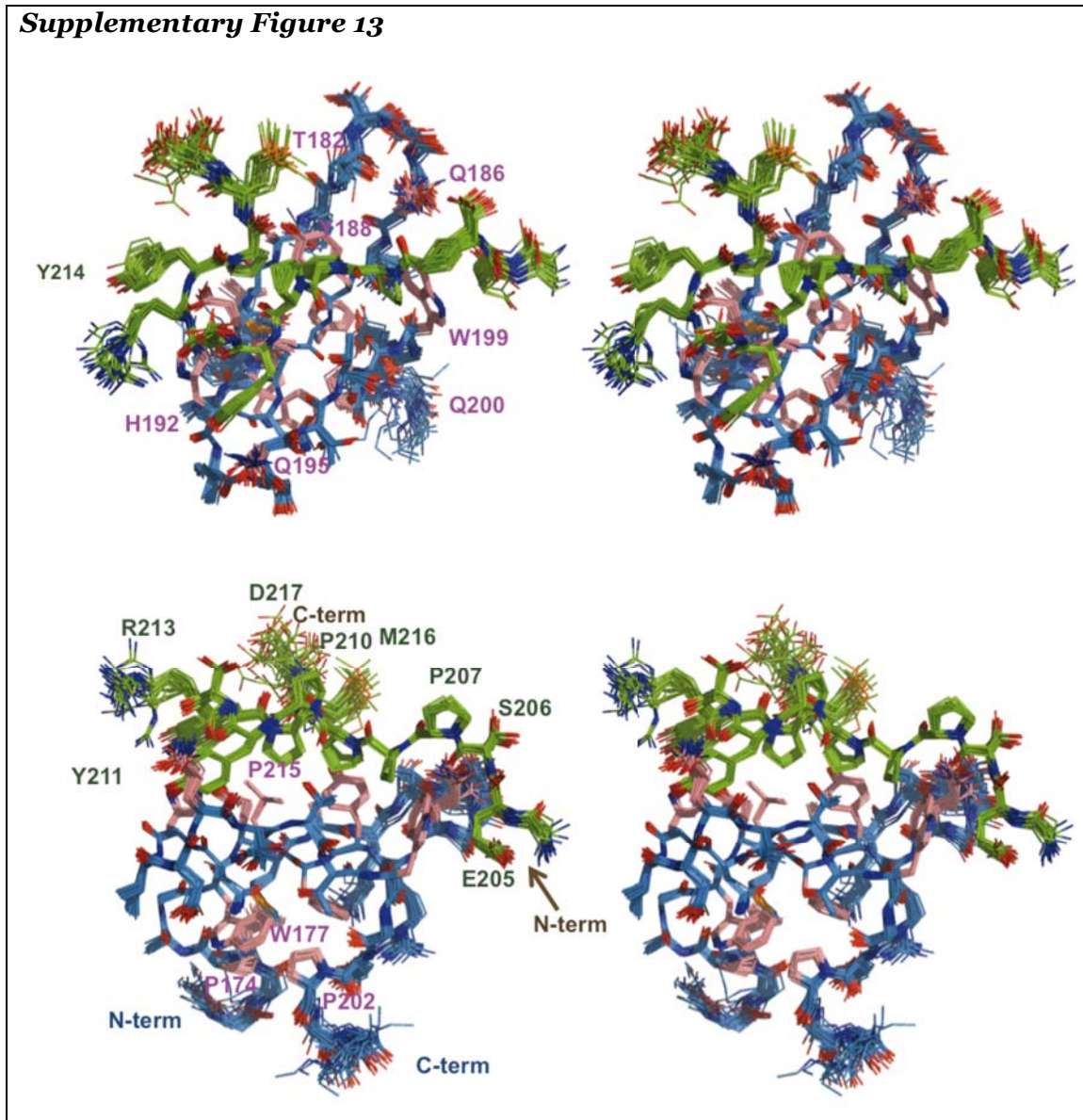
Smurf1WW1/2 and Smad7		Smurf2WW2/3 and Smad7	
Mass	Specie	Mass	Specie
1560.70	D2L15	1219.00	D1L8
1955	Trim2L17	1369.44	Trim3L26
1958.26	D1L11	1485.04	ML8*
2336.91	ML5/ D2L10	2585.85	D2L9
2344.95	Tetra8L23	2841.58	Trim1L11
2694.00	D1L8	2389.23	D1L9
2696.00	Trim3L8	2742.00	D1L8
3079.9	D1L7	3332	D2L7
3338.86	D2L7	3754.01	T2L9
3903.67	Trim1L8		

Supplementary Figure 11

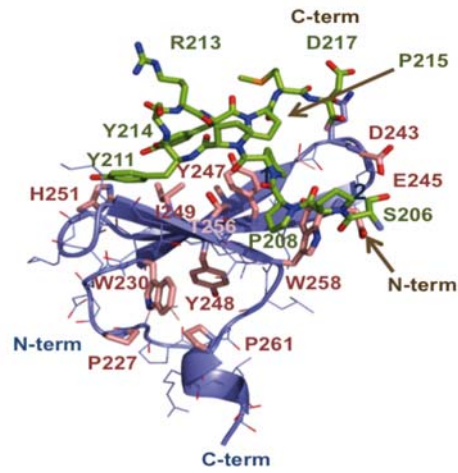
25-Refined NMR structures of the Smurf1 WW2 domain (residues 277-314, orange) bound to the 203-217 segment of Smad7 (green), oriented as in Figure 32. The structures were fitted using PyMol (0.4 Å for backbone heavy atoms of the domain and the Smad7 peptide). A few selected residues are labeled in green (peptide) and in gray (WW domain).

Supplementary Figure 12

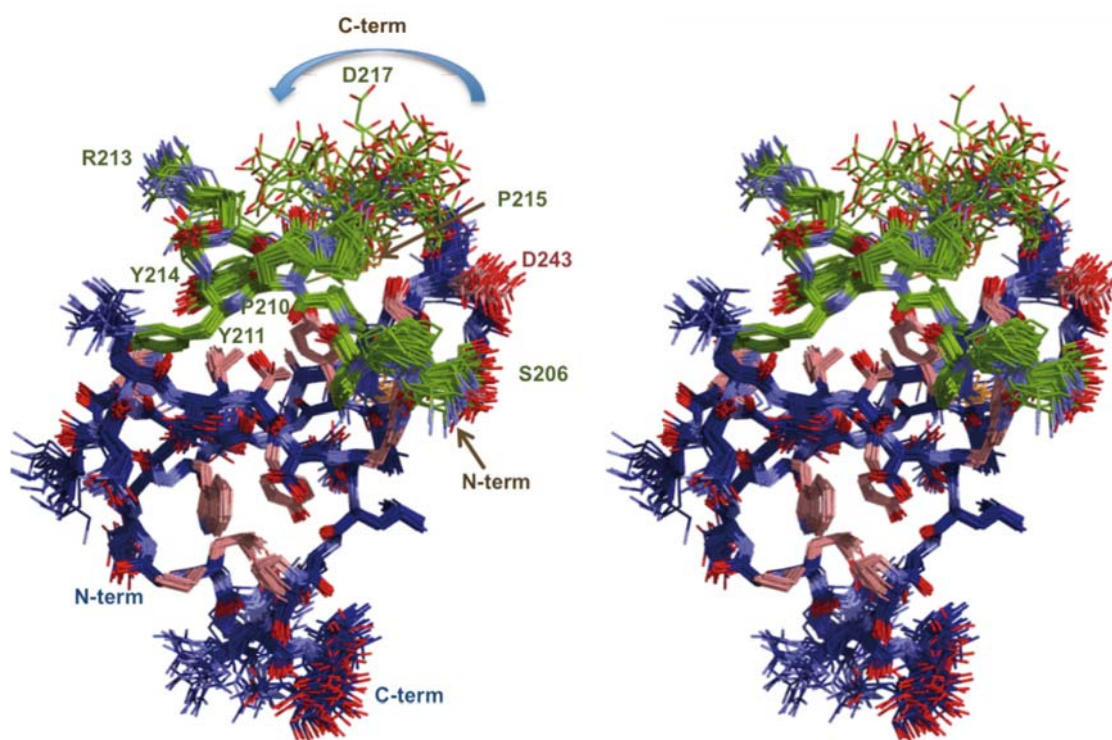
25-Refined NMR structures of the Smurf2 WW3 domain (residues 297-333, yellow) bound to the 203-217 segment of Smad7 (green) oriented as in Figure 35. The structures were fitted using PyMol (0.3 Å for backbone heavy atoms of the domain excluding the flexible ends and the 203-215 fragment of the Smad7 peptide). A few selected residues are labeled in green (peptide) and in black (WW domain). Residues in the WW domain are highlighted in orange.

Supplementary Figure 13

25-Refined NMR structures of the YAP WW1 domain (residues 163-206, orange) bound to the 203-217 segment of Smad7 (green) oriented as in Figure 36 and with a 90 rotation. The structures were fitted using PyMol (0.3 Å for backbone heavy atoms of the domain and the Smad7 peptide). A few selected residues are labeled in green (peptide) and in violet (WW domain).

Supplementary Figure 14

Cartoon representation of the complex between YAP WW2 (cobalt blue) and Smad7 (green). A few selected residues are labeled in green (peptide) and in orange (WW domain).

Supplementary Figure 15

25-Refined NMR structures of the YAP WW2 domain (residues 232-314, cobalt blue) bound to the 203-217 segment of Smad7 (green) oriented as in Figure 37. The structures were fitted using PyMol (0.3 Å for backbone heavy atoms of the domain excluding the flexible ends and the 203-215 fragment of the Smad7 peptide). A few selected residues are labeled in green (peptide) and in orange (WW domain). Residues in the WW domain are highlighted in orange.

In all cases the structures were calculated exclusively using experimental dihedral and unambiguously assigned NOE restraints. Secondary structure elements are based on NOEs and carbon chemical shift differences with respect to random coil values.

5.5 Assignments and Coordinates

For each of the five complexes (short names as in Supplementary Table S2) we have deposited a family of structures in the Protein Data Bank and the list of restraints and chemical shifts in the BioMagResBank database. For the complexes the corresponding PDB and BMRB codes are respectively: 2ltv/18498 (YAPWW1_S7), 2ltw/18499 (YAPWW2_S7), 2ltx/18500 (Smurf1WW2_S7), and 2ltz/18502 (Smurf2WW3_S7). Statistics of the five NMR refined complexes analyzed with Procheck and are collected in Table 2, appendix.

Discussion

6. Discussion

Fifty thousand papers after the identification of the TGF- β family of cytokines, the mystery of how and why the consequences of TGF- β signaling are so unpredictable continues to be open. Indeed, the TGF- β hormone can apparently trigger, -with similar ability-, a given function and the opposite (it can inhibit cell proliferation and promote cell growth or enhance cell pluripotency and differentiation, reviewed in (Massagué, 2012)). These apparent paradoxes are caused by the broad spectrum of signals that are initiated by the TGF- β hormone, which are critical for the correct functioning of multicellular organisms. One of the best-studied networks involve a family of transcription factors named receptor activated Smads, which act as mediators in the transmission of the signal created by the interaction of the TGF- β hormone with its membrane receptor. Smads interact directly with the inner part of the receptor for activation, with nucleoporins and with importin- β for nuclear import and export processes, as well as with DNA and with several transcription factors, transcription coactivators or corepressors, which fine-tune the function of Smads in every context (reviewed in (Massagué, 2012; Massagué et al., 2005)). Thus, using a simple approximation, Smad proteins can be seen as a conserved system present in all cells, which can be easily adapted to form versatile functional complexes. In turn, each of these specific complexes carries out numerous tasks required by healthy animal cells to survive and die. Consequently, it is in this plasticity and complexity where the power of the Smad organization resides and where the beauty and risks of the TGF- β signaling cascade are encoded. However, since a few keys to the TGF- β code are still missing -or only partially understood-, the messages we receive are sometimes confusing. With the present work and thanks to the combination of *cell and structural biology approaches* we have investigated a

few critical aspects of the TGF- β signaling network aiming to brighten up some of the dark sides of Smads, specifically how the switch between Smads activation or degradation is defined and how the competition between R-Smads and I-Smads is regulated.

In the first part of the work it is shown that pairs of WW domains in Smad binding proteins function as readers of a phosphoserine code that dictates Smad peak transcriptional action as well as the subsequent elimination of Smad molecules that participate in transcription. The code of this action-turnover switch is written by kinases CDK8/9 and GSK3 acting on the linker region of activated, transcriptionally poised Smad proteins. In our model, CDK8/9 create binding sites that are preferentially recognized by WW-containing Smad transcriptional cofactors. These phosphorylations additionally prime Smads for subsequent GSK3-mediated phosphorylation, which creates sites for ubiquitin ligase binding at the expense of transcriptional cofactor binding sites. Thus, GSK3 switches the phosphorylation code in the Smad linker region from one that favors Smad action to one that favors Smad destruction. As a result, TGF- β /BMP signal delivery becomes coupled to Smad turnover.

In order to explain these observations we propose that degradation is a price that Smad molecules pay for participating in transcription. CDK8 and 9 are components of the CDK8/CyclinC/Med12/Med13 transcriptional mediator complex and the P-TEF β CDK9/CyclinT elongation complex that regulate RNA polymerase II during transcription (Durand et al., 2005; Komarnitsky et al., 2000; Malik and Roeder, 2000). CDK8/9 have access to Smad molecules on the chromatin but not to receptor-activated Smad molecules that fail to engage in transcriptional complexes (Alarcón et al., 2009).

Our work provides a structural and functional basis for the involvement of tandem WW domains in these protein-protein interactions. On one hand,

the WW domains of Smurf1 and YAP proteins achieve overall high specificity in target recognition by acting in pairs, extending the interacting surface to recognize not only the canonical PY site but also the adjacent pS/pTP motifs. The interactions with the pS sites were unexpected from predictions based on sequence conservation since only the WW domain of Pin1 has been described as a pS/pTP binding motif, and the Pin1 residues involved in the phosphate interaction are not strictly conserved in YAP and Smurf1 WW domains.

On the other hand, since the linker region of R-Smads contains a set of CDK phosphorylation sites, and a set of CDK-primed GSK3 sites, the phospho-Ser motifs can be tuned to optimally bind either transcriptional cofactors (mono-phosphorylated) or ubiquitin ligases (di-phosphorylated). The WW2 domains of Smurf1 and YAP recognize a canonical PY motif in Smad1. The PY-independent, mono-phosphoserine motif of Smad1 is recognized by the WW1 domain of YAP, whereas the atypical WW1 domain of Smurf1 domain recognizes di-phosphoserine motifs in Smads 1. The ability of WW domains to recognize the pS(-4)pS motif was previously unknown.

The interaction of these proteins with Smads is determined not only by the binding specificity of individual WW domains but also by the configuration of the two domains. For example, Smurf1 cannot bind Smad3 because the N-terminal position of the pT[PY] motif relative to the pSer cluster in Smad3 is opposite to the orientation required by the WW1 and WW2 domains of Smurf1. The present findings expand the known structural and functional versatility of WW domains as protein-protein interaction modules (Macias et al., 2002).

The recognition of distinct Smad phosphorylation codes by different WW domain proteins provides ample opportunities for regulation. As an adaptor that binds to the WW1-WW2 connector of Smurf1 (Lu et al., 2008), CKIP1 may enforce an optimal orientation of these WW domains for contact with

their cognate sites on the Smad1 linker. A Smurf1 isoform with a longer WW1-WW2 connector exists that may differ in this regulatory function (Schultz et al., 2000). In mammalian cells and in *Drosophila*, YAP enhances certain BMP responses but not others (Alarcón et al., 2009). Similarly, Pin1 enhances certain effects of TGF- β on mammalian cell migration but not other effects (Matsuura et al., 2009). Different factors may fulfill these roles in other contexts or on other target genes. A larger repertoire of Smad linker code-reading factors than presently known may therefore exist.

The action-turnover switch delineated here involves a remarkable concentration of opposing protein-binding functions in a discrete region of the Smad proteins (Figure 9). The fulfillment of dual roles under a phosphorylation-dependent switch is also characteristic of another key component of this pathway, the regulatory GS region of type I TGF- β receptor kinase (Huse et al., 2001). The core TGF- β /Smad pathway is therefore characterized by the economical use that it makes of the structural elements that switch key pathway components from one activation state into another.

In the second part of the work presented here, we have investigated the interaction of YAP and of three ubiquitin ligases with the inhibitor Smad7. The rules governing target recognition by HECT type E3 ubiquitin ligases are open questions, and the work to date reflects a more complex scenario than originally expected. Some family members recognize PY motifs using a single WW domain, as is the case of Itch binding to the Epstein Barr virus protein LMP2A (Morales et al., 2007) and Nedd4 binding to the voltage gated sodium channel (Kanelis et al., 2001) and to Commissureless (Kanelis et al., 2006). In other cases, such as the binding of Smurf1 and Nedd4L to R-Smads, the proteins use a pair of WW domains to expand the binding interface with a composite binding site that includes pSer/pThr-Pro elements in addition to a canonical PY motif, a combination that allows regulation of the interaction by

input-driven protein kinases (Alarcón et al., 2009; Aragón et al., 2011; Gao et al., 2009).

The work here presented show a further versatility of WW domains depending on the target protein in that a given HECT E3 ubiquitin ligase can act singly (this work) or in a combinatorial manner (Alarcón et al., 2009; Aragón et al., 2011; Gao et al., 2009), depending on the target. The interactions of Smurf1 and Smurf2 ubiquitin ligases with Smad7 involve a unique WW domain, whereas Smurf1 binds Smad1 using two WW domains to recognize a PY motif and a phosphorylated motif in the linker region. We observed neither the contacts between the two WW domains nor the contacts between the first WW domain and Smad7 that were described in a recent report (Chong et al., 2011). We observed instead that the Smurf1 and Smurf2 WW-WW pairs have a high tendency to form homo-dimers *via* the WW1 domain in case of Smurf1 and *via* the WW2 domain in Smurf2. The presence of Smad7 peptide did not prevent these dimerizations since the WW-WW domain pairs bind the Smad7 peptide mainly through contacts with the WW2 domain of Smurf1 and with the WW3 domain of Smurf2. It has recently been reported that full-length Smurf1 forms homo-dimers and oligomers *in vitro* and *in vivo* through inter-molecular contacts mapped to a fragment containing the C2 and the WW domains of one molecule and to the HECT domain of the partner (Wan et al., 2011). Furthermore, intra-molecular contacts between the C2 domain and the HECT domain of Smurf2 have also been characterized (Wiesner et al., 2007). In both Smurf1/2 ligases the close conformation inhibits the mechanism of protein self-ubiquitination (Wan et al., 2011). It is possible that in addition to these reported interactions between the C2 and HECT domains, the inter WW-WW contacts detected in our work could also contribute to the formation of the dimers and oligomers *in vivo*, and to the stabilization of the close conformation of Smurf1 and 2. In the presence of two Smad7 equivalents, a reverse reaction may occur with the

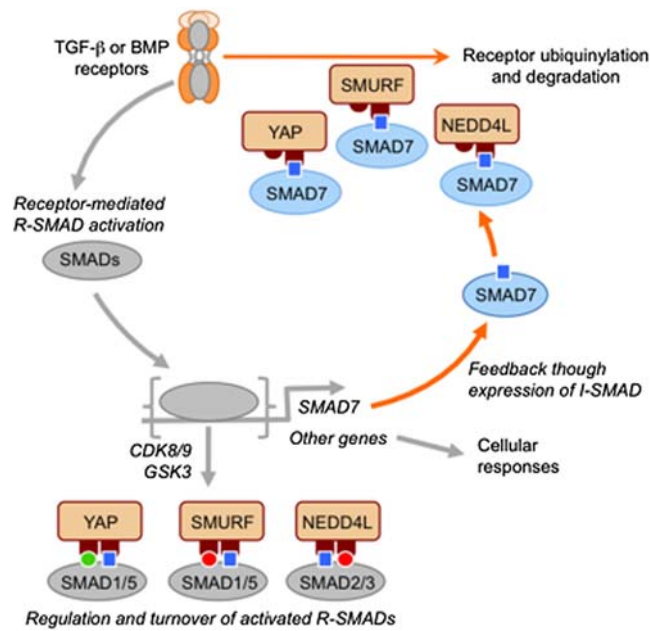
MH1 and PY sites of the Smad7 protein competing for the Smurf1 HECT and the WW2 domains respectively, pulling apart the WW1-WW1 dimer. The result of this reaction would be the generation of two activated Smurf1-Smad7 complexes, starting from the close and inactive full-length dimer. A similar mechanism may occur with the Smurf1 and Smad1 interaction. These possibilities notwithstanding, the WW-WW homodimers here detected could be solely a result of the experimental conditions used *in vitro*.

Our work also shows that YAP requires only the WW1 domain for binding to the Smad7 PY region. Notably, in the interaction with Smad1, YAP uses this WW domain for binding a phosphoserine motif, and instead uses the WW2 domain for binding the Smad1 PY motif (Aragón et al., 2011). By solving the structures of both YAP complexes with Smad7 we discerned the reasons for these differences with respect to the interaction with Smad1. The WW2 domain contains negatively charged residues in the area where the E205 side-chain is normally accommodated, destabilizing the interaction. The affinity of the YAP WW1-WW2 pair for the composed pSP-PY site of Smad1 is 8x higher than that of the WW1 domain for the PY site of Smad7. However, the concentration of Smad7 in the nucleus is high and it could compete *in vivo* with Smad1 for YAP binding, providing a new scenario for the inhibitory role of Smad7.

We propose that WW-WW pairs in these Smad regulators form functional units that evolved to recognize PY containing regions of variable length and complexity, including composite PY/phospho-Ser/Thr motifs in R-Smads and simple PY motifs in Smad7. These features expand the functional versatility of E3 ubiquitin ligases by optimizing the interacting surface depending on the needs. With Smad7, Nedd4L and Smurf1/2 act as partners in targeting TGF- β receptors for ubiquitination. Smad7 may also act as a constitutive YAP sequestration or reservoir protein. In contrast, R-Smads are

direct targets of the ubiquitin ligases and functional partners of YAP only in specific stages of the Smad signaling cycle (Aragón et al., 2011). The absence of a requirement for phosphorylation in the interaction with Smad7 argues that YAP, Nedd4L and Smurf1/2 are constitutive partners of Smad7 whereas they are conditional, phosphorylation dependent regulators of R-Smads in TGF- β and BMP signal transduction (see the Schematic representation of these results in Figure 40). The features of Smad7 defined here provide a structural basis for its central role as a hub for negative feedback and crosstalk regulation in TGF- β signaling.

Figure 40: Schematic representation of the results presented in this work.

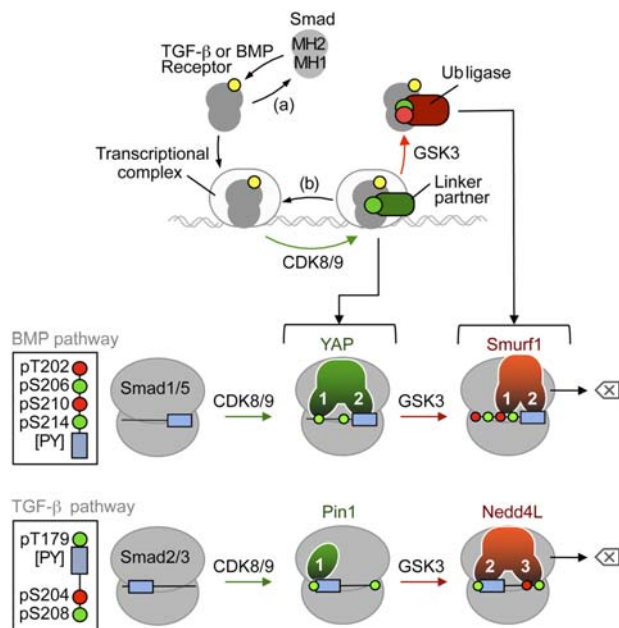


The Smad action turnover switch in the BMP and TGF- β pathways: Structural basis for how regulators that use WW domain pairs for selective interactions with phosphorylated R-Smads, resort to one single WW domain for binding Smad7 to centralize regulation in the TGF- β and BMP pathways.

Following receptor-mediated phosphorylation, Smad proteins translocate to the nucleus and assemble transcriptional complexes, which are phosphorylated at CDK8/9 sites (green circle) in the MH1–MH2 interdomain linker region. This phosphorylation creates high-affinity binding sites for transcriptional partners (such as YAP in the case of the BMP mediator Smad1, Pin1 in the case of the TGF- β mediator Smad3, and probably others), thus achieving peak transcriptional action. Phosphorylation by CDK8/9 also primes the Smads for GSK3-mediated phosphorylation (red symbol) at the -4

position with respect to the CDK8/9 site, which favors the binding of ubiquitin ligases Smurf1 (BMP pathway) and Nedd4L (TGF- β pathway), leading to proteasome-dependent degradation of Smad molecules that participate in transcription. In addition to degradation, the activity of R-Smads is also regulated by the competition of Smad7, which is one of the few genes whose expression is activated by the Smad pathway in most cell types. Smad7 inhibits the function of R-Smads blocking several steps of their function. It binds cofactors as YAP, thus reducing the total amount of available cofactors that facilitate transcription. Smad7 also blocks the receptor site that is occupied by R-Smads for activation, thus reducing the population of activated Smads required for transcription. Furthermore, when Smad7 is bound to the Smurf1/2 and Nedd4L Ubiquitin ligases helps to localize them in the proximity of the TGF- β receptor and by doing so it contributes to receptor degradation.

Figure 41: Schematic of the Smad linker phospho-amino acid codes (insets) and WW domain code readers.



The conserved CDK8/9 phosphorylation sites (green circles) and GSK3 sites (red circles) are located at the indicated positions relative to the PY box (slate box). Amino acid positions correspond to Smad1 and Smad3. In the BMP pathway, the YAP WW1 domain binds to pS206 in Smad1, as long as p210 is not phosphorylated. The Smurf1 WW1 domain binds with higher affinity to the pS210–pS214 motif. The WW2 domains bind the [PY] motif. In the TGF- β pathway, the sole WW domain of Pin1 binds the pT179[PY] motif, as does the WW2 domain of Nedd4L. However, the Nedd4L WW3 domain increases the binding affinity by recognizing the pS204–pS208 motif. See the text for additional details and citations on the known roles of these WW domain proteins in Smad signal transduction. Alternatively, C-terminal Smad phosphatases (a) and linker phosphatases (b) reverse these phosphorylation states. See the text for details and citations. (Bottompanels)

Conclusions

7. Conclusions

Our data reveal a surprisingly independence from phosphorylation in the interactions of Smad7 while phosphorylation is critical for the interaction of R-Smads, with YAP, Smurf1, Smurf2 and Nedd4L proteins. In addition, while pairs of WW domains recognize R-Smads, binding to the inhibitor Smad7 requires the presence of single WW domains. These observations illuminate the functional versatility of WW domain containing proteins and also of the WW domains as mediators of specific interactions with Smad proteins.

The results of this work are highlighted with the following four conclusions:

1. E3 ubiquitin ligases of the HECT type and the transcription coactivator YAP can use single or multiple WW domains to select Smad protein targets.
2. The targets can require multiple binding sites, as in R-Smads or unique sites as in I-Smads.
3. Binding sites in R-Smads are phosphorylation dependent while the I-Smad site is independent of phosphorylation.
4. In Smurf1 WW-WW homodimers detected by NMR and IM-MS techniques can stabilize the close conformation of the ligase, previously described *in vivo*.

7.1 REFERENCES

Alarcón, C., Zaromytidou, A.I., Xi, Q., Gao, S., Yu, J., Fujisawa, S., Barlas, A., Miller, A.N., Manova-Todorova, K., Macias, M.J., *et al.* (2009). Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways. *Cell* 139, 757-769.

Aragón, E., Goerner, N., Zaromytidou, A.I., Xi, Q., Escobedo, A., Massagué, J., and Macias, M.J. (2011). A Smad action turnover switch operated by WW domain readers of a phosphoserine code. *Genes Dev.* 25, 1275-1288.

Bai, S., and Cao, X. (2002). A nuclear antagonistic mechanism of inhibitory Smads in transforming growth factor-beta signaling. *J. Biol. Chem.* 277, 4176-4182.

Bartels, C., Xia, T.H., Billeter, M., Güntert, P., and Wüthrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* 5, 1-10.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.* (1998). Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* 54, 905-921.

Chong, P.A., Lin, H., Wrana, J.L., and Forman-Kay, J.D. (2006). An expanded WW domain recognition motif revealed by the interaction between Smad7 and the E3 ubiquitin ligase Smurf2. *J. Biol. Chem.* 281, 17069-17075.

Chong, P.A., Lin, H., Wrana, J.L., and Forman-Kay, J.D. (2011). Coupling of tandem Smad ubiquitination regulatory factor (Smurf) WW domains modulates target specificity. *Proc. Natl. Acad. Sci. USA* 107, 18404-18409.

Cohen, P., and Frame, S. (2001). The renaissance of GSK3. *Nat. Rev. Mol. Cell. Biol.* 2, 769-776.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277-293.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System.

Durand, L.O., Advani, S.J., Poon, A.P., and Roizman, B. (2005). The carboxyl-terminal domain of RNA polymerase II is phosphorylated by a complex containing cdk9 and infected-cell protein 22 of herpes simplex virus 1. *J. Virol.* *79*, 6757-6762.

Ebisawa, T., Fukuchi, M., Murakami, G., Chiba, T., Tanaka, K., Imamura, T., and Miyazono, K. (2001). Smurfi interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation. *J. Biol. Chem.* *276*, 12477-12480.

Eichhorn, P.J., Rodón, L., González-Junca, A., Dirac, A., Gili, M., Martínez-Sáez, E., Aura, C., Barba, I., Peg, V., Prat, A., *et al.* (2012). USP15 stabilizes TGF-beta receptor I and promotes oncogenesis through the activation of TGF-beta signaling in glioblastoma. *Nat. Med.* *18*, 429-435.

Ferrigno, O., Lallemand, F., Verrecchia, F., L'Hoste, S., Camonis, J., Atfi, A., and Mauviel, A. (2002). Yes-associated protein (YAP65) interacts with Smad7 and potentiates its inhibitory activity against TGF-beta/Smad signaling. *Oncogene* *21*, 4879-4884.

Fuentealba, L.C., Eivers, E., Ikeda, A., Hurtado, C., Kuroda, H., Pera, E.M., and De Robertis, E.M. (2007). Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal. *Cell* *131*, 980-993.

Gao, S., Alarcón, C., Sapkota, G., Rahman, S., Chen, P.Y., Goerner, N., Macias, M.J., Erdjument-Bromage, H., Tempst, P., and Massagué, J. (2009). Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF-beta signaling. *Mol. Cell* *36*, 457-468.

Hata, A., Lagna, G., Massagué, J., and Hemmati-Brivanlou, A. (1998). Smad6 inhibits BMP/Smad1 signaling by specifically competing with the Smad4 tumor suppressor. *Genes Dev.* *12*, 186-197.

Hayashi, H., Abdollah, S., Qiu, Y., Cai, J., Xu, Y.Y., Grinnell, B.W., Richardson, M.A., Topper, J.N., Gimbrone, M.A., Jr., Wrana, J.L., *et al.* (1997). The MAD-related protein Smad7 associates with the TGFbeta receptor and functions as an antagonist of TGFbeta signaling. *Cell* *89*, 1165-1173.

He, W., Li, A.G., Wang, D., Han, S., Zheng, B., Goumans, M.J., Ten Dijke, P., and Wang, X.J. (2002). Overexpression of Smad7 results in severe pathological alterations in multiple epithelial tissues. *EMBO J.* *21*, 2580-2590.

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* *4*, 435-447.

- Huse, M., Muir, T.W., Xu, L., Chen, Y.G., Kuriyan, J., and Massagué, J. (2001). The TGF beta receptor activation process: an inhibitor- to substrate-binding switch. *Mol. Cell* *8*, 671-682.
- Inman, G.J., Nicolas, F.J., and Hill, C.S. (2002). Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity. *Mol. Cell* *10*, 283-294.
- Kanelis, V., Bruce, M.C., Skrynnikov, N.R., Rotin, D., and Forman-Kay, J.D. (2006). Structural determinants for high-affinity binding in a Nedd4 WW3* domain-Comm PY motif complex. *Structure* *14*, 543-553.
- Kanelis, V., Rotin, D., and Forman-Kay, J.D. (2001). Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nat. Struct. Biol.* *8*, 407-412.
- Kavsak, P., Rasmussen, R.K., Causing, C.G., Bonni, S., Zhu, H., Thomsen, G.H., and Wrana, J.L. (2000). Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. *Mol. Cell* *6*, 1365-1375.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* *14*, 2452-2460.
- Kretzschmar, M., Doody, J., and Massagué, J. (1997). Opposing BMP and EGF signalling pathways converge on the TGF-beta family mediator Smad1. *Nature* *389*, 618-622.
- Kretzschmar, M., Doody, J., Timokhina, I., and Massagué, J. (1999). A mechanism of repression of TGFbeta/ Smad signaling by oncogenic Ras. *Genes Dev.* *13*, 804-816.
- Kuratomi, G., Komuro, A., Goto, K., Shinozaki, M., Miyazawa, K., Miyazono, K., and Imamura, T. (2005). NEDD4-2 negatively regulates TGF-beta signaling by inducing ubiquitin-mediated degradation of Smad2 and TGF-beta type I receptor. *Biochem. J.* *386*, 461-470.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* *8*, 477-486.
- Lin, X., Duan, X., Liang, Y.Y., Su, Y., Wrighton, K.H., Long, J., Hu, M., Davis, C.M., Wang, J., Brunicardi, F.C., *et al.* (2006). PPM1A functions as a Smad phosphatase to terminate TGFbeta signaling. *Cell* *125*, 915-928.
- Lo, R.S., and Massagué, J. (1999). Ubiquitin-dependent degradation of TGF-beta-activated smad2. *Nat. Cell Biol.* *1*, 472-478.

Lu, K., Yin, X., Weng, T., Xi, S., Li, L., Xing, G., Cheng, X., Yang, X., Zhang, L., and He, F. (2008). Targeting WW domains linker of HECT-type ubiquitin ligase Smurf1 for activation by CKIP-1. *Nat. Cell Biol.* 10, 994-1002.

Macias, M.J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* 382, 646-649.

Macias, M.J., Wiesner, S., and Sudol, M. (2002). WW and SH₃ domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.* 513, 30-37.

Malik, S., and Roeder, R.G. (2000). Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends Biochem. Sci.* 25, 277-283.

Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labelling of recombinant proteins. *J. Biomol. NMR* 20, 71-75.

Massagué, J. (1998). TGF-beta signal transduction. *Annu. Rev. Biochem.* 67, 753-791.

Massagué, J. (2012). TGF-beta signaling in context. *Nature reviews. Molecular cell biology* 13, 616-630.

Massagué, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. *Genes Dev.* 19, 2783-2810.

Matsuura, I., Chiang, K.N., Lai, C.Y., He, D., Wang, G., Ramkumar, R., Uchida, T., Ryo, A., Lu, K., and Liu, F. (2009). Pini promotes transforming growth factor-beta-induced migration and invasion. *J. Biol. Chem.* 285, 1754-1764.

Morales, B., Ramirez-Espain, X., Shaw, A.Z., Martin-Malpartida, P., Yraola, F., Sanchez-Tillo, E., Farrera, C., Celada, A., Royo, M., and Macias, M.J. (2007). NMR structural studies of the ItchWW₃ domain reveal that phosphorylation at T30 inhibits the interaction with PPxY-containing ligands. *Structure* 15, 473-483.

Nakao, A., Afrakhte, M., Moren, A., Nakayama, T., Christian, J.L., Heuchel, R., Itoh, S., Kawabata, M., Heldin, N.E., Heldin, C.H., *et al.* (1997). Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. *Nature* 389, 631-635.

Nilges, M., Macias, M.J., O'Donoghue, S.I., and Oschkinat, H. (1997). Automated NOESY Interpretation with Ambiguous Distance Restraints: The Refined NMR Solution Structure of the Pleckstrin Homology Domain from beta-Spectrin. *J. Mol. Biol.* 269, 408-422.

Ohnishi, S., Guntert, P., Koshiba, S., Tomizawa, T., Akasaka, R., Tochio, N., Sato, M., Inoue, M., Harada, T., Watanabe, S., *et al.* (2007). Solution structure of an

atypical WW domain in a novel beta-clam-like dimeric form. *FEBS Lett.* *581*, 462-468.

Pierce, M.M., Raman, C.S., and Nall, B.T. (1999). Isothermal titration calorimetry of protein-protein interactions. *Methods* *19*, 213-221.

Pires, J.R., Taha-Nejad, F., Toepert, F., Ast, T., Hoffmuller, U., Schneider-Mergener, J., Kuhne, R., Macias, M.J., and Oschkinat, H. (2001). Solution structures of the YAP65 WW domain and the variant L30 K in complex with the peptides GTPPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. *J. Mol. Biol.* *314*, 1147-1156.

Ramirez-Espain, X., Ruiz, L., Martin-Malpartida, P., Oschkinat, H., and Macias, M.J. (2007). Structural characterization of a new binding motif and a novel binding mode in group 2 WW domains. *J. Mol. Biol.* *373*, 1255-1268.

Sapkota, G., Alarcón, C., Spagnoli, F.M., Brivanlou, A.H., and Massagué, J. (2007). Balancing BMP signaling through integrated inputs into the Smad1 linker. *Mol. Cell* *25*, 441-454.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.* *34*, 93-158.

Schmierer, B., Tournier, A.L., Bates, P.A., and Hill, C.S. (2008). Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. *Proc. Natl. Acad. Sci. USA* *105*, 6608-6613.

Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., and Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* *28*, 231-234.

Shi, Y., and Massagué, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* *113*, 685-700.

Shi, Y., Wang, Y.F., Jayaraman, L., Yang, H., Massagué, J., and Pavletich, N.P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell* *94*, 585-594.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* *25*, 4876-4882.

Toepert, F., Pires, J.R., Landgraf, C., Oschkinat, H., and Schneider-Mergener, J. (2001). Synthesis of an Array Comprising 837 Variants of the hYAP WW Protein Domain This work was supported by the DFG (INK 16/B1-1), by the Fonds der

Chemischen Industrie, and by the Universitätsklinikum Charité Berlin. *Angew. Chem. Int. Ed. Engl.* 40, 897-900.

Topper, J.N., DiChiara, M.R., Brown, J.D., Williams, A.J., Falb, D., Collins, T., and Gimbrone, M.A., Jr. (1998). CREB binding protein is a required coactivator for Smad-dependent, transforming growth factor beta transcriptional responses in endothelial cells. *Proc. Natl. Acad. Sci. USA* 95, 9506-9511.

Ulloa, L., Doody, J., and Massagué, J. (1999). Inhibition of transforming growth factor-beta/SMAD signalling by the interferon-gamma/STAT pathway. *Nature* 397, 710-713.

Wan, L., Zou, W., Gao, D., Inuzuka, H., Fukushima, H., Berg, A.H., Drapp, R., Shaik, S., Hu, D., Lester, C., *et al.* (2011). Cdh1 regulates osteoblast function through an APC/C-independent modulation of Smurf1. *Mol. Cell* 44, 721-733.

Wiesner, S., Ogunjimi, A.A., Wang, H.-R., Rotin, D., Sicheri, F., Wrana, J.L., and Forman-Kay, J.D. (2007). Autoinhibition of the HECT-Type ubiquitin ligase Smurf2 through its C2 domain. *Cell* 130, 651-662.

Wrighton, K.H., Willis, D., Long, J., Liu, F., Lin, X., and Feng, X.H. (2006). Small C-terminal domain phosphatases dephosphorylate the regulatory linker regions of Smad2 and Smad3 to enhance transforming growth factor-beta signaling. *J. Biol. Chem.* 281, 38365-38375.

Wu, D., and Pan, W. (2010). GSK3: a multifaceted kinase in Wnt signaling. *Trends Biochem. Sci.* 35, 161-168.

Wu, J.W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D.J., Kyin, S., Muir, T.W., Fairman, R., *et al.* (2001). Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. *Mol. Cell* 8, 1277-1289.

Xu, L., Kang, Y., Col, S., and Massagué, J. (2002). Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGFbeta signaling complexes in the cytoplasm and nucleus. *Mol. Cell* 10, 271-282.

Yan, X., and Chen, Y.G. (2011). Smad7: not only a regulator, but also a cross-talk mediator of TGF-beta signalling. *Biochem. J.* 434, 1-10.

Appendix

8. APPENDIX

8.1 Table 1

	<i>Smurf1-Smad1</i>			<i>YAP-Smad1</i>		
	WW1-pS214	WW1- pS210pS214	WW2-PY	WW1- pS206	WW1- pT202pS20 6	WW2-PY
Restraints used for the calculation <SA>^(a)						
Interdomain	41	75	80	33	23	66
Sequential (i-j = 1)	180	186	171	172	170	158
Medium range (1 < i-j ≤ 4)	87	89	51	81	81	71
Long range (i-j > 4)	219	253	308	258	248	234
Dihedrals	63	64	91	71	68	90
Hydrogen bonds	10	10	10	10	10	10
All restraints (unambiguous)	559	677	711	625	600	629
Restraint ratio (WW-ligand)	15.5 (30-6)	16.9 (30-10)	16.9 (30-12)	14.9 (33-9)	15.8 (33-5)	14.3 (32-12)

R.m.s. deviation (Å) from experimental (b)						
NOE (x10 ⁻³):	14.1+/-0.6	9.1+/-0.3	7.1+/-0.3	3.1+/-0.4	3.3+/-0.3	18.1+/-0.5
Bonds (x10 ⁻³) (Å)	8.9+/-0.3	7.0+/-0.2	6.4+/-0.2	4.1+/-0.2	5.1+/-0.2	9.7+/-0.3
Angles (°)	0.9+/-0.03	0.8+/-0.02	0.7+/-0.02	0.5+/-0.03	0.6+/-0.03	0.9+/-0.03
Coordinate Precision (Å) (c)						
Backbone, all residues in the complex	0.35	0.24	0.47	0.46	0.84	0.5
CNS potential energy (kcal mol⁻¹)						
Total energy ^(d)	-1079 +/- 16.18	-1300 +/- 25.34	-1342 +/- 30.77	-1644 +/- 34.48	-1737 +/- 28.57	-821.6 +/- 41.91
Electrostatic	-1593 +/- 28.52	-1749 +/- 33.27	-1747 +/- 44.04	-1840 +/- 37.95	-1946 +/- 44.54	-1519 +/- 49.21
Van der Waals	-50.02 +/- 13.35	-57.61 +/- 10.71	-101.1 +/- 9.308	-149.6 +/- 9.842	-139.5 +/- 14.95	-29.89 +/- 22.3
Bonds	52.37 +/- 3.993	33.45 +/- 1.932	31.81 +/- 2.211	11.88 +/- 1.538	12.27 +/- 1.12	69.53 +/- 4.007
Angles	150.1 +/- 6.577	139.9 +/- 8.802	120.3 +/- 7.151	64.7 +/- 8.816	65.67 +/- 6.167	155.4 +/- 12.58
Structural quality (%residues) 20 best structures						
In most favored region of Ramachandran plot	85.3	84.4	87.5	86.2	87.2	80
In additionally allowed region	14.7	15.6	12.2	13.7	12.2	20

8.2 Table 2

	<i>Smurf1 WW2-Smad7</i>	<i>Smurf2WW3-Smad7</i>	<i>YAPWW1-Smad7</i>	<i>YAPWW2-Smad7</i>
Interdomain	75	80	33	23
Sequential ($ i-j = 1$)	186	171	172	170
Medium range ($1 < i-j \leq 4$)	89	51	81	81
Long range ($ i-j > 4$)	253	308	258	248
Dihedrals	64	91	71	68
Hydrogen bonds	10	10	10	10
All restraints (unambiguous)	677	711	625	600
Restraint ratio (WW-ligand)	16.9 (30-10)	16.9 (30-12)	14.9 (33-9)	15.8 (33-5)
Bonds ($\times 10^{-3}$) (Å)	9.1 \pm 0.3	7.1 \pm 0.3	3.1 \pm 0.4	3.3 \pm 0.3
Angles (°)	7.0 \pm 0.2	6.4 \pm 0.2	4.1 \pm 0.2	5.1 \pm 0.2
NOE ($\times 10^{-3}$):	0.8 \pm 0.02	0.7 \pm 0.02	0.5 \pm 0.03	0.6 \pm 0.03
Backbone, all residues in the complex	0.24	0.47	0.46	0.84
Total energy ^(d)	-1300 \pm 25.34	-1342 \pm 30.77	-1644 \pm 34.48	-1737 \pm 28.57
Electrostatic	-1749 \pm 33.27	-1747 \pm 44.04	-1840 \pm 37.95	-1946 \pm 44.54
Van der Waals	-57.61 \pm 10.71	-101.1 \pm 9.308	-149.6 \pm 9.842	-139.5 \pm 14.95
Bonds	33.45 \pm 1.932	31.81 \pm 2.211	11.88 \pm 1.538	12.27 \pm 1.12
Angles	139.9 \pm 8.802	120.3 \pm 7.151	64.7 \pm 8.816	65.67 \pm 6.167
In most favored region of Ramachandran plot	84.4	87.5	86.2	87.2
In additionally allowed region	15.6	12.2	13.7	12.2

8.3 *Curriculum Vitae*

Personal Information

Name : Eric Aragón Altarriba

Birthplace and date

31-05-77, Barcelona, Spain

Electronic mail

eric.aragon@irbbarcelona.org

Education

Degree on Biology, Universitat de Barcelona, 2002

Degree on Biochemistry, Universitat de Barcelona, 2002

DEA

Job Experience

At present, I am a Research Associated at Dr. Maria J. Macias Group, at the Institute for Research in Biomedicine.

Scientific Publications

Aragón E, Goerner N, Xi Q, Gomes T, Gao S, Massagué J, Macias MJ.

Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF- β Pathways

Structure. 2012 Oct 10;20(10):1726-36

Martín-Gago P, Gomez-Caminals M, Ramón R, Verdaguer X, Martin-Malpartida P, **Aragón E**, Fernández-Carneado J, Ponsati B, López-Ruiz P, Cortes MA, Colás B, Macias MJ, Riera A.

Fine-tuning the π - π aromatic interactions in peptides: somatostatin analogues containing mesityl alanine

Angewandte Chemie International Edition Engl. 2012 Feb 20;51(8):1820-5

Aragón E., Goerner, N., Zaromytidou, A.I., Xi, Q., Escobedo A., Massagué, J. and Macias, M.J.

A Smad action-turnover switch operated by WW domain readers of a phosphoserine code

Genes & Development 2011 Jun 15: Vol.25 No.12 1275-1288

Bonet R, Ruiz L, **Aragón E**, Martin-Malpartida P, Macias MJ.

NMR structural studies on human p190-A RhoGAPFF1 revealed that domain phosphorylation by the PDGF-receptor alpha requires its previous unfolding

Journal of Molecular Biology 2009 Jun 5;389(2):230-7

Miguel Feliz, Jesús García, **Eric Aragón**, Miquel Pons

Fast 2D-NMR ligand screening using Hadamart spectroscopy

Journal of the American Chemical Society. 2006 Jun 7;128(22):7146-7

Conferences and Workshops

NMR course Universidad de Barcelona. Barcelona 22 March-2 April 2004

Taller de Resonancia Magnética Nuclear de sistemas paramagnéticos y diamagnéticos. Interacciones Moleculares. Barcelona 20, 21 y 22 de April, 2004

Dinàmica molecular de proteïnes i àcids nucleics. Barcelona 7 -9 de february, 2005

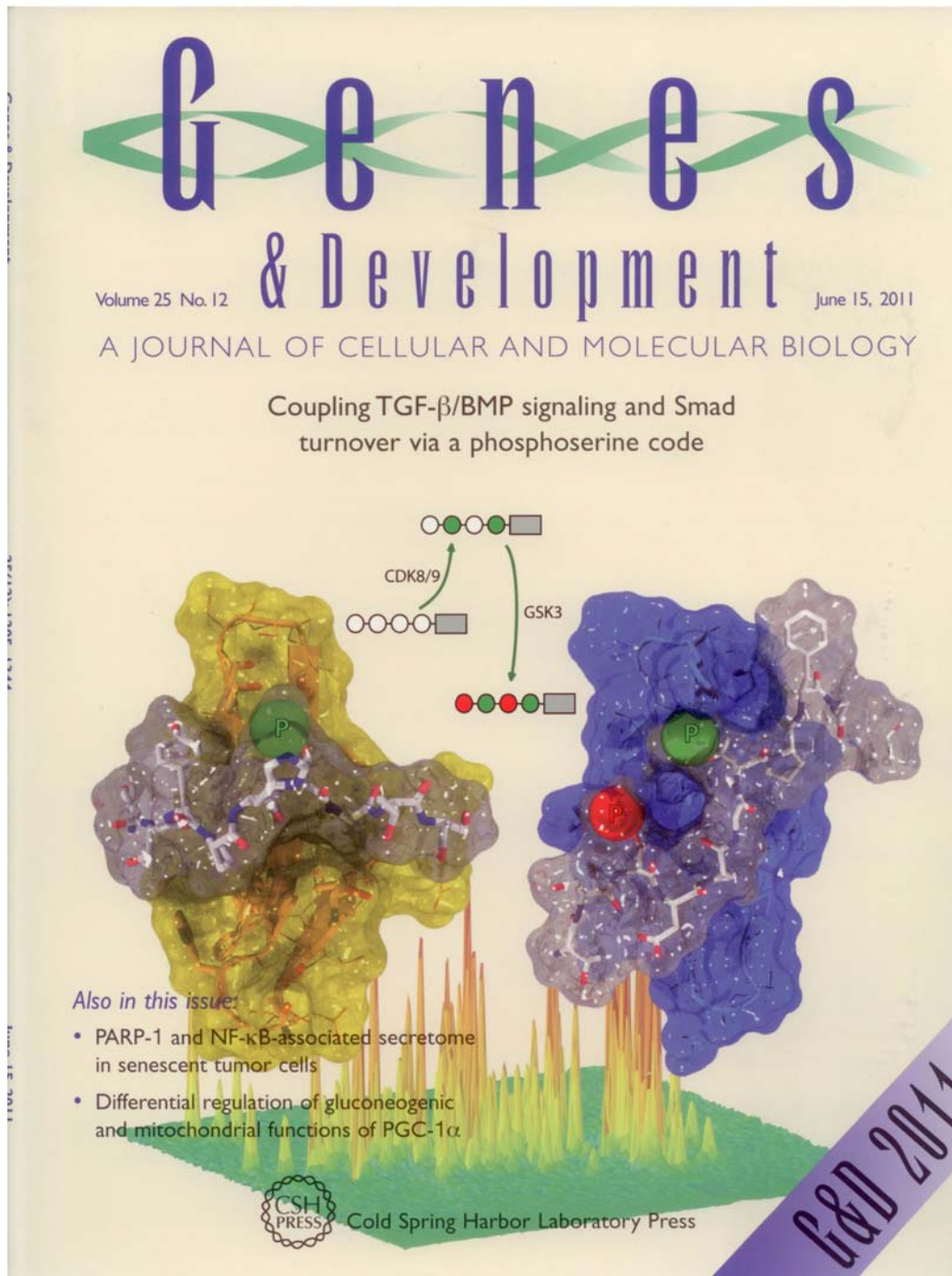
Espectroscopía de RMN Hadamard. Barcelona 22 de February de 2005

Curso avanzado de Resonancia Magnética Nuclear. Jaca 13-17 june 2005

Embo Practical Course: High throughput methods for protein production and cristallization AFMB Marseille 4-13 July 2011

8.4 Reprints

A Smad action-turnover switch operated by WW domain readers of a phosphoserine code



A phosphoserine code regulates Smad protein function and turnover in TGF- β /BMP pathways. Shown here is a composite image depicting this process: Upon activation by BMP or TGF- β receptors, Smad transcription factors (grey space-filling models) are phosphorylated by CDK8/9 (green) and GSK3 (red) kinases. This phosphoserine code is read by the WW domains of transcriptional co-factors such as YAP (yellow), and then targeted for degradation by WW domain-containing ubiquitin ligases, like Smurf1(blue). Shown underneath is a 3D plot of the NMR spectrum acquired for a complex of the YAP WW1-WW2 fragment bound to the phosphorylated Smad1 linker region.

A Smad action turnover switch operated by WW domain readers of a phosphoserine code

Eric Aragón,^{1,5} Nina Goerner,^{1,5} Alexia-Ileana Zaromytidou,² Qiaoran Xi,² Albert Escobedo,¹ Joan Massagué,^{2,3,6,7} and Maria J. Macias^{1,4,6}

¹Structural and Computational Biology Programme, Institute for Research in Biomedicine, 08028 Barcelona, Spain; ²Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA; ³Howard Hughes Medical Institute (HHMI), Chevy Chase, Maryland 20185, USA; ⁴Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

When directed to the nucleus by TGF- β or BMP signals, Smad proteins undergo cyclin-dependent kinase 8/9 (CDK8/9) and glycogen synthase kinase-3 (GSK3) phosphorylations that mediate the binding of YAP and Pin1 for transcriptional action, and of ubiquitin ligases Smurf1 and Nedd4L for Smad destruction. Here we demonstrate that there is an order of events—Smad activation first and destruction later—and that it is controlled by a switch in the recognition of Smad phosphoserines by WW domains in their binding partners. In the BMP pathway, Smad1 phosphorylation by CDK8/9 creates binding sites for the WW domains of YAP, and subsequent phosphorylation by GSK3 switches off YAP binding and adds binding sites for Smurf1 WW domains. Similarly, in the TGF- β pathway, Smad3 phosphorylation by CDK8/9 creates binding sites for Pin1 and GSK3, then adds sites to enhance Nedd4L binding. Thus, a Smad phosphoserine code and a set of WW domain code readers provide an efficient solution to the problem of coupling TGF- β signal delivery to turnover of the Smad signal transducers.

[*Keywords:* signal transduction; transcription; Smad; BMP; TGF- β ; WW domains]

Supplemental material is available for this article.

Received April 19, 2011; revised version accepted May 16, 2011.

An important gap in the current understanding of cytokine-driven transcriptional control is about the processes that remove mediator molecules that have participated in gene regulation. A case in point is the Smad transcription factors, central mediators of the TGF- β and BMP pathways (Li and Flavell 2008; Massagué 2008; Wu and Hill 2009). Fundamental aspects of metazoan embryo development and tissue homeostasis are controlled by TGF- β and BMP through Smad-mediated transcription of master regulator genes. In the course of this action in the nucleus, Smad proteins undergo certain phosphorylation events that enable peak transcriptional activity but also mark the proteins for destruction (Alarcon et al. 2009; Gao et al. 2009). These findings presented a paradox, but also an opportunity to define how the delivery of TGF- β and BMP signals is coupled to the turnover of the Smad signal transducers.

Several key phosphorylations drive the Smad signaling process. The ligand cytokines activate receptor serine/threonine protein kinases that phosphorylate Smad proteins at the C terminus. The BMP receptors act on Smad1, Smad5, and Smad8, and the receptors for the TGF- β /nodal/activin/myostatin group of ligands act mainly on Smad2 and Smad3 (Shi and Massagué 2003). The phosphorylated C terminus provides a binding site for Smad4, which is an essential component in the assembly of target-specific transcriptional complexes. These phosphorylations are reversed by protein phosphatases that limit the general pool of activated Smad molecules (Inman et al. 2002; Xu et al. 2002; Lin et al. 2006; Schmierer et al. 2008).

Receptor-activated Smad proteins that associate with Smad4 and bind to target genes undergo a second set of phosphorylations; these are catalyzed by the transcriptional cyclin-dependent kinases CDK8 and CDK9 (Alarcon et al. 2009; Gao et al. 2009) and glycogen synthase kinase-3 (GSK3) (Fuentelba et al. 2007; Sapkota et al. 2007; Alarcon et al. 2009). CDK8 and CDK9 are part of the transcriptional Mediator and Elongation complexes, respectively (Komarnitsky et al. 2000; Malik and Roeder 2000; Durand et al. 2005). GSK3 is a Wnt- and PI3K-regulated kinase (Cohen and Frame 2001; Wu and Pan

⁵These authors contributed equally to this work.

⁶These authors contributed equally to this work.

⁷Corresponding author.

E-mail j-massague@ski.mskcc.org.

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.2060811>.

Freely available online through the *Genes & Development* Open Access option.

2010). CDK8/9 phosphorylation of Smad serves as the priming event for phosphorylation by GSK3. These phosphorylations are clustered in an interdomain linker region and enable peak activation of Smads, but also mark the proteins for polyubiquitination and proteasome-mediated degradation (Alarcon et al. 2009; Gao et al. 2009). Degradation of agonist-activated Smads (Lo and Massagué 1999; Alarcon et al. 2009) occurs alongside dephosphorylation of the linker (Wrighton et al. 2006; Sapkota et al. 2007). Whereas dephosphorylation recycles the Smad proteins for repeated rounds of signaling, action-coupled destruction of Smad depletes the pool of signal transducer. In a different context, the Smad linker region is phosphorylated by MAP kinases and cell division CDKs in response to mitogens and stresses to constrain TGF- β and BMP signaling (Kretzschmar et al. 1997, 1999; Matsuura et al. 2009).

Four proteins are known to bind specifically to linker phosphorylated Smads during BMP and TGF- β signal transduction. The HECT domain ubiquitin ligase Smurf1 (Sapkota et al. 2007) and the transcriptional effector of the Hippo pathway YAP bind to linker phosphorylated Smad1/5 (Alarcon et al. 2009), whereas the Smurf1-related protein Nedd4L (Gao et al. 2009) and the peptidyl-prolyl *cis/trans* isomerase Pin1 (Matsuura et al. 2009) bind to linker phosphorylated Smad2/3. YAP cooperates with Smad1 to activate *ID* genes that suppress neural differentiation in mouse embryonic stem cells in response to BMP signals (Alarcon et al. 2009). Pin1 cooperates with Smad2/3 to stimulate cancer cell migration in response to TGF- β (Matsuura et al. 2009). Smurf1 and Nedd4L target activated Smad1/5 and Smad2/3, respectively, for polyubiquitination and proteasome-dependent degradation. Common to this set of Smad-binding proteins is the presence of WW domains: one in Pin1, two in Smurf1 and YAP, and four in Nedd4L. WW domains are 38- to 40-amino-acid residue units characterized by two highly conserved tryptophans and folded as a three-strand β sheet that typically binds proline-rich sequences (e.g., PPxY or "PY box") or, in the case of Pin1, phospho-SP motifs (Macias et al. 2002). A PY box is located near the CDK/GSK3 phosphorylation sites in the linker region of Smad proteins.

These lines of evidence present a scenario in which different nuclear protein kinases phosphorylate agonist-activated Smads to create docking sites for competing transcriptional cofactors and ubiquitin ligases. The outcome of these interactions governs Smad function, and is therefore important in BMP and TGF- β signal transduction. However, the convergence of activation and turnover functions on a clustered set of Smad modifications raises questions about how Smads get to act before undergoing disposal. We postulated that a mechanism must exist that ensures the orderly sequence of events in this process by somehow switching Smad proteins from binding transcriptional cofactors to binding ubiquitin ligases. Combining the power of functional and structural approaches, we uncovered such a switch mechanism and defined the basis for its operation and specificity in the BMP and TGF- β pathways.

Results

GSK3 switches the Smad1 binding preference from YAP to Smurf1

Smad proteins consist of a globular N-terminal MH1 (Mad Homology 1) domain with DNA-binding activity, a C-terminal MH2 domain that mediates key protein-protein interactions, and an interdomain linker region with a conserved cluster of phosphorylation sites adjacent to a PY motif (Fig. 1A,B; Shi and Massagué 2003). Phosphorylation of these sites follows TGF- β - and BMP-driven C-terminal phosphorylation and nuclear translocation of Smads, as seen in human cell lines, mouse embryonic stem cells, the mouse embryo, and the *Xenopus* embryo (Supplemental Fig. 1A–C; Fuentealba et al. 2007; Sapkota et al. 2007; Alarcon et al. 2009). In Smad1, CDK8/9 phosphorylate S206 and S214, which prime T202 and S210, respectively, for phosphorylation by GSK3. To dissect this process, we tested the effect of pharmacological inhibitors of CDK8/9 and GSK3 in human embryonic kidney 293 (HEK293) cells expressing epitope-tagged Smurf1 or YAP constructs. A catalytically inactive Smurf1 mutant (Smurf1DD) (Ebisawa et al. 2001) was used in order to avoid confounding the effects of Smurf1-dependent Smad degradation. The BMP inhibitor noggin was added to the culture medium in order to block endogenous BMP and thus set a basal state. Incubation of the cells with BMP rapidly induced the formation of Smad1–YAP and Smad1–Smurf1 complexes (Fig. 1C,D). The CDK8/9 inhibitor flavopiridol, which inhibits all BMP-induced linker phosphorylations (Alarcon et al. 2009), prevented the formation of both complexes (Fig. 1C,D). Addition of LiCl, which inhibits GSK3 site phosphorylation (Fuentealba et al. 2007), also prevented the Smad1–Smurf1 interaction (Fig. 1C). Interestingly, LiCl did not inhibit, but rather increased, the level of Smad1–YAP complex (Fig. 1D). These results suggested that the formation of the YAP–Smad1 complex in response to BMP requires CDK8/9 but not GSK3, whereas the formation of the Smurf1–Smad1 complex requires both kinase activities.

We performed isothermal titration calorimetry (ITC) binding assays to investigate the interaction between recombinant WW1–WW2 segments and Smad1 linker phosphopeptides (Fig. 1E,F). We tested versions of the Smad1 199–233 linker region with no phosphorylation, with phosphorylation at CDK8/9 sites S206 and S214, or with additional phosphorylation at GSK3 sites T202 and S210 (Fig. 1F). The YAP WW1–WW2 segment bound the unphosphorylated Smad1 peptide with $K_D = 19.0 \pm 3 \mu\text{M}$ and the CDK8/9-phosphorylated peptide with $K_D = 8.4 \pm 1 \mu\text{M}$. Notably, this gain in affinity was fully erased by phosphorylation at the GSK3 sites ($K_D = 60.6 \pm 7$) (Fig. 1E,F). In contrast, the affinity of the Smurf1 WW1–WW2 segment for the Smad1 peptide was increased by phosphorylation at the CDK8/9 sites, and was further increased by phosphorylation at the GSK3 sites (Fig. 1E,F).

Further refinement of the Smurf1–Smad1 interaction revealed a strong preference of Smurf1 for pS214 over pS206, achieving the highest affinity ($K_D = 1.2 \pm 0.3 \mu\text{M}$) with a Smad1 208–233 peptide containing pS210 and

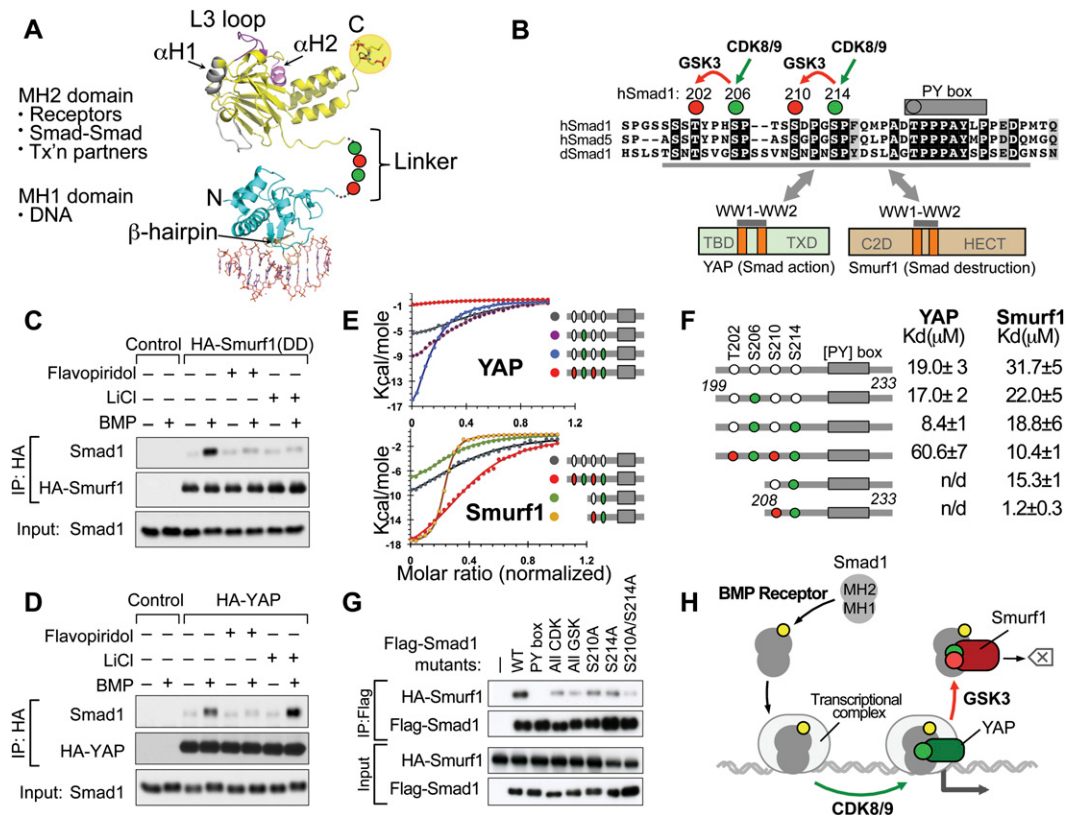


Figure 1. GSK3 switches the binding preference of Smad1 from YAP to Smurf1. (A) Schematic representation of the Smad protein domains and their main functions. The MH1 domain (cyan) contains a β hairpin that mediates binding to dsDNA (orange) (PDB code: 1MHD) (Shi et al. 1998). The MH2 domain (yellow) binds to the type I TGF- β receptor, which involves the L3 loop (magenta); to Smad4 via the phosphorylated C terminus (highlighted) and the α -helix 1 (gray); and to various DNA-binding cofactors and histone-modifying enzymes (PDB code: 1KHJ) (Wu et al. 2001). The interdomain linker region (dotted line) contains CDK8/9 and GSK3 phosphorylation sites, represented by green and red circles, respectively. (B) Sequence alignment of the linker region of human Smad1 and Smad5 and *Drosophila* MAD (dSmad1) proteins, with conserved residues highlighted. The conserved CDK8/9 sites (green) and CDK8/9-primed GSK3 sites (red) and the PY box are shown. The Smad1 (199–232) segment used in this study is underlined. The domain composition of Smurf1 and YAP proteins and the regions that mediate binding to linker phosphorylated Smad1 are indicated. (C) BMP-dependent formation of a complex between HA-Smurf1(DD) and endogenous Smad1 in HEK293 cells, and effects of flavopiridol and LiCl on the formation of this complex. (D) BMP-dependent formation of a complex between HA-YAP and endogenous Smad1 in HEK293 cells, and effects of flavopiridol and LiCl on the formation of this complex. (E) ITC curves for the binding of Smurf1 and YAP WW1–WW2 segments to Smad1 synthetic peptides. (F) Synthetic Smad1 (phospho-)peptides and their affinity for recombinant WW1–WW2 segments of YAP and Smurf1. Colored circles denote phosphorylation of the residues. (G) Effect of alanine mutations in the PY box and the indicated phosphorylation sites on the ability of Flag-tagged Smad1 constructs to bind HA-Smurf1(DD) in HEK293 cells. (H) Schematic summary of the Smad action turnover switch operated by CDK8/9 and GSK3 in combination with YAP and Smurf1.

pS214 (Fig. 1E,F). Moreover, mutation of S210 or S214 to alanine inhibited the Smad1–Smurf1 interaction in HEK293 cells, and mutation of both residues further decreased binding (Fig. 1G). Collectively, these results suggest that CDK8/9-mediated phosphorylation of the Smad1 linker creates binding sites for competing YAP and Smurf1 WW1–WW2 domains, and GSK3 switches this balance in favor of Smurf1 binding and at the expense of the YAP interaction (Fig. 1H).

Structure of the Smurf1 WW pair bound to the Smad1 linker

We used NMR spectroscopy to calculate the structure of the Smurf1 WW1–WW2 segment bound to the Smad1

linker peptide (208–233) diphosphorylated at S210 and S214 in solution. Triple-resonance NMR spectroscopy was applied to assign the WW1–WW2 pair in this complex, whereas homonuclear and half-filter spectra were used to assign the Smad1 peptide and its contacts with Smurf1 (Fig. 2A; Supplemental Fig. 2A,B). In the complex, each of the WW domains adopts the typical triple-stranded anti-parallel β -sheet fold, even in the case of WW1 that lacks the first highly conserved tryptophan. The WW domains do not contact each other, but each contacts a portion of the linker. This arrangement provides enough freedom for the WW domains to adopt an anti-parallel orientation, forming a continuous binding surface that smoothly cradles the phosphorylated Smad1 linker. The Smad1 linker adopts an extended conformation, with

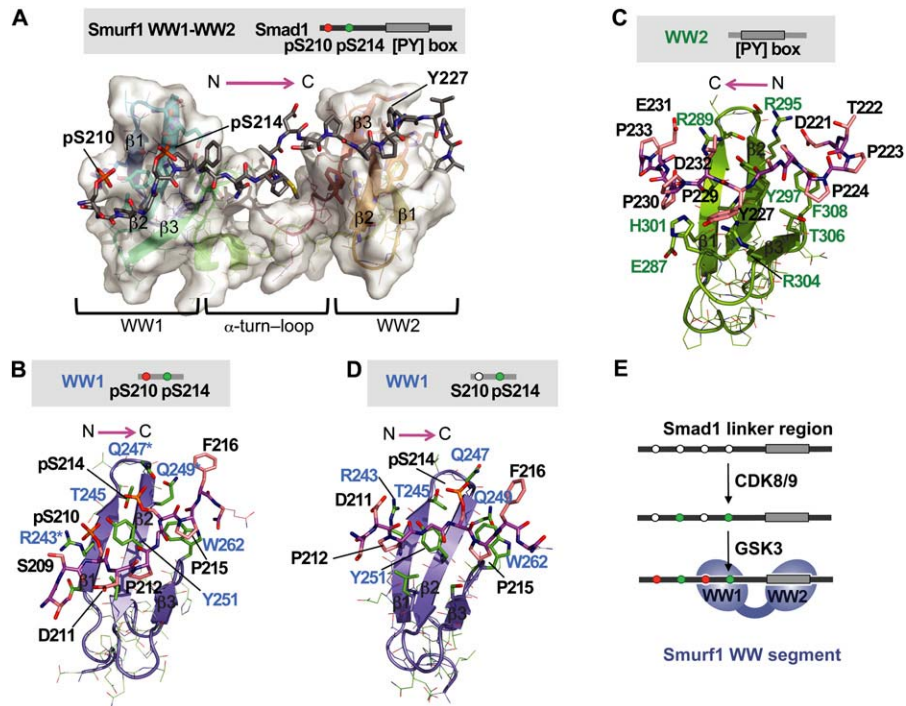


Figure 2. Structure of the Smurf1 WW1-WW2 segment bound to the Smad1 linker. (A) NMR model of the complex between the human Smurf1 WW1-WW2 pair (residues 232–314) and the 208–233 segment of the Smad1 linker diphosphorylated at S210 and S214. Smurf1 is shown as a semitransparent surface, with all elements of secondary structure represented. The Smad1 peptide is shown with a stick representation, with the backbone colored in gray. There are several relative orientations of the WW domains that satisfy all experimental NMR restraints (shown in Supplemental Fig. 2), and, due to this, we call this complex the NMR model. (B) Detailed view of the refined structure of the Smurf1 WW1 domain (slate) bound to the diphosphorylated pS210/pS214 region of the Smad1 linker. Key residues in Smad1 (black) and Smurf1 (blue) are indicated. (Asterisks) Three residues that, when jointly mutated to alanine, decreased the binding affinity of the complex by ~25-fold. (C) Detailed view of the refined structure of the Smurf1 WW2 domain (green) bound to the PY motif of Smad1. Key residues in Smad1 (black) and Smurf1 (green) are indicated. (D) Detailed view of the refined structure of the Smurf1 WW1 domain (slate) bound to the monophosphorylated pS214 region of the Smad1 linker. (E) Schematic representation of the mode of binding of Smurf1 to the Smad1 linker region.

the diphosphorylated T208–P215 segment bound to the WW1 domain and the PY motif bound to the WW2 domain (Fig. 2A). The segment between P215 and the PY motif forms a turn defined by interactions between the F217 and A220 backbone atoms. The 10-residue segment connecting the two Smurf1 WW domains adopts a helical structure in its first half (Fig. 2A). This configuration allows access to CKIP1 (casein kinase 2-interacting protein-1), a protein that binds to this region to enhance the Smurf1–Smad1 interaction (Lu et al. 2008).

To facilitate the presentation in the text, we use the one-letter amino acid notation for Smad residues and the three-letter notation for residues in its binding partners. The Smurf1 WW1 domain binds the Smad1 pS210 residue through contacts with Tyr251, Arg243, and Leu253 side chains. The Tyr251 hydroxyl and the Arg243 guanidinium groups jointly coordinate the phosphate group of pS210. pS214 also contacts Tyr251, and the phosphate group is coordinated by the hydroxyl of Thr245 and the side chains of Gln247 and Gln249. P215 is packed parallel to the aromatic ring of Trp262, and P212 is sandwiched between the Leu253 and Ser260 side chains in a cavity perpendicular to the β sheet (Fig. 2B). The pS210–D211–

P212–G213 segment forms a turn favored by a D211–P212 *cis* bond, whereas pS214–P215 is in *trans*. Single-alanine mutations of Arg243, Gln247, or Gln249 decreased the affinity to K_D values of ~30 μ M, confirming the importance of these residues in the interaction of WW1 with the pS210 and pS214 phosphate groups.

The Smurf1 WW2 domain binds to the PY motif in a manner similar to canonical group 1 WW complexes (Macias et al. 2002). P224 and P225 contact Tyr297 and Phe308, respectively, and Y227 binds between His301 and Arg304 (Fig. 2C; Supplemental Fig. 2B). The six residues after the tyrosine in the PY motif fold over the first strand of WW2. Abundant contacts are observed between P229 and P230 and His301 and Glu287, respectively. The side chain of E231 points toward the Tyr297 hydroxyl and shows contacts with the Arg289 side chain.

We also solved the structure of the Smurf1 WW1-WW2 segment bound to the Smad1 linker monophosphorylated at pS214 (Supplemental Fig. 2B). Most of the contacts between the two molecules are like those in Smurf1 WW1-WW2 bound to the pS210/pS214 diphosphorylated peptide, except that both the D211–P212 and pS214–P215

bonds are in *trans* and are bound differently (Fig 2, cf. D and B). S210 is less ordered than in the phosphorylated state, and only weak contacts are observed between D211 and Arg243 side chains (Fig. 2D). These weaker contacts explain the intermediate affinity of the WW1–WW2 domain for the Smad1 linker monophosphorylated at S214. In conclusion, formation of the Smurf1–Smad1 complex involves recognition of the Smad1 PY motif by the Smurf1 WW2 domain, and of the Smad1 GSK3 and CDK phosphorylated sites by the WW1 domain (Fig. 2E).

Structure of the YAP WW pair bound to the Smad1 linker

Alanine mutations in either the WW1 domain or WW2 domain in YAP almost completely abolished the interaction of overexpressed YAP and Smad1 in transiently transfected human cells (Supplemental Fig. 3A), suggesting that both domains are essential for this interaction. Given the high affinity of the YAP WW1–WW2 module for a Smad1 (199–233) linker peptide phosphorylated at the CDK8/9 sites pS206 and pS214 (see Fig. 1F), we solved the structure of this complex first. We used double- and triple-labeled WW1–WW2 samples (YAP 163–266 seg-

ment) to assign the protein resonances in combination with filtered and homonuclear experiments to obtain the chemical shifts of the bound peptide and the contacts between both molecules. We also used independent domains to assist in the assignment. The 25-residue connector between the two WW domains adopts a helix–loop–helix structure, as determined on the basis of the detected nuclear Overhauser effects (NOEs) and carbon chemical shift analysis. Several contacts are present between each WW domain and this connector, but these contacts do not prevent the WW domains from adopting an optimal orientation for interactions with the Smad1 linker. No contacts were observed between the WW domains.

In the complex between the WW1–WW2 pair and the Smad1 34-residue peptide, both WW domains adopt the canonical fold and participate in the interaction with Smad1 (Fig. 3A; Supplemental Fig. 3B). The WW1 domain contacts the pS206 region and the WW2 domain contacts the PY motif. The Smad1 pS206 and P207 side chains are accommodated in the aromatic cavity formed by Tyr188 and Trp199 in the YAP WW1 domain (Fig. 3B). The pS206 phosphate group is at a hydrogen bond distance from the hydroxyl groups of Thr182 and Tyr188 and the Gln186

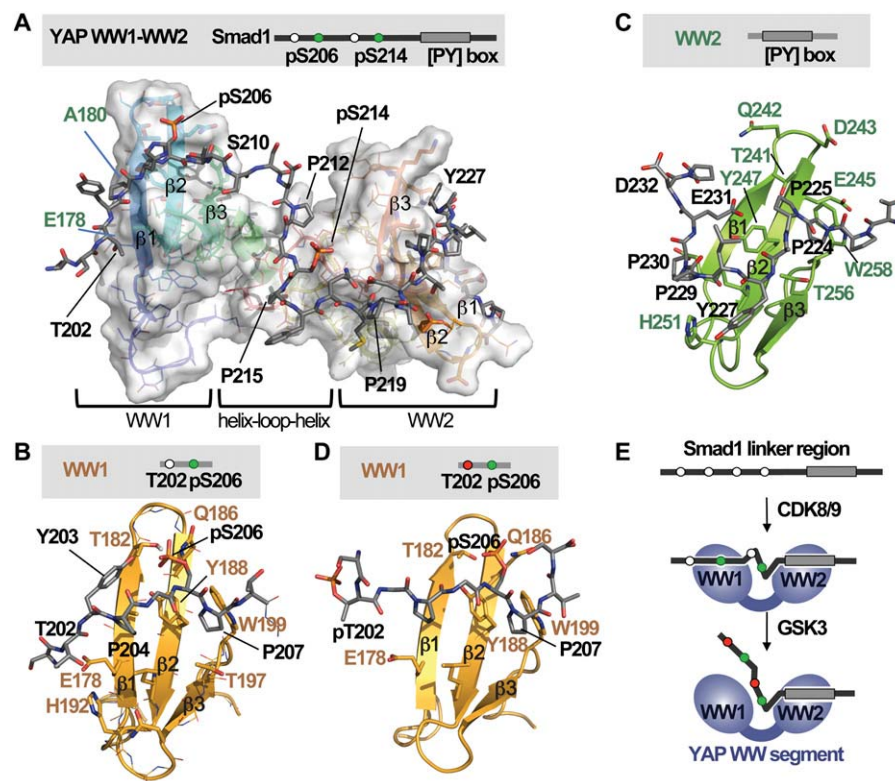


Figure 3. Structure of the YAP WW1–WW2 pair bound to the Smad1 linker. (A) NMR model of the complex between the human YAP WW1–WW2 pair (residues 163–266) and the 199–233 segment of the Smad1 linker diphosphorylated at S206 and S214. YAP is shown as a semitransparent surface, and Smad1 is shown as gray sticks. (B) Detailed view of the refined structure of the YAP WW1 domain (gold) bound to the mono-pS206 phosphorylation site of Smad1 (gray). Key residues in Smad1 (black) and YAP (brown) are indicated. (C) Detailed view of the refined structure of the YAP WW2 domain (green) bound to the PY motif region of Smad1 (gray), with the key residues indicated. (D) Detailed view of the refined structure of the YAP WW1 domain (gold) bound to the diphenylated pT202, pS206 region of the Smad1 linker. (E) Schematic representation of the mode of binding of YAP to the Smad1 linker region.

side chain. Trp199 is also involved in a network of contacts with residues comprised between P207 and S209. The structure of the WW2 domain bound to the Smad1 PY motif resembles that of Smurf1 WW2 bound to this region. The interaction is well defined, involving eight Smad1 residues between D221 and D232 and nine out of 13 residues on the WW2 domain surface (Fig. 3C). We also observed NOEs from the E231 side chain with T241 and Tyr247 and from D232 to Gln242, suggesting the presence of intermolecular salt bridges between these residues. Thus, formation of the YAP–Smad1 complex involves recognition of the Smad1 PY and the CDK phosphorylated site pS206 by the WW2 and WW1 domains, respectively.

The GSK3 phosphorylated Smad1 linker avoids YAP binding

Next we analyzed the interactions between the YAP WW1–WW2 pair and Smad1 linker peptides containing GSK3 site phosphorylations. NMR-based titrations with a peptide containing pT202, pS206, pS210, and pS214 require a fourfold to fivefold peptide excess to induce chemical shift changes in the YAP WW1–WW2 pair (Supplemental Fig. 5F), corroborating the weak interaction measured by ITC and precluding the determination of the complex structure. In a complex of YAP WW1–WW2 domains with a peptide containing pT202, pS206,

pS214, and the PY site, but no phosphorylation at S210, the NOEs detected from the N-terminal end of the peptide to the WW1 domain of YAP were weak and the structure is defined only for the P204–pS206–P207 site (Fig. 3D; Supplemental Fig. 3B). In particular, the Glu178 side chain and the His192 ring in the WW1 domain fail to contact the methyl of T202. Y203 and P204 are only partially ordered. Thus, the presence of a phosphate group in T202 destabilizes the interaction of Smad1 with the YAP WW1 domain, and the presence of phosphates at both T202 and at S210 drastically reduces the interaction between YAP and the Smad1 linker.

To better illuminate the different binding preferences of the Smurf1 and YAP complexes, we compared the charge distribution on the surfaces of the Smurf1 and YAP WW1 domains. Both WW1 domains contain Gln residues in the surroundings of pS214 and pS206, respectively, but the positively charged patch of Smurf1 WW1 that interacts with pS210 (Fig. 4A,B) is absent in YAP. Instead, YAP contains a negatively charged region suited to interact with T202 but incompatible with the presence of a phosphate group at T202 (Fig. 4C,D).

Notably, in the complex of YAP WW1–WW2, the Smad1 linker segment between residues S210 and D221 runs across the inter-WW connector with three prolines (P212, P215, and P219) in *trans* (see Fig. 3A). The *trans* configuration of the D211P and pS214P bonds favors the formation of two β turns that facilitate the interaction of

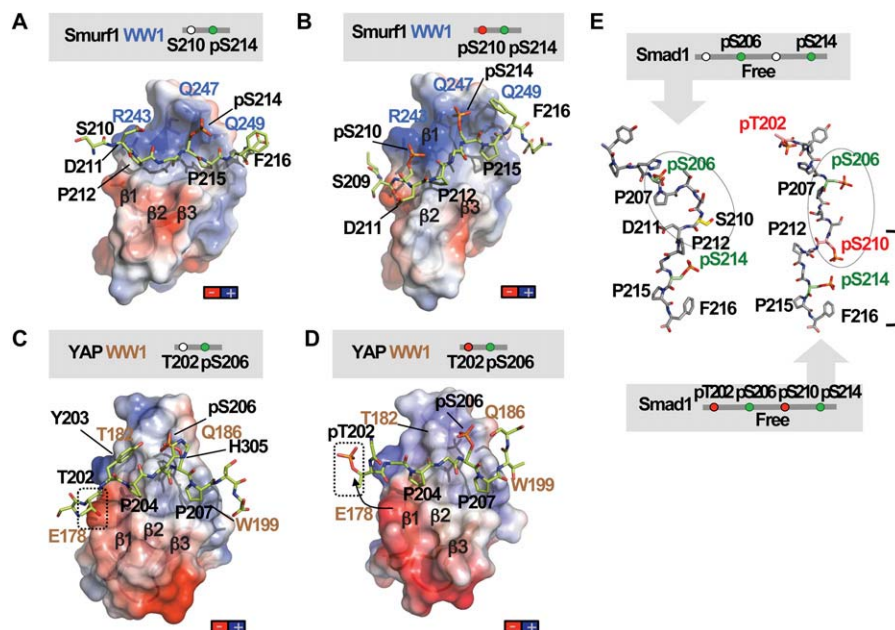


Figure 4. The GSK3 phosphorylated Smad1 linker prevents YAP binding. (A,B) Charge distribution on the surface of the Smurf1 WW1 domain in complex with the Smad1 linker monophosphorylated at S214 (A) or diphosphorylated at S210 and S214 (B). Negatively charged patches are shown in red, and positively charged patches are shown in dark blue. Smurf1 WW1 is shown as a semitransparent surface, and Smad1 is shown as green sticks. Key residues in Smad1 (black) and Smurf1 WW1 (blue) are shown. The complex is shown in the same orientation as that of Figure 2. (C,D) Charge distribution on the surface of the YAP WW1 domain in complex with the Smad1 linker monophosphorylated at S206 (C) or diphosphorylated at T202 and S206 (D). The YAP WW1 domain is shown as a semitransparent surface and with the same orientation as in Figure 3. The position of T202 is shown in a box. The conformational change observed in pT202 is represented with an arrow. (E) Molecular simulations performed on two peptides corresponding to Smad1 phosphorylated at S206 and S214 (left) or at T202, S206, S210, and S214 (right). Key residues are labeled.

the Smad1 pS206 site with YAP. This feature likely explains the higher affinity of YAP for the pS206/pS214 diphosphorylated peptide ($8.4 \pm 1 \mu\text{M}$) compared with its affinity for the pS206 monophosphorylated peptide ($17.0 \pm 2 \mu\text{M}$). The negative effect of pS210 on the YAP–Smad1 interaction observed by ITC and NMR titration experiments could arise from a conformational change in the Smad1 fragment forced by electrostatic repulsion between the pS210 phosphate group and the negatively charged D211. To test the potential impact of this phosphate, we performed NMR experiments and molecular dynamic simulations of peptides containing pS214, pS210, and pS206. The simulations revealed that pS210 favors an extended conformation without the β turn centered at D211–P212 (Fig. 4E) that would decrease the likelihood of WW1 interacting with the pS206 site.

Collectively these observations suggest that phosphorylation of Smad1 by CDK8/9 creates a binding site for the YAP WW1 domain in pS206, and the downstream all-trans configuration imposed by D211P and pS214P favors

this binding interaction. GSK3 phosphorylation of the Smad1 linker at T202 and, particularly, at S210 creates a conformation that avoids recognition by the YAP WW1 domain (Fig. 3E) while favoring recognition by the Smurf1 WW1 domain.

A Smad action turnover switch in the TGF- β pathway

Several clues suggested that a similar Smad action turnover switch, with its own phospho-amino acid code and set of WW domain code readers, may operate in the TGF- β /nodal version of the pathway. The linker regions of Smad2 and Smad3 contain a conserved PY motif, three agonist-dependent CDK phosphorylation sites, and one CDK-primed GSK3 site (Fig. 5A; Alarcon et al. 2009; Wang et al. 2009). Significant differences exist in the configuration of these elements compared with Smad1. In the Smad3 linker, a T179 immediately preceding the PY motif is rapidly phosphorylated by CDK8/9 in response to TGF- β . This is followed by phosphorylation of two

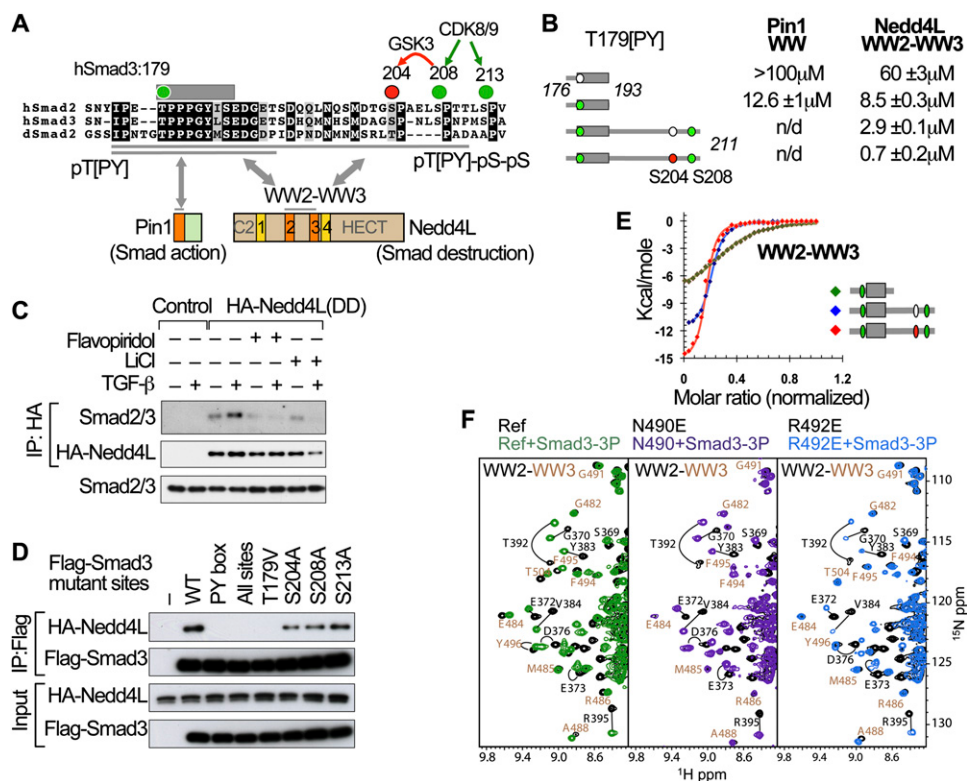


Figure 5. Elements of a Smad action turnover switch in the TGF- β pathway. (A) Sequence alignment of the linker regions of human Smad2 and Smad3 and *Drosophila* Smad2, with the conserved residues highlighted. The conserved CDK8/9 sites (green) and CDK8/9-primed GSK3 site (red) and the PY box (blue) are shown. Two Smad3 segments (176–193 and 176–211) used in this study are underlined. The domain composition of Pin1 and Nedd4L proteins and the regions that mediate binding to linker phosphorylated Smad3 are indicated. (B) Synthetic Smad3 (phospho-)peptides and their affinity for the recombinant Pin1 WW domain and Nedd4L WW2–WW3 pair. Colored circles denote phosphorylation of the indicated residues. (n.d.) Not determined. (C) TGF- β -dependent formation of a complex between HA-Nedd4L(DD) and endogenous Smad3, and effects of flavopiridol and LiCl on the formation of this complex. (D) Effect of alanine mutations in the PY box and the indicated phosphorylation sites on the ability of HA-Nedd4L(DD) to bind Flag-tagged Smad3 constructs to bind HA-Nedd4L(DD) in HEK293 cells. (E) ITC curves and corresponding fitting to pairs of Nedd4L WW domains and the indicated Smad3 (phospho-)peptides. (F) NMR titrations of WW2–WW3 pairs (wild type in green) with point mutations introduced in two residues that coordinate the pS204pS208 site (violet and royal blue). Residues that belong to WW2 and WW3 are labeled in black and camel, respectively.

CDK8/9 sites (S208 and S213) and a single GSK3 site (S204), all of which are located downstream from the PY motif in this case. The pT179[PY] motif is recognized by the single WW domain of Pin1 (Fig. 5A; Matsuura et al. 2009). Using ITC titrations, we corroborated that pT179 within a Smad3 176–193 peptide is the preferred binding site for the Pin1 WW domain ($K_D = 12.6 \pm 1 \mu\text{M}$) (Fig. 5B), as compared with pS204/pS208 within a Smad3 202–211 peptide ($K_D = 23.4 \pm 7 \mu\text{M}$) or pS208/pS213 within a Smad3 206–215 peptide ($K_D = 49 \pm 11 \mu\text{M}$).

The pT179[PY] motif is also recognized by the Nedd4L WW2 domain (Gao et al. 2009). A recombinant WW2 domain bound this motif with $K_D = 8.5 \pm 0.3 \mu\text{M}$ (Fig. 5B). However, formation of the TGF- β -dependent Nedd4L–Smad3 complex was inhibited not only by the CDK inhibitor flavopiridol, but also by the GSK3 inhibitor LiCl (Fig. 5C). Furthermore, alanine mutation of the GSK3 site S204 diminished the binding of Smad3 to Nedd4L in HEK293 cells, as did mutation of the GSK3-priming CDK site S208 (Fig. 5D). Mutation of S213 had no discernible effect on binding. These results suggest that the pS208-primed, GSK3-mediated phosphorylation of S204 further augments the affinity of Nedd4L for Smad3. Thus, we postulated that another WW domain in Nedd4L must recognize the pS204–pS208 site and collaborate with the binding of WW2 to the pT179[PY] motif.

To test this hypothesis, we determined whether pairs of Nedd4L WW domains could bind to an extended region of Smad3, including the pT179[PY] motif as well as the phosphorylated S208 site or the phosphorylated S204 and S208 sites (Fig. 5B). Of the three consecutive pairs present in Nedd4L, only the WW2–WW3 pair showed high affinity for the diphosphorylated peptide ($K_D = 2.9 \pm 0.1 \mu\text{M}$) and a further gain in affinity for the triphosphorylated peptide ($K_D = 0.7 \pm 0.2 \mu\text{M}$) (Fig. 5B,E). This affinity is 15-fold stronger than the affinity ($K_D = 10.1 \pm 0.7 \mu\text{M}$) of the same peptide for the WW2–WW3 region of another family member, Smurf2. NMR titrations with increasing concentrations of the pT[PY]-pS-pS peptide corroborated chemical shift changes in both domains of a recombinant WW2–WW3 pair. The residues affected in WW2 were the same as those affected in titrations with the pT179[PY] peptide, whereas residues affected in WW3 were located in and around the first loop and in the third strand (Fig. 5F). The effects of point mutations in WW2 (Leu384 to Tyr) or in WW3 (Ile496 to Tyr) provided further evidence that WW2 recognizes the pT179[PY] motif and WW3 recognizes the pS204–pS208 sites (Supplemental Fig. 4A), as summarized in Figure 5A.

Basis for Nedd4L and Pin1 recognition of the Smad3 linker phosphorylation code

Determination of the solution structure of the Nedd4L WW2–WW3 module (Nedd4L 364–512 segment) bound to the Smad3 linker peptide pT[PY]-pS-pS (Smad3 176–211 segment triphosphorylated at T179, S204, and S208) (Fig. 6A) posed a challenge owing to poor NMR signal dispersion of the 80-amino-acid region connecting WW2 and WW3. To simplify the interpretation of the NMR data, we

prepared the WW2–WW3 module as two separate fragments, which allowed the use of sequential isotope labeling. A disulfide bond protein ligation strategy was used to connect these two fragments (Fig. 6B; Baca et al. 1995; Nair and Burley 2003). Applying a stepwise protocol to solve the structure of the complex (see the Materials and Methods), we were able to fully assign the WW2 and WW3 domains and ~80% of the residues in the linker (Fig. 6A). The unassigned residues are located in a proline-rich segment immediately upstream of the WW3 domain. The α and β ^{13}C chemical shifts analysis revealed that the segment connecting the WW2–WW3 pair (residues 400–480) lacks elements of a secondary structure in both the free and bound conformations, and is unaffected by binding of the Smad3 peptide.

The Nedd4L WW2 domain recognizes the EpTPPPGY segment in this complex as well as in a complex with a shorter Smad3 peptide (pT[PY] peptide, Smad3 176–193 segment) that we also solved (Fig. 6C). P181 and P182 bind in a cavity formed by the Tyr382 and Trp393 aromatic rings, whereas Y184 binds to His386, Arg389, and backbone atoms in loop 2. The Lys378 and Arg380 side chains coordinate the phosphate group of pT179, while the Trp393 aromatic ring contacts the methyl and β protons. The electrostatic interactions of Lys378 and Arg380 with the phosphate group explain the high affinity of this interaction and compensate for an absence of contacts with residues downstream from the PY motif observed in YAP and Smurf1 complexes. Versions of the Nedd4L WW2 domain with Glu mutations in Lys378 or Arg380 decreased the affinity for the pT[PY] peptide from $K_D = 4.1 \pm 0.3 \mu\text{M}$ to $K_D = 20\text{--}23 \mu\text{M}$, and mutation of both residues caused a further decrease ($K_D = 44.6 \pm 15 \mu\text{M}$), confirming the importance of these residues in binding (Fig. 6C). We solved the structures of the mutant proteins to verify that the mutations do not affect the overall fold of the WW domain, but that they alter the charge distribution of loop 1, involved in the phosphate recognition of the pT179 site (Supplemental Fig. 5C–F). Negatively charged residues in and around loop1 are highly unusual in WW sequences (SMART database of protein domains). Interestingly, the wild-type sequence of the YAP and TAZ WW2 domains are unique for having these negatively charged residues. The weak affinity of the YAP WW1 domain for PY motifs (Macias et al. 1996; Pires et al. 2001; Toepert et al. 2001) and the presence of a negatively charged patch in the YAP WW2 domain that prevents binding to a pT[PY] motif explain our previous result that YAP does not interact efficiently with linker phosphorylated Smad3 (Supplemental Fig. 5F–H; Alarcon et al. 2009).

The Nedd4L WW3 domain binds pS208 between the Asn490 side chain and the guanidinium group of Arg492, and also contacts the aromatic ring of Phe494 (Fig. 6D). Additional contacts of P209 and P211 to the aromatic ring of Trp505 help position pS208. pS204 is bound by the Arg486 guanidinium group and the hydrophobic part of Glu484, and the ring of P205 is accommodated between the side chains of Glu484, Arg486, and Tyr496. Indeed, alanine mutation of either Arg486 or Arg492

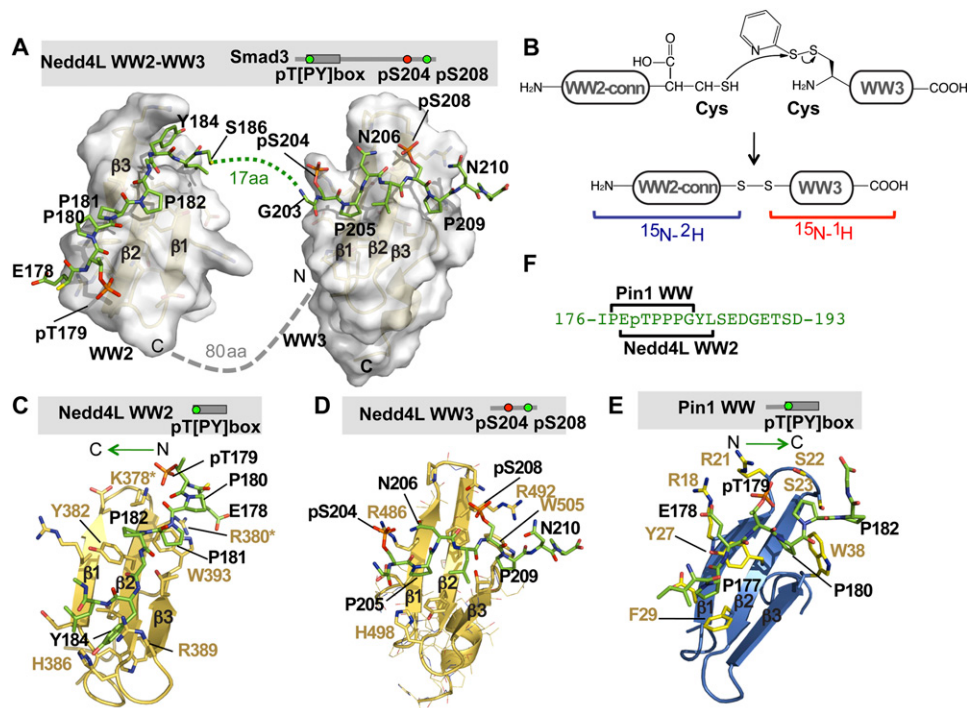


Figure 6. Basis for Nedd4L and Pin1 recognition of the phosphorylated Smad3 linker. (A) Model of the complex structure between the human Nedd4L WW2–WW3 pair (residues 364–512) and the 176–211 segment of the Smad3 linker triphosphorylated at T179, S204, and S208. Nedd4L is shown as a semitransparent surface, and Smad3 is shown as green sticks. Smad3 residues involved in the interaction with the Nedd4L WW2–WW3 pair are indicated. The 80-amino-acid region connecting the WW2 and WW3 domains (dotted line) does not adopt a defined secondary or tertiary structure, as indicated by near-random ^{13}C chemical shifts. Due to the complexity that this long unstructured part adds to the calculation of the complex, the model has been calculated using three independent molecules (WW2, WW3, and the Smad3 peptide) without the 80-amino-acid region. Three possible orientations of the WW2 and WW3 pair were obtained using a set of RDC experiments. The orientation that yields the best view of the bound Smad3 peptide is shown here. In the bound Smad3, the segment between S186 and G203 does not contact either WW2 or WW3 and is not represented. (B) Schematic of the protein ligation strategy employed to prepare the WW2–WW3 module as two separate fragments for sequential isotope labeling. A fully deuterated WW2 and connector (WW2-conn) segment of Nedd4L was ligated to the protonated WW3 domain, as shown. Using this strategy, the signals from the first part of the protein were filtered, and the analysis of data corresponding to the WW3 domain bound to Smad3 and of that of Smad3 itself was simplified. In the calculation of the complex, this information was combined with that of the WW2 in complex with the pT[PY] peptide and with residual dipolar coupling data obtained from the entire WW2–WW3 segment. (C) Detailed view of the Nedd4L WW2 domain (gold) bound to the phosphorylated PY motif (pT179[PY]) of Smad3 (green). Key residues in Smad3 (black) and Nedd4L (brown) are indicated. This complex has been refined using the data of the WW2–WW3 pair that corresponds to the second WW and the fragment of 176–190 of Smad3. (Asterisks) Two residues that, when jointly mutated to alanine, decreased the binding affinity of the complex by ~ 11 -fold. (D) Detailed view of the Nedd4L WW3 domain (gold) bound to the diphosphorylated pS204–pS208 sites of Smad3 (green). Key residues in Smad3 (black) and Nedd4L (brown) are indicated. This complex has been refined using the data of the WW2–WW3 that corresponds to the WW3 site and the fragment of 203–211 of Smad3. (E) Solution structure of the Pin1 WW domain bound to the Smad3 pT179[PY] motif. The WW domain is shown as a ribbon representation, shown in marine. Key residues in Smad3 (black) and Pin1 (light brown) are indicated. This complex is displayed using the same orientation as that of the Nedd4L WW2 complex (shown in C) to highlight that these WW domains bind to the pT179[PY] site in opposite orientations. (F) Representation of the distinct portions of the pT179[PY] motif of Smad3 that provide contacts with Pin1 and Nedd4L.

caused a fourfold reduction in the affinity of the WW2–WW3 construct for the pT[PY]-pSpS peptide. Overall, the GSK3-CDK8/9 diphosphorylated motifs of Smad3 and Smad1 are recognized by the corresponding WW domains of Nedd4L and Smurf1 in a structurally similar manner, with the phospho-serines positioned close to arginine or glutamine side chains.

The structure of the Pin1 WW domain bound to a Smad3 pT[PY] peptide (Fig. 6E) revealed similarities but also significant differences compared with how Nedd4L binds to this motif (Fig. 6F). The Pin1 WW

domain recognizes the PEpTPPP motif of Smad3, with P177 bound between the Tyr27 and Phe29 aromatic rings, and P180, P181, and P182 all in *trans* and contacting Trp38. The proline aromatic stacking used by Pin1 provides a different binding strategy compared with the network of van der Waals interactions used by the Nedd4L WW2 domain to bind these prolines (Fig. 6, cf. C and E). The pT179 phosphate is coordinated by arginine side chains (Arg18 and Arg21) in β strand 1 and loop 1 of the Pin1 WW domain, which also differs from the coordination of this phosphate group by β strands 2 and 3 in

Aragón et al.

the case of Nedd4L WW2. These different modes of recognition of the Smad3 pT179-PY motif by Pin1 and Nedd4L may reflect the different outcomes of these two interactions: peak transcriptional action in the case of Pin1, and Smad3 polyubiquitination in the case of Nedd4L.

Discussion

We show here that tandem WW domains in Smad-binding proteins function as readers of a phospho-serine code that dictates Smad peak transcriptional action as well as the subsequent elimination of Smad molecules that participate in transcription. The code of this action turnover switch is written by kinases CDK8/9 and GSK3 acting on the linker region of activated, transcriptionally poised Smad proteins. In our model (Fig. 7), CDK8/9 create binding sites that are preferentially recognized by WW-containing Smad transcriptional cofactors. These phosphorylations additionally prime Smads for subsequent GSK3-mediated phosphorylation, which creates sites for ubiquitin ligase binding at the expense of transcriptional cofactor-binding sites. Thus, GSK3 switches the phosphorylation code in the Smad linker region from one that favors Smad action to one that favors Smad destruction. As a result, TGF- β /BMP signal delivery becomes coupled to Smad turnover.

We propose that degradation is a price that Smad molecules pay for participating in transcription. CDK8 and CDK9 are components of the CDK8/CyclinC/Med12/Med13 transcriptional mediator complex and P-TEFb Cdk9/CyclinT elongation complex that regulate RNA polymerase II during transcription (Komarnitsky et al. 2000; Malik and Roeder 2000; Durand et al. 2005). CDK8/9 have access to Smad molecules on the chromatin but not to receptor-activated Smad molecules that fail to engage in transcriptional complexes (Alarcon et al. 2009). Another fate for activated Smads is C-terminal dephosphorylation, which mediates Smad recycling for new rounds of signaling and thereby links the duration of signaling to the presence of activated TGF- β or BMP receptors (Batut et al. 2008; Schmierer et al. 2008). However, Smad C-terminal dephosphorylation does not require prior participation of the molecule in transcription. Smad C-terminal dephosphorylation may modulate the global pool of activated Smads, whereas the action-coupled turnover process elucidated here eliminates Smad molecules as a function of their exposure to CDK8/9 during target gene regulation.

Our study provides a structural and functional basis for the involvement of tandem WW domains in these protein-protein interactions. On the one hand, the different WW domains of Smurf1, Nedd4L, and YAP proteins achieve overall high specificity in target recognition by acting in pairs, extending the interacting surface to recognize not only the canonical PY site, but also the adjacent pS/pTP motifs. The interactions with the pS sites were unexpected from predictions based on sequence conservation, since only the WW domain of Pin1 has been described as a pS/pTP-binding motif, and the Pin1 residues involved in the phosphate interaction

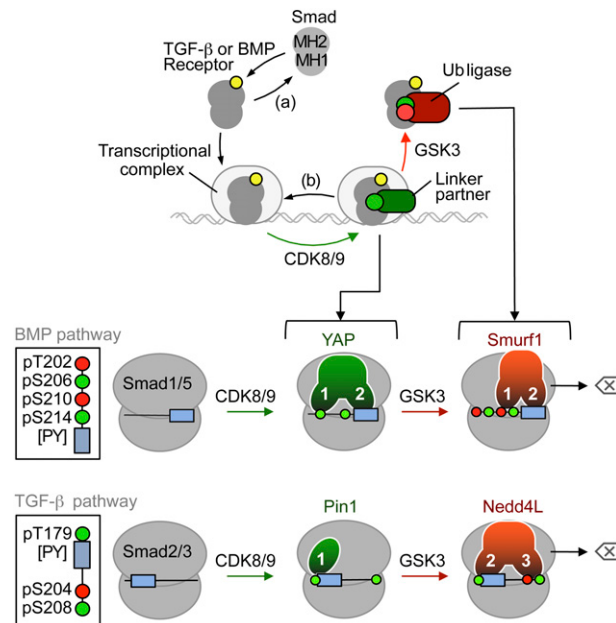


Figure 7. The Smad action turnover switch in the BMP and TGF- β pathways: pSer codes and WW domain code readers. (*Top panel*) Schematic summary of the Smad action turnover switch in the BMP and TGF- β pathways. Following receptor-mediated phosphorylation (yellow circle), Smad proteins translocate to the nucleus and assemble transcriptional complexes, which are phosphorylated at CDK8/9 sites (green circle) in the MH1-MH2 interdomain linker region. This phosphorylation creates high-affinity binding sites for transcriptional partners (such as YAP in the case of the BMP mediator Smad1, Pin1 in the case of the TGF- β mediator Smad3, and probably others), thus achieving peak transcriptional action. Phosphorylation by CDK8/9 also primes the Smads for GSK3-mediated phosphorylation (red symbol) at the -4 position, which favors the binding of ubiquitin ligases Smurf1 (BMP pathway) and Nedd4L (TGF- β pathway), leading to proteasome-dependent degradation of Smad molecules that participate in transcription (erase symbol). Alternatively, C-terminal Smad phosphatases (a) and linker phosphatases (b) reverse these phosphorylation states. See the text for details and citations. (*Bottom panels*) Schematic of the Smad linker phospho-amino acid codes (*insets*) and WW domain code readers. The conserved CDK8/9 phosphorylation sites (green circles) and GSK3 sites (red circles) are located at the indicated positions relative to the PY box (slate box). Amino acid positions correspond to Smad1 and Smad3. In the BMP pathway, the YAP WW1 domain binds to pS206 in Smad1, as long as p210 is not phosphorylated. The Smurf1 WW1 domain binds with higher affinity to the pS210-pS214 motif. The WW2 domains bind the [PY] motif. In the TGF- β pathway, the sole WW domain of Pin1 binds the pT179[PY] motif, as does the WW2 domain of Nedd4L. However, the Nedd4L WW3 domain increases the binding affinity by recognizing the pS204-pS208 motif. See the text for additional details and citations on the known roles of these WW domain proteins in Smad signal transduction.

are not strictly conserved in YAP, Smurf1, and Nedd4L WW domains. On the other hand, since the linker region of Smads contains a set of CDK phosphorylation sites and a set of CDK-primed GSK3 sites, the phospho-Ser motifs

can be tuned to optimally bind either transcriptional cofactors (monophosphorylated) or ubiquitin ligases (diphosphorylated). The WW2 domains of Smurf1 and YAP recognize a canonical PY motif in Smad1, whereas the WW2 domain of Nedd4L and the Pin1 WW domain recognize a phospho-threonine-PY motif in Smad3. The PY-independent, monophospho-serine motif of Smad1 is recognized by the WW1 domain of YAP, whereas the atypical WW1 domain of Smurf1 and the Nedd4L WW3 domain recognize diphospho-serine motifs in Smad1 and Smad3, respectively. The ability of WW domains to recognize both pT[PY] and pS(-4)pS motifs was previously unknown. The interaction of these proteins with Smads is determined not only by the binding specificity of individual WW domains, but also by the configuration of the two domains. For example, Smurf1 cannot bind Smad3 because the N-terminal position of the pT[PY] motif relative to the pSer cluster in Smad3 is opposite to the orientation required by the WW1 and WW2 domains of Smurf1. The present findings expand the known structural and functional versatility of WW domains as protein-protein interaction modules (Macias et al. 2002).

The recognition of distinct Smad phosphorylation codes by different WW domain proteins provides ample opportunities for regulation. As an adaptor that binds to the WW1–WW2 connector of Smurf1 (Lu et al. 2008), CKIP1 may enforce an optimal orientation of these WW domains for contact with their cognate sites on the Smad1 linker. A Smurf1 isoform with a longer WW1–WW2 connector exists that may differ in this regulatory function (Schultz et al. 2000). The ability of serum/glucocorticoid-regulated kinase 1 (SGK1) to phosphorylate the WW2–WW3 connector and inhibit Nedd4L binding to Smad3 (Gao et al. 2009) may similarly be based on a change in the binding ability of these WW domains. In response to mitogens and stresses, MAPKs primarily phosphorylate Smad3 at S208/S213, not at T179 (Gao et al. 2009). The poor ability of Nedd4L to recognize such MAPK phosphorylated Smad3 species (Gao et al. 2009) can be explained by the key role of the pT179 phosphate group in contacting the Nedd4L WW2 domain. Furthermore, the signals that inhibit GSK3 and increase the half-life of Smad proteins (Fuentelba et al. 2007; Guo et al. 2008; Wang et al. 2009) can now be regarded as inputs that extend the period of CDK8/9-dependent Smad peak performance. In mammalian cells and *Drosophila*, YAP enhances certain BMP responses but not others (Alarcon et al. 2009). Similarly, Pin1 enhances certain effects of TGF- β on mammalian cell migration but not other effects (Matsuura et al. 2009). Different factors may fulfill these roles in other contexts or on other target genes. A larger repertoire of Smad linker code-reading factors than presently known may therefore exist.

The action turnover switch delineated here involves a remarkable concentration of opposing protein-binding functions in a discrete region of the Smad proteins. The fulfillment of dual roles under a phosphorylation-dependent switch is also characteristic of another key component of this pathway: the regulatory GS region of type I TGF- β receptor kinase (Huse et al. 2001). The core TGF- β /Smad

pathway is therefore characterized by the economical use that it makes of the structural elements that switch key pathway components from one activation state into another.

Materials and methods

Mammalian cell expression vectors

The plasmids encoding Smad1, Smad3, Smurf1(DD), and Nedd4L(DD) are described elsewhere (Sapkota et al. 2007; Alarcon et al. 2009; Gao et al. 2009). The linker phosphorylation site mutant Smad3 (all sites) was denoted previously as Smad3 (EPSM) (Kretzschmar et al. 1999). PCR-based site-directed mutagenesis was employed to generate all of the other mutants of the linker phosphorylation sites of Smad1 and Smad3 using the primers containing the desired mutation. The YAP construct was obtained from M.B. Yaffe (Massachusetts Institute of Technology). The YAP mutants with WW domain mutations were generated by PCR-based site-directed mutagenesis. The primers are YAPWW1AA-F (CAGACAACAACAGCGCAGGACGCCAGG AAGGCCATG), YAPWW1AA-R (CATGGCCTTCCTGGCGTCC TGCCTGTTGTTGTCTG), YAPWW2AA-F (AAGACCACCT TGCCTAGACGCAAGGCTTGACCCT), and YAPWW2AA-R (AGGGTCAAGCCTTGCGTCTAGCGCAGAGGTGGTCTT).

The human *PIN1* cDNA was from Open Biosystems, and the XhoI, EcoRI fragment was cloned into the PCI vector (Promega). Sequences of all constructs generated and used in this study were verified by DNA sequencing.

Transfection, immunoprecipitation, and immunoblotting

Transfections of the indicated plasmids were performed as described previously (Alarcon et al. 2009; Gao et al. 2009). Cells were used 36 h later. HEK293T cells were incubated as indicated with BMP2 (25 ng/mL; R&D Systems), TGF- β 1 (100 pM; R&D Systems), noggin (50 ng/mL; R&D Systems), or SB431542 (10 μ M; TOCRIS). When used, flavopiridol (0.6 mM; National Cancer Institute) and LiCl (15 mM) were added to the cells for 1 h prior to the addition of growth factors. Immunoprecipitations and Western immunoblotting were done as described previously (Sapkota et al. 2007). Antibodies raised against full-length Smad1 were from R&D Systems or were generated in-house. Antibodies against phospho-tail Smad1 were raised against Smad1 (pS463/pS465) (Cell Signaling). Antibodies against Smad2/3, Smad1pS206 were produced in-house (Sapkota et al. 2007), and Smad1 pS214 was produced in-house. Antibodies against Smad1 pS210 were a gift from E. De Robertis (Fuentelba et al. 2007). Other antibodies used include rat monoclonal anti-HA-peroxidase (Roche), mouse monoclonal anti-Flag M2-peroxidase antibody (Sigma), agarose-conjugated mouse monoclonal anti-HA clone 7 (Sigma), and anti-Flag M2 affinity gel (Sigma).

Recombinant products

The three consecutive WW domain pairs in human Nedd4L (WW1–WW2 193–400, WW2–WW3 364–512, and WW3–WW4 476–563) and the WW domains of human Smurf1 (WW1 232–267, WW2 277–314, and WW1–WW2 232–314) were cloned into a pETM11 vector. The WW domains of human YAP (WW1 163–206, WW2 227–266, and WW1–WW2 163–266) and Pin1 (1–40) were cloned into a pETM30 vector (a gift from the EMBL-Heidelberg Protein Expression Facility) using NcoI and HindIII sites. All point mutations described in the text were introduced using the QuickChange site-directed mutagenesis kit (Stratagene)

Aragón et al.

with the appropriate complementary mutagenic primers. All wild-type and mutant constructs were confirmed by sequencing.

Generation of Nedd4L WW2–WW3 segment by protein ligation

We selected a segmental labeling approach that requires the protein to be expressed as two fragments, each produced with a different labeling pattern, and then combined by chemical ligation. For the ligation, we used the formation of a disulfide bridge built from the side chains of two cysteine residues—one at the C terminus of fragment 1 (WW2-connector-C471) and the other at the N terminus of fragment 2 (C472-WW3)—using mutants I515Y and H517A as templates. In order to guide the reaction toward the formation of the hetero-disulfide WW2–WW3 product, the (C472-WW3) site was activated prior to the ligation reaction (Baca et al. 1995). The pyridylsulfenyl-cysteine WW3 product was identified as a new peak with a higher retention time in the reverse-phase high-performance liquid chromatography (RP-HPLC) profile of the reaction, and its mass was corroborated by MALDI-TOF mass spectroscopy. The product was purified by semipreparative RP-HPLC and lyophilized. ^2H , ^{15}N -labeled Nedd4L WW2-linker-C471 (0.4 μmol) and ^1H , ^{15}N -labeled Nedd4L PyS-(C472)-WW3 domain (0.3 μmol) were dissolved in 600 μL of 6 M guanidine-HCl and sodium acetate buffer (20 mM, 100 mM NaCl at pH 4.5) and stirred overnight at 4°C. The ligation product was identified by RP-HPLC/MALDI-TOF mass spectroscopy. After purification, the product was lyophilized.

Proteins

Unlabeled, ^{15}N -labeled, ^{13}C , ^{15}N -labeled, and ^2H , ^{13}C , ^{15}N -labeled proteins were expressed in *Escherichia coli* BL21 (DE3) in LB medium or minimal medium (M9) using either H_2O or D_2O (99.99%, CortecNet) enriched with $^{15}\text{NH}_4\text{Cl}$ and/or D -[^{13}C] glucose as the sole sources of carbon and nitrogen, respectively (Marley et al. 2001). *E. coli* extracts were lysed using an Emulsi-Flex-C5 (Avestin) cell disrupter equipped with an in-house-developed Peltier temperature controller system. Soluble fusion proteins were purified by nickel-affinity chromatography (HiTrap Chelating HP column, GE Healthcare Life Science), and samples were eluted using buffer A with EDTA. Smurf1 WW1–WW2 protein was mostly in the insoluble fraction after centrifugation of *E. coli* lysates, and was solubilized with 6 M guanidine-hydrochloride and then purified using the HiTrap HP column. After buffer exchange, fusion tags were removed by overnight TEV protease digestion at 4°C followed by a second nickel-affinity binding step. All proteins were further purified with an additional gel filtration chromatography step using HiLoad Superdex 30, 75, or 200 16/60 prep-grade columns (GE Healthcare), depending on the protein size. Fractions containing the purified proteins were concentrated to 1–2 mM for NMR experiments. To ensure the presence of a 1:1 protein:peptide ratio, and to avoid formation of aggregates or misfolded samples, Smurf1 and YAP proteins were concentrated in the presence of the Smad1 peptides prior to NMR experiments. The NMR buffer was 20 mM deuterated Tris-HCl (pH 7.2–7.4), 100 mM NaCl, 0.01% NaN_3 , and 2.5 mM deuterated dithiothreitol in the presence of 10% D_2O .

Peptides

All peptides were prepared using Fmoc solid-phase peptide synthesis with 0.10–0.15 mmol FastMoc protocols as described (Alarcon et al. 2009; Gao et al. 2009). The syntheses of Smad3 3P and Smad1 3P and 4P peptides were optimized by combining manual and automated strategies. Crude peptides were purified

by RP-HPLC using a Vydac C18 or C4 Sephasil preparative columns and an ÄKTApurifier10 (GE Healthcare) or in a Waters HPLC delta 600 system using a high preparative Waters SunFire C18 column. Fractions containing the desired peptides were identified by MALDI-TOF mass spectrometry.

ITC

ITC experiments were performed using a VP-ITC MicroCalorimeter (MicroCal) at 10°C and 25°C essentially as described (Alarcon et al. 2009; Gao et al. 2009). ITC isotherms were fit to the simplest model with MicroCal's ORIGIN software.

NMR spectroscopy

All experiments were recorded on a Bruker Avance III 600-MHz spectrometer equipped with a Z pulse field gradient unit and either triple (^1H , ^{13}C , ^{15}N) or quadruple (plus ^{31}P) resonance probe heads. Double- and/or triple-labeled samples were prepared to obtain sequence-specific [HNCACB/HN(CO)CACB or CBCA(CO)NH/CBCANH] experiments. Side chain resonance assignments were obtained using standard triple-resonance experiments [^{15}N -TOCSY, CC(CO)NH, HCCH-TOCSY (12-msec mixing time) and ^{15}N , ^{13}C NOESY experiments (with 100- and 150-msec mixing times depending on the protein size)]. Intramolecular proton distance restraints were obtained from peaks assigned in 2D-NOESY, ^{15}N -NOESY, ^{13}C -NOESY, and half-filtered experiments. All spectra were processed with the NMRPipe/NMRDraw (Delaglio et al. 1995) software and were analyzed with CARAMBA (Bartels et al. 1995). Spectra used for the calculation were integrated with the batch integration method of the XEASY package. $^3\text{J}_{\text{HN-H}\alpha}$ scalar couplings were obtained from HNHA experiments. One-bond N- ^1H RDCs were determined by using the IPAP ^{15}N HSQC sequence. Hydrogen bonds were obtained by acquiring a set of ^1H / ^{15}N -HSQC experiments after dissolving the lyophilized protein in D_2O . Double- and half-filtered experiments were run to assign the peptides in the bound state (Sattler et al. 1999).

NMR titration experiments with peptides

For the ^{15}N -HSQC experiments, ^{15}N -labeled protein domains were prepared at 0.25 mM concentration in the same buffer as described above, and unlabeled ligand was added to the ^{15}N -labeled sample up to a final molar ratio of 3:1. Measurements were performed at 285 K or 295 K on a Bruker Avance III 600 MHz.

Structure calculation

For the structure calculation, distance restraints derived from NOESY experiments, $^3\text{J}(\text{H}^{\text{N}}, \text{H}^{\text{A}})$ obtained from HNHA spectra, and hydrogen bond restraints determined by D_2O exchange were used. The structures were calculated using CNS (Brünger et al. 1998) with an in-house-modified protocol of Aria 1.2 (Nilges et al. 1997). Since only unambiguously assigned restraints were used, the protocol was reduced to two iterations of one and 120 structures, respectively, using 100,000 cooling steps. All calculated structures were submitted to water refinement, and were ranked based on minimum values of energy terms and violations. The water refinement protocol was also modified by weighing the value of unambiguous NOEs, hydrogen bonds, and dihedral restraints by a factor of 10. In this way, all experimental restraints are used during the refinement process, and the obtained structures are in better agreement with the experimental data while retaining good Ramachandran values. The Nedd4L WW2–WW3 complex was refined using residual

dipolar couplings obtained from aligned samples in different alkyl-poly(ethylene glycol) mixtures (Ruckert and Otting 2000).

Analysis of the quality of the lowest-energy structures was performed using PROCHECK-NMR (Laskowski et al. 1996). The statistics from the analysis are shown in Supplemental Table S1.

Molecular dynamic simulations

Molecular dynamic simulations were performed with the Gromacs package (Hess et al. 2008). Prior to the simulations, we generated an extended model of each molecule with CNS, surrounded by a charged-equilibrated, periodic cubic water box. Then, the system was energy-minimized and short position-restrained molecular dynamics was performed to equilibrate the water molecules. Finally, a 40-nsec molecular dynamics in explicit solvent with Particle Mesh Ewald electrostatics was carried out. Calculated structures and the results of the molecular dynamic simulations were analyzed with PyMOL (<http://www.pymol.org>). Sequence alignments were performed using ClustalX (Thompson et al. 1997) and BoxShade 3.21 (http://www.ch.embnet.org/software/BOX_form.html).

Accession numbers

For each of the nine complexes (short names in bold, as in Supplemental Table S1), we deposited 20 structures in the Protein Data Bank (PDB), and the list of restraints and chemical shifts in the BioMagResBank (BMRB) database. For the Smurf1-Smad1 complexes, the corresponding PDB and BMRB codes are, respectively, **WW1-pS214**: 21az, 17541; **WW1-pS210pS214**: 21b0, 17542; and **WW2-PY**: 21b1, 17543. For the YAP-Smad1 complexes, the corresponding PDB and BMRB codes are, respectively, **WW1-pS206**: 21ay, 17540; **WW1-pT202pS206**: 21ax, 17539; and **WW2-PY**: 21aw, 17538. For the Nedd4L-Smad3 complexes, the corresponding PDB and BMRB codes are, respectively, **WW2-pTPY**: 21b2, 17544; and **WW3-pS204pS208**: 21aj, 17529. For the PIN1-Smad3 complex, the corresponding PDB and BMRB codes are, respectively, **WW1-pT**: 21b3, 17545.

Acknowledgments

We thank M. Royo (Parc Científic de Barcelona) for advice on protein chemical ligation and peptide synthesis. N.G. and A.E. have an IRB and FPU-AP2009-3242 PhD fellowship, respectively. This work was supported by SMSI-BFU2008-02795 grant (to M.J.M.) and by NIH grant CA34610 (to J.M.). M.J.M. is an ICREA Programme Investigator. J.M. is an Investigator of the Howard Hughes Medical Institute.

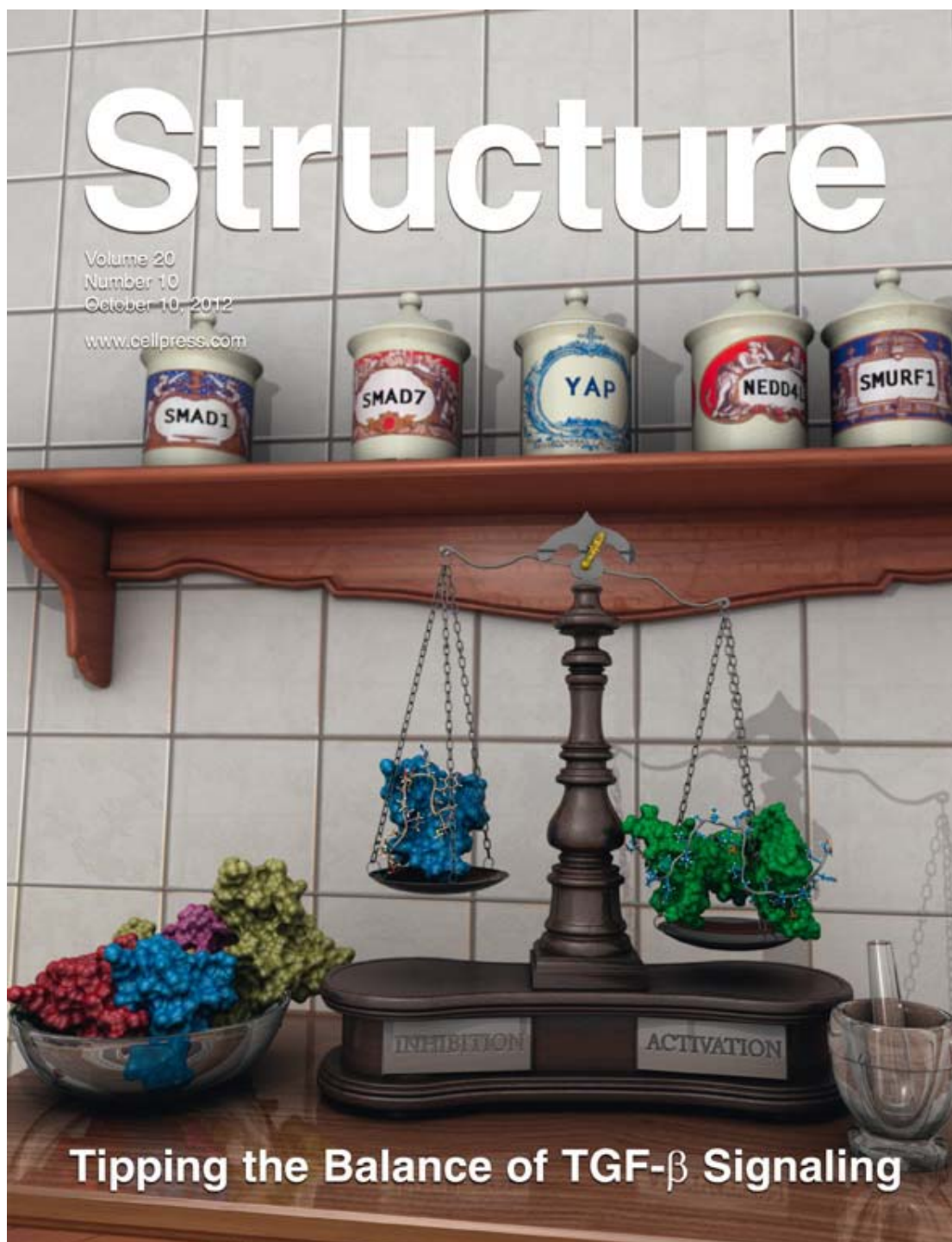
References

- Alarcon C, Zaromytidou AI, Xi Q, Gao S, Yu J, Fujisawa S, Barlas A, Miller AN, Manova-Todorova K, Macias MJ, et al. 2009. Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF- β pathways. *Cell* **139**: 757–769.
- Baca M, Muir TW, Schnölzer M, Kent SBH. 1995. Chemical ligation of cysteine-containing peptides: synthesis of a 22 kDa tethered dimer of HIV-1 protease. *J Am Chem Soc* **116**: 10797–10798.
- Bartels C, Xia T-H, Billeter M, Güntert P, Wüthrich K. 1995. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biol NMR* **5**: 1–10.
- Batut J, Schmierer B, Cao J, Raftery LA, Hill CS, Howell M. 2008. Two highly related regulatory subunits of PP2A exert opposite effects on TGF- β /Activin/Nodal signalling. *Development* **135**: 2927–2937.
- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**: 905–921.
- Cohen P, Frame S. 2001. The renaissance of GSK3. *Nat Rev Mol Cell Biol* **2**: 769–776.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biol NMR* **6**: 277–293.
- Durand LO, Advani SJ, Poon AP, Roizman B. 2005. The carboxyl-terminal domain of RNA polymerase II is phosphorylated by a complex containing cdk9 and infected-cell protein 22 of herpes simplex virus 1. *J Virol* **79**: 6757–6762.
- Ebisawa T, Fukuchi M, Murakami G, Chiba T, Tanaka K, Imamura T, Miyazono K. 2001. Smurf1 interacts with transforming growth factor- β type I receptor through Smad7 and induces receptor degradation. *J Biol Chem* **276**: 12477–12480.
- Fuentealba LC, Eivers E, Ikeda A, Hurtado C, Kuroda H, Pera EM, De Robertis EM. 2007. Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal. *Cell* **131**: 980–993.
- Gao S, Alarcon C, Sapkota G, Rahman S, Chen PY, Goerner N, Macias MJ, Erdjument-Bromage H, Tempst P, Massagué J. 2009. Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF- β signaling. *Mol Cell* **36**: 457–468.
- Guo X, Waddell DS, Wang W, Wang Z, Liberati NT, Yong S, Liu X, Wang XF. 2008. Ligand-dependent ubiquitination of Smad3 is regulated by casein kinase 1 γ 2, an inhibitor of TGF- β signaling. *Oncogene* **27**: 7235–7247.
- Hess B, Kutzner C, van der Spoel D, Lindahl E. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **4**: 435–447.
- Huse M, Muir TW, Xu L, Chen YG, Kuriyan J, Massagué J. 2001. The TGF β receptor activation process: an inhibitor-to substrate-binding switch. *Mol Cell* **8**: 671–682.
- Inman GJ, Nicolas FJ, Hill CS. 2002. Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF- β receptor activity. *Mol Cell* **10**: 283–294.
- Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**: 2452–2460.
- Kretzschmar M, Doody J, Massagué J. 1997. Opposing BMP and EGF signalling pathways converge on the TGF- β family mediator Smad1. *Nature* **389**: 618–622.
- Kretzschmar M, Doody J, Timokhina I, Massagué J. 1999. A mechanism of repression of TGF β /Smad signaling by oncogenic Ras. *Genes Dev* **13**: 804–816.
- Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biol NMR* **8**: 477–486.
- Li MO, Flavell RA. 2008. TGF- β : a master of all T cell trades. *Cell* **134**: 392–404.
- Lin X, Duan X, Liang YY, Su Y, Wrighton KH, Long J, Hu M, Davis CM, Wang J, Brunicaardi FC, et al. 2006. PPM1A functions as a Smad phosphatase to terminate TGF β signaling. *Cell* **125**: 915–928.
- Lo RS, Massagué J. 1999. Ubiquitin-dependent degradation of TGF- β -activated smad2. *Nat Cell Biol* **1**: 472–478.
- Lu K, Yin X, Weng T, Xi S, Li L, Xing G, Cheng X, Yang X, Zhang L, He F. 2008. Targeting WW domains linker of HECT-type

Aragón et al.

- ubiquitin ligase Smurf1 for activation by CKIP-1. *Nat Cell Biol* **10**: 994–1002.
- Macias MJ, Hyvonen M, Baraldi E, Schultz J, Sudol M, Saraste M, Oschkinat H. 1996. Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* **382**: 646–649.
- Macias MJ, Wiesner S, Sudol M. 2002. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett* **513**: 30–37.
- Malik S, Roeder RG. 2000. Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends Biochem Sci* **25**: 277–283.
- Marley J, Lu M, Bracken C. 2001. A method for efficient isotopic labelling of recombinant proteins. *J Biol NMR* **20**: 71–75.
- Massagué J. 2008. TGF β in cancer. *Cell* **134**: 215–230.
- Matsuura I, Chiang KN, Lai CY, He D, Wang G, Ramkumar R, Uchida T, Ryo A, Lu K, Liu F. 2009. Pin1 promotes transforming growth factor- β -induced migration and invasion. *J Biol Chem* **285**: 1754–1764.
- Nair SK, Burley SK. 2003. X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* **112**: 193–205.
- Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H. 1997. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β -spectrin. *J Mol Biol* **269**: 408–422.
- Pires JR, Taha-Nejad F, Toepert F, Ast T, Hoffmuller U, Schneider-Mergener J, Kuhne R, Macias MJ, Oschkinat H. 2001. Solution structures of the YAP65 WW domain and the variant L30K in complex with the peptides GTPPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. *J Mol Biol* **314**: 1147–1156.
- Ruckert M, Otting G. 2000. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* **122**: 7793–7797.
- Sapkota G, Alarcon C, Spagnoli FM, Brivanlou AH, Massagué J. 2007. Balancing BMP signaling through integrated inputs into the Smad1 linker. *Mol Cell* **25**: 441–454.
- Sattler M, Schleucher J, Griesinger C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* **34**: 93–158.
- Schmierer B, Tournier AL, Bates PA, Hill CS. 2008. Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. *Proc Natl Acad Sci* **105**: 6608–6613.
- Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. 2000. SMART: a Web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* **28**: 231–234.
- Shi Y, Massagué J. 2003. Mechanisms of TGF- β signaling from cell membrane to the nucleus. *Cell* **113**: 685–700.
- Shi Y, Wang YF, Jayaraman L, Yang H, Massagué J, Pavletich NP. 1998. Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF- β signaling. *Cell* **94**: 585–594.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.
- Toepert F, Pires JR, Landgraf C, Oschkinat H, Schneider-Mergener J. 2001. Synthesis of an array comprising 837 variants of the hYAP WW protein domain. *Angew Chem Int Ed Engl* **40**: 897–900.
- Wang G, Matsuura I, He D, Liu F. 2009. Transforming growth factor- β -inducible phosphorylation of Smad3. *J Biol Chem* **284**: 9663–9673.
- Wrighton KH, Willis D, Long J, Liu F, Lin X, Feng XH. 2006. Small C-terminal domain phosphatases dephosphorylate the regulatory linker regions of Smad2 and Smad3 to enhance transforming growth factor- β signaling. *J Biol Chem* **281**: 38365–38375.
- Wu MY, Hill CS. 2009. TGF- β superfamily signaling in embryonic development and homeostasis. *Dev Cell* **16**: 329–343.
- Wu D, Pan W. 2010. GSK3: a multifaceted kinase in Wnt signaling. *Trends Biochem Sci* **35**: 161–168.
- Wu JW, Hu M, Chai J, Seoane J, Huse M, Li C, Rigotti DJ, Kyin S, Muir TW, Fairman R, et al. 2001. Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF- β signaling. *Mol Cell* **8**: 1277–1289.
- Xu L, Kang Y, Col S, Massagué J. 2002. Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGF β signaling complexes in the cytoplasm and nucleus. *Mol Cell* **10**: 271–282.

**Structural basis for the versatile interactions of Smad7
with regulator WW domains in TGF- β Pathways.**



On the cover: Smad transcription factors are central mediators of the effects of TGF-beta (Transforming growth factor-beta) and BMP (bone morphogenetic protein) on numerous aspects of metazoan development and tissue homeostasis. The TGF-beta pathway is tightly regulated. Its regulation includes a feedback process where two sets of Smads play complementary roles in the signaling cascade, as they can either inhibit (Smad7) or activate gene transcription (Smad1), depending on the physiological needs of the tissue or organism. We provide a structural basis for how regulators that use WW domain pairs for selective interactions with phosphorylated R-Smads, resort to one single WW domain for binding Smad7 to centralize regulation in the TGF- β and BMP pathways. The cover uses the analogy of a chemist's shop equipped with jars storing some of the proteins participating in TGF signaling. The feedback process is represented as a balance loaded with YAP bound to Smad1 (on the right) and to Smad7 (on the left). For more details, see the article by Aragón et al. (pp. 1726-36). Illustration designed by P. Martín-Malpartida (PMM) and M.J. Macias, and created by PMM.

Structural Basis for the Versatile Interactions of Smad7 with Regulator WW Domains in TGF- β Pathways

Eric Aragón,¹ Nina Goerner,^{1,5} Qiaoran Xi,² Tiago Gomes,¹ Sheng Gao,^{2,6} Joan Massagué,^{2,3,*} and Maria J. Macias^{1,4,*}

¹Structural and Computational Biology Programme, Institute for Research in Biomedicine, Baldori Reixac 10-12, 08028 Barcelona, Spain

²Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

³Howard Hughes Medical Institute (HHMI), Chevy Chase, MD 20185, USA

⁴Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010-Barcelona, Spain

⁵Present address: Biocrates Life Sciences AG, Innrain 66, 6020 Innsbruck, Austria

⁶Present address: Cardiovascular Diseases, Merck Research Laboratory, Merck & Co., 126 E. Lincoln Avenue, Rahway, NJ 07065, USA

*Correspondence: j-massague@ski.mskcc.org (J.M.), maria.macias@irbbarcelona.org (M.J.M.)

<http://dx.doi.org/10.1016/j.str.2012.07.014>

SUMMARY

Transforming growth factor (TGF)- β and BMP signaling is mediated by Smads 1–5 (R-Smads and Co-Smads) and inhibited by Smad7, a major hub of regulation of TGF- β and BMP receptors by negative feedback and antagonistic signals. The transcription coactivator YAP and the E3 ubiquitin ligases Smurf1/2 and Nedd4L target R-Smads for activation or degradation, respectively. Pairs of WW domain in these regulators bind PY motifs and adjacent CDK/MAPK and GSK3 phosphorylation sites in R-Smads in a selective and regulated manner. In contrast, here we show that Smad7 binds YAP, Smurf1, Smurf2, and Nedd4L constitutively, the binding involving a PY motif in Smad7 and no phosphorylation. We also provide a structural basis for how regulators that use WW domain pairs for selective interactions with R-Smads, resort to one single versatile WW domain for binding Smad7 to centralize regulation in the TGF- β and BMP pathways.

INTRODUCTION

Smad transcription factors are key mediators of the transforming growth factor beta (TGF- β) and bone morphogenetic proteins (BMP) pathways in the control of stem cell pluripotency and differentiation, embryo development, tissue regeneration, and differentiated tissue homeostasis (Massagué, 1998). According to their function, Smad proteins are classified as receptor regulated Smads (R-Smads), which include Smads 1, 5, and 8 in the BMP-driven version of the SMAD pathway, and Smads 2 and 3 in the TGF- β /Nodal/Activin pathway. R-Smads form complexes with the common coactivator Smad (Co-Smad) Smad4. Two inhibitory Smads (I-Smads), Smad6 and Smad7, provide critical negative regulation to these powerful and ubiquitous pathways.

R-Smads and Smad4 consist of two Mad Homology domains MH1 and MH2 connected by a linker region. This linker contains

a cluster of phosphorylation sites adjacent to a proline rich PY motif. MH1 domains of R-Smads and Smad4 bind to DNA, whereas the MH2 domain and the linker function as scaffolds for receptors, regulator proteins, and transcription cofactors to interact and determine the outcome of the signal (Shi and Massagué, 2003).

Several key phosphorylation steps regulate the activation and turnover of R-Smads during the TGF- β and BMP signaling cycles. Binding of TGF- β or BMPs to their receptors triggers the receptor-mediated phosphorylation of R-Smads at their C termini. This phosphorylation generates a docking site for Smad4 for the assembly of a heterotrimeric transcriptional complex. Once the complex is in the nucleus, a second round of phosphorylations occurs in the linker region of R-Smads, creating binding sites that interact with the WW domains of activators such as YAP and Pin1, as well as with the WW domains of the HECT-type E3 ubiquitin ligases Nedd4L, Smurf1, and Smurf2 that prime R-Smads for degradation (Alarcón et al., 2009; Fan et al., 2009; Fuentealba et al., 2007; Gao et al., 2009; Kuratomi et al., 2005). These binding interactions depend on WW domain contacts with the PY motif and the phosphorylated sites, and with the exception of Pin1 always involve two WW domains (Aragón et al., 2011).

Compared to R-Smads and Co-Smads, the I-Smads have low sequence similarity in the MH1 domain but conserve an MH2 domain and a linker region with a characteristic PY motif (Figure 1A). I-Smads are expressed in response to TGF- β or BMPs to provide negative feedback in the pathway (Bai and Cao, 2002; Hata et al., 1998; He et al., 2002; Kavsak et al., 2000; Nakao et al., 1997; Yan et al., 2009) and in response to other pathways such as STAT to oppose TGF- β signaling (Ulloa et al., 1999). Smad6 interferes with the formation of Smad1-Smad4 complexes (Hata et al., 1998), whereas Smad7 inhibits TGF- β and BMP receptors (Hayashi et al., 1997; Topper et al., 1998).

Work in recent years has revealed Smad7 as a central hub for negative regulation of activated TGF- β or BMP receptors (Yan and Chen, 2011). Receptor-bound Smad7 recruits ubiquitin ligases Nedd4L, Smurf1, and Smurf2 to mediate receptor poly-ubiquitination and route the receptor to degradative endocytosis (Ebisawa et al., 2001; Kavsak et al., 2000; Kuratomi et al., 2005).

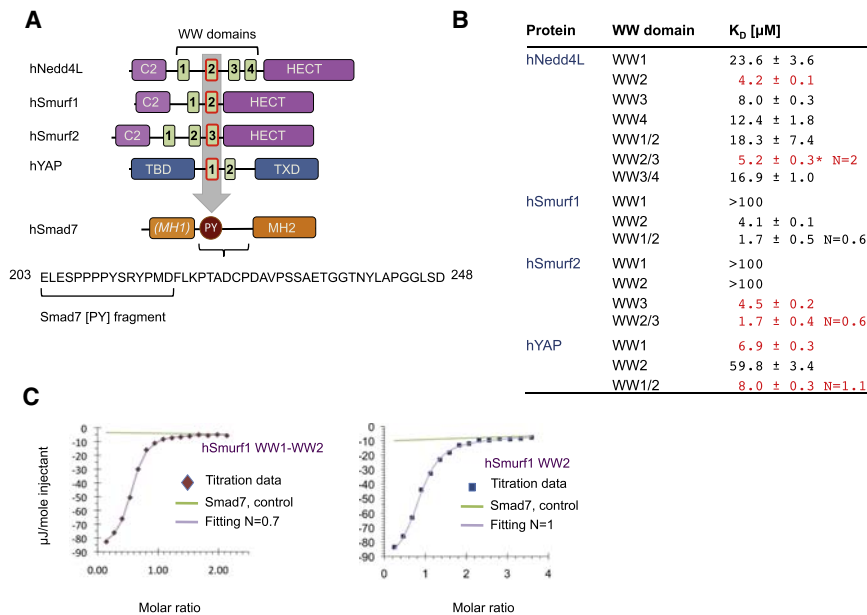


Figure 1. Domain Composition of the Human Nedd4L, Smurf1, Smurf2, YAP, and Smad7 Proteins and Binding Affinities, Determined by Isothermal Titration Calorimetry

(A) The three-ubiquitin ligases contain the characteristic C2 domain, a central region with a variable number of WW domains and the catalytic HECT domain, each domain represented as a rectangle. The human Yes Associated Protein (YAP) contains a TEA binding domain, two WW domains and a transactivator domain (TXD). Smad7 contains a canonical MH2 group and a divergent MH1. To highlight this divergence we have labeled the MH1 domain using italics and in brackets. The detailed sequence of the Smad7 linker is shown (residues 203–248). The synthesized PY peptide is underlined. The WW domains that mediate the interaction with the PY motif of Smad7 are indicated with an arrow.

(B) ITC affinity values for the recombinant fragments of Nedd4L, Smurf1, Smurf2, and YAP and the Smad7 peptide. Binding experiments have been performed at least three times, using two protein expression batches, and different buffers and temperatures. Values hereby presented were obtained at 15°C.

(C) ITC curves for the Smurf1 WW1-WW2 and the independent WW2 fragments in the presence of the Smad7 peptide. The data represented here correspond to values acquired at 15°C in Tris buffer. The affinity values were 1.7 ± 0.3 and 4.1 ± 0.6 μ M with stoichiometries of 0.7 and 1, respectively.

Moreover, Smad7 can simultaneously bind Smurf2 and the protein deubiquitinase USP15, recruiting both enzymes to the TGF- β receptor complex for an integrated control of receptor polyubiquitination as a function of ligand concentration (Eichhorn et al., 2012). Smad7 also binds YAP (Ferrigno et al., 2002) providing a mechanism for sequestration of this mediator of Hippo and BMP signaling (Alarcón et al., 2009). These protein interactions involve the linker region of Smad7 and the WW domain region of Nedd4L, Smurf1/2, and YAP.

In order to advance our understanding of the functional capacities of Smad7 as a hub for TGF- β or BMP pathway regulation, we investigated the interactions between the Smad7 PY motif region and the WW domains of its client proteins, using NMR and ITC, ion mobility mass spectrometry, and protein interaction analysis in mammalian cells. Our data reveal a surprisingly absence of selectivity and independence from phosphorylation in the interactions of Smad7 with these proteins, illuminating the versatility of different WW domains as mediators of convergent interactions with a common Smad7 target, in addition to their discriminating interactions with different R-Smad proteins.

RESULTS

Nedd4L, Smurf1, Smurf2, and YAP Use One Single WW Domain to Bind Smad7

In order to characterize the protein regions involved in the interaction with the Smad7 linker we performed isothermal titration calorimetry (ITC) binding assays using recombinant proteins, either containing independent WW domains or all consecutive pairs. The binding ligand was a 15 residue synthetic peptide corresponding to amino acid residues E203-D217 of Smad7, and including the entire PY motif (Figure 1A).

In the interaction of the Smad7 peptide with Nedd4L WW domains, ITC experiments revealed that each of the four independent WW domains and the three possible WW-WW pairs bind the peptide with dissociation constants in the μ M range (Figure 1B). The stoichiometry of these interactions was 1:1 in all cases except with the WW2-WW3 pair, which bound two equivalents of peptide per protein. To address this complication, we used two previously characterized WW2-WW3 mutant constructs that bind only one PY site while maintaining other possible binding sites active (Aragón et al., 2011). These WW2-WW3 constructs showed peptide affinity values close to those of the single domains. Collectively, the data show that Nedd4L preferentially uses WW2 to interact with the Smad7 PY motif, and that the presence of domain pairs does not increase the affinity in vitro.

We used a similar experimental approach to investigate the interactions of the two E3 ligases Smurf1 and Smurf2 with the Smad7 peptide. The Smurf1 WW1 and Smurf2 WW2 domains show very low affinity for the Smad7 peptide whereas the Smurf1 WW2 and Smurf2 WW3 showed binding dissociation constants of 4.1 μ M and a 1:1 stoichiometry with the Smad7 peptide. The affinity of the Smurf1 WW1-WW2 and Smurf2 WW2-WW3 pairs is 0.7 ± 0.3 μ M at 5°C, 1.7 ± 0.3 μ M at 15°C and 5.0 ± 0.3 μ M at 25°C (Figures 1B and C). These values are in agreement with previous reports for other WW interactions (Aragón et al., 2011; Chong et al., 2006, 2010; Gao et al., 2009; Kanelis et al., 2006; Pires et al., 2001; Ramirez-Espain et al., 2007). However, the affinity increase due to the presence of the WW domain pair is about 2-fold with respect to the values obtained with the Smurf1 WW2 or with the Smurf2 WW3 domains at a given temperature, in contrast to previous observations that suggest an improvement of about 10-fold (Chong et al., 2010). Furthermore, with both protein pairs the affinity is calculated with

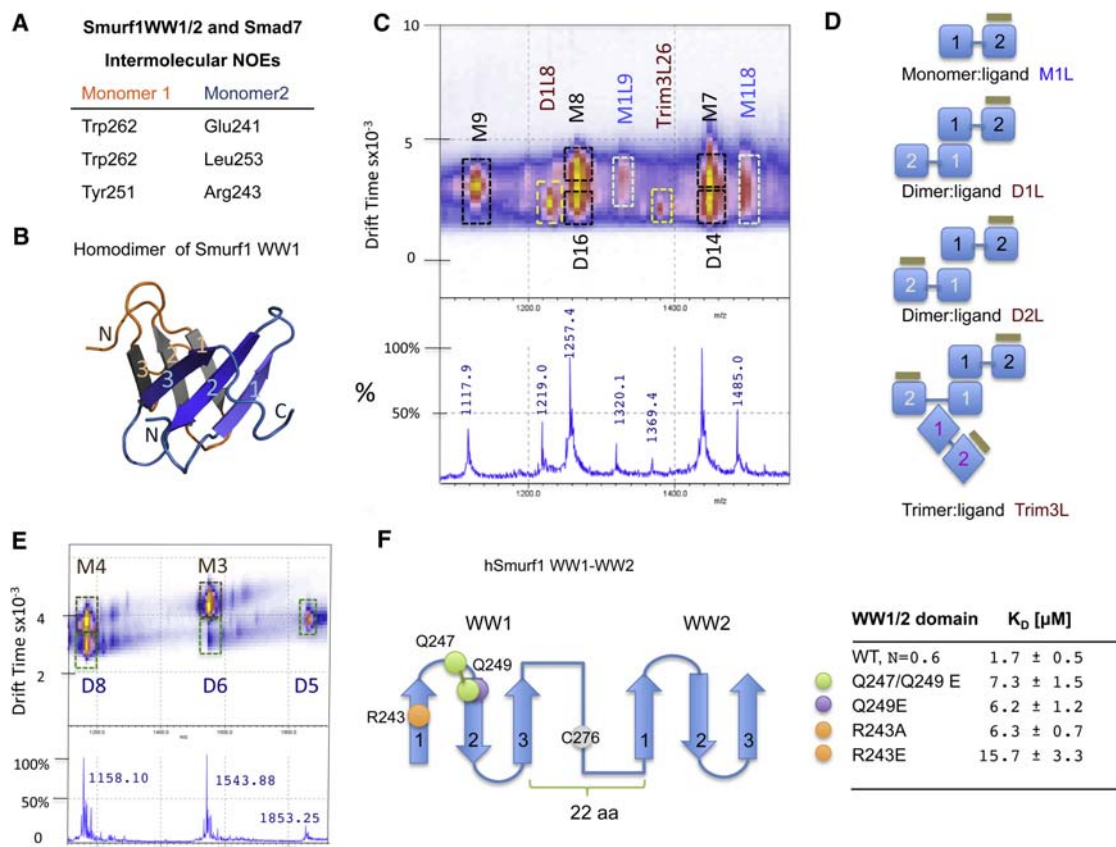


Figure 2. Smurf1WW1-WW2 and Smurf2WW2-WW3 Form Different Species in the Presence of Smad7 Peptide

(A) Assigned NOEs defining the hSmurf1WW1-WW1 dimer interface. Schematic representations of the possible species present in the complexes of Smurf1 WW1-WW2 pair with Smad7.

(B) Schematic representation of the Smurf1 WW1 dimer. Secondary structure elements are shown in gray and in blue shades and numbered. The model has been generated using the structure of the published Smurf1 WW1 domain (Aragón et al., 2011) pdb entry 2laz and the set of assigned intermolecular NOEs.

(C) An expansion of the IM-MS data obtained for the complexes of Smurf2WW2-WW3 with the Smad7 peptide displaying nine characterized species (the full spectrum is shown as Figure S1). Each ion was assigned to a given species based on its characteristic mobility. Abbreviations used are ML (monomer with one ligand), D1L and D2L (dimer with one or two ligands respectively), Trim1L, Trim2L, and Trim3L (trimers with one, two, or three ligands). Numbers following the species' name reflect the protonation state. We have unambiguously detected dimers with one or with two bound ligands and trimers in both Smurf1 and Smurf2 complexes.

(D) Schematic representation of the species identified by IM-MS for the complexes of Smurf1 WW1-WW2 and Smurf2WW2-WW3 pairs with the Smad7 PY site. The WW domains are represented as blue rectangles (labeled in black, white and violet to represent monomer, dimer and trimers, respectively). The Smad7 peptide is represented as a green thick-line on top of the WW2 domain. Contacts involving two WW1 domains or between the WW1 domain of one molecule with the linker connecting the WW pair of a second molecule are based on experimental NOEs.

(E) A region of the Ion Mobility-Mass Spectrometry data obtained for the Smurf1 WW1 dimer, (the full spectrum is shown as Figure S1). As in (C), each ion was assigned to a given species based on its characteristic drift-time.

(F) Schematic representation of the pair of WW domains and the linker present in Smurf1. The mutated positions used in the ITC binding experiments and the three strands of each WW domain are labeled. Next to it is the list of the measured affinity values.

a stoichiometry below 1 at all temperatures (Figure 1C) (0.6–0.8 range). ITC experiments measured in two different buffer solutions (tris and ammonium acetate [pH 7]) yielded similar values and stoichiometries.

We considered the possibility that formation of WW-WW protein aggregates could affect the interpretation of the stoichiometry and affinity of Smad7 binding. Indeed, using nuclear magnetic resonance (NMR), we observed nuclear Overhauser effects (NOEs) in Smurf1 samples containing either the WW1 independent domain or the WW1-WW2 pair that define a dimer via WW1-WW1 contacts (Figures 2A and 2B). Since NMR and ITC experiments were carried out at different concentration

ranges (millimolar versus micromolar), we made use of ion mobility-mass spectrometry (IM-MS) to investigate the potential presence of aggregates at the protein concentrations used in ITC experiments (30–50 μ M). Using this technique we identified the presence of monomeric complexes, protein dimers bound to only one ligand, dimers bound to two ligands and higher order species (Figures 2C and 2D; Figure S1 available online). The dimers, trimers, and other higher order species were identified based on their different specific ionization masses and/or on their characteristic drift-times. Dimers were also observed by IM-MS analysis performed with samples containing either the WW1 or the WW2 domains of Smurf1 (Figure 2E). In the presence

of the Smad7 peptide the dimer population was reduced in the sample containing the WW2 domain, while that of the WW1 domain was unaffected (Figure S1). Thus, the <1 stoichiometry observed with the pairs under these experimental conditions may result from the coexistence of protein monomers, dimers, trimers, and tetramers each binding one equivalent of Smad7 peptide, plus the presence of other species bound to two, three, or four Smad7 equivalents (Figures 2C and 2D). Since the presence of the ligand does not prevent the formation of the higher order species, we interpret that Smurf1 WW1 and Smurf2 WW2 domains have a minor role in binding to short PY containing sequences, but an important role in protein oligomerization and aggregation. This observation differs from a previously reported interpretation of Smurf2 WW2-WW3 bound to Smad7 (Chong et al., 2010), where the WW2 domain also participates in hydrophobic and electrostatic interactions with the Smad7 peptide. To characterize further the role of Smurf1 WW1 domain, we introduced mutations in the WW1-WW2 pair domain in equivalent positions to those that in Smurf2 were proposed to contact Smad7, and also two additional control mutations in a position that did not participate in the interaction of Smurf2 and Smad7 but in the dimer formation. In all cases, single and double mutations (Arg243Ala, Gln249Glu, Gln247, and Gln249 to Glu) reduced the affinity by 3-fold while the Arg243Glu mutation reduced the affinity by 10-fold (Figure 2F). The results suggest that these residues do not play a key role in binding but that they may participate in protein homodimerization. Based on this interpretation of the binding data, we conclude that Smurf1 and Smurf2 use their WW2 domain and WW3 domain, respectively, as their primary binding sites for the Smad7 PY peptide.

ITC titrations performed at 15°C with the YAP WW domains revealed that the YAP WW1 domain preferentially binds to the Smad7 peptide, with a dissociation constant of $6.90 \pm 0.28 \mu\text{M}$, while WW2 binds with a dissociation constant 9-fold weaker and the WW1-WW2 pair binds slightly worse than the isolated WW1 ($9.8 \pm 0.9 \mu\text{M}$ and $N = 1$; Figure 1B). Thus, the interaction with the Smad7 peptide mainly involves the YAP WW1 domain.

All together, these results suggest that in each protein a specific, single WW domain is sufficient for high-affinity recognition of the Smad7 PY site. The WW domains that mediate the interaction with the PY motif of Smad7 are indicated with an arrow in the Figure 1A.

Smad7 Binds to Nedd4L WW2 Domain, Forming a Long beta Hairpin Independent of Phosphorylation

We used NMR spectroscopy to characterize with atomic detail, the interaction of the Nedd4L WW2 domain bound to the Smad7 linker peptide (203–217). Triple resonance NMR spectroscopy was applied to assign the WW2 domain in this complex, whereas filter and homonuclear spectra were used to assign the bound Smad7 peptide and its contacts with the domain.

In the complex structure, the Nedd4L WW2 domain adopts the canonical WW fold, while the bound peptide forms an ordered hairpin from E205 until D217, with a turn centered at positions Y211-S212-R213. The complex is well defined, based on abundant contacts detected from the Smad7 peptide with residues located in the three strands of the domain (Figure 3A and Fig-

ure S2; Table 1). Y211 participates in many contacts with Val384, His386, Arg389, and Thr391 residues in the protein, while P207, P208, and P209 interact with the side chains of Thr391, Trp393, and Tyr382 (we use the one letter amino acid notation for Smad7 residues and the three letter notation for residues in the proteins). P215 is perpendicular to the beta-sheet plane and is bound by the Tyr391, Val384, and Arg374 side chains. We also observed interactions between E205, located upstream of the PY motif, with the side chains of Arg380 and of D217 with Lys378 and Arg380 (Figure 3B). The Smad7 PY motif includes a potential phosphorylation site serine (Ser206-Pro), equivalent to the Thr179-Pro of Smad3 whose phosphorylation by CDK8/9 kinases is critical for Smad3 recognition by Pin1 and Nedd4L (Gao et al., 2009; Matsuura et al., 2010).

In contrast to the complex formed by Nedd4L and the Smad3 pT179[PY] motif (Aragón et al., 2011), we did not detect contacts between S206 (equivalent to T179) and the protein. Remarkably, the interaction of E205 and also D217 with Arg380 and Lys378 resemble the electrostatic interactions used to interact with the phosphorylated T179 in the Nedd4L WW2-Smad3 pT179[PY] complex (Aragón et al., 2011). To clarify the relevance that the E205 and D217 contacts have for the complex, we designed protein variants in which either Arg380 or Lys378, or both, were replaced by a negatively charged residue (Glu), and quantified the changes in affinity using ITC. Single changes reduce the affinity for the Smad7 peptide 4- to 5-fold when compared to the wild-type (Arg380Glu 15.4 ± 3.5 , Lys378Glu $23.5 \pm 1.5 \mu\text{M}$, respectively) and more than 6-fold in the double mutant (Lys378/Arg380 to Glu, $K_D = 27.3 \pm 2.2 \mu\text{M}$), corroborating that both residues in the WW2 domain participate in the interaction with the Smad7 fragment (Figure 3C).

To further investigate the role of S206 in Smad7 binding interactions, we addressed this question with the full-length proteins in the context of HEK293T cell line. Immunoprecipitation experiments were performed using plasmids encoding flag epitope-tagged Smad7 or the Smad7 mutants S206A and AAPY. These constructs were coexpressed with a HA epitope-tagged Nedd4L construct in which the HECT domain catalytic Cys was mutated to Asp in order to prevent autoubiquitination, and degradation of the protein (Gao et al., 2009). The results of protein immunoprecipitation followed by western immunoblotting showed that the Nedd4L-Smad7 interaction in mammalian cells does not require phosphorylation of the PY motif (Figure 3D). This result is in sharp contrast to the important role of the corresponding phosphorylation in the interaction between Nedd4L and Smad3 (Gao et al., 2009). The different binding modes of Nedd4L with Smad3 and with Smad7 are schematically summarized in Figure 3E.

Structure of the Smad7 PY Motif Bound to Different E3 Ubiquitin Ligase WW Domains

To compare the binding modes of the E3 ubiquitin ligases Nedd4L, Smurf1, and Smurf2, we investigated the interactions between the Smurf1 WW1-WW2 pair and the independent WW2 domain with the Smad7 peptide using triple resonance NMR spectroscopy. Under these conditions the NMR assignment of the complexes reveals that, both Smurf1 and Smurf2 proteins interact with the Smad7 peptide using the WW2 domain, in the case of Smurf1, and the WW3 for Smurf2. As observed during the ITC titrations and IM-MS experiments, the Smurf1

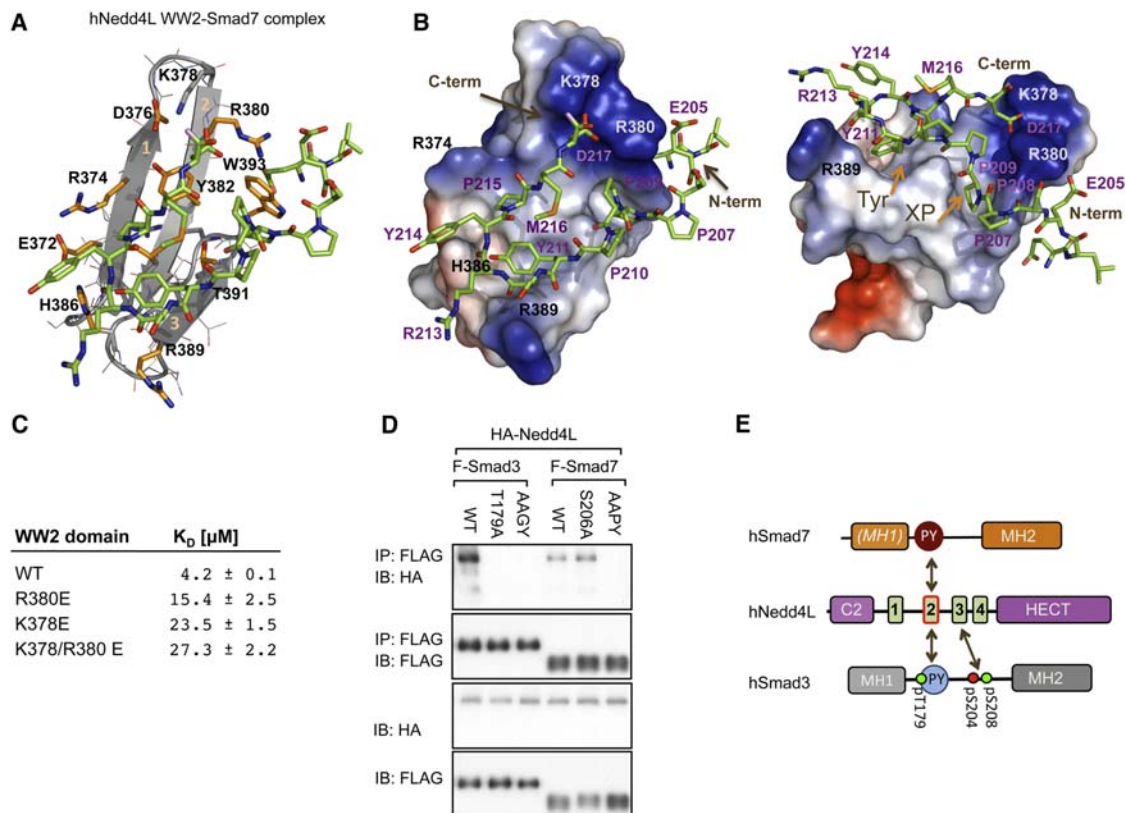


Figure 3. Structure of the Nedd4L WW2 Domain Bound to the Smad7 Linker

(A) Detailed view of the refined structure of the Nedd4L WW2 domain (364–403) in complex with the Smad7 synthesized PY fragment (203–217). The elements of secondary structure (graphite) are numbered and key residues of the Nedd4L WW2 domain that participate in the interaction with the peptide are labeled and highlighted in orange.

(B) Electrostatic potential surface of the Nedd4L WW2 domain with the bound Smad7 peptide (stick representation) colored in green. The lowest energy structure displayed in (A) and (B) and the family of 25-calculated structures are shown as Figures S2A and S2B. Positively charged sites are colored in blue and negatively charged sites in red. Key residues in Smad7 and Nedd4L are indicated (violet and black, respectively). The N- and C-terminal sites of the Smad7 peptide are indicated with an arrow.

(C) Two residues that recognize the N-terminal part of Smad7 were both independently and jointly mutated to glutamic acid, and the effect of the change in the interaction affinity was determined by ITC. The binding affinity decreased by approximately 7-fold with respect to the wild-type Nedd4L WW2 domain.

(D) Human HEK293T cells transduced with vectors encoding the indicated epitope-tagged proteins were subjected to immunoprecipitation (IP) with anti-Flag antibodies and western immunoblotting with antibodies against the indicated tags (upper panels). Aliquots of cell lysates were directly subjected to immunoblotting as loading controls (bottom panels).

(E) Schematic representation of the different binding modes displayed by Nedd4L with Smad3 and Smad7. Nedd4L binds preferentially to the PY site of Smad7 using its WW2 domain (this work) while binding to Smad3 region including the diphosphorylated site and pT[PY] motif requires the WW2-WW3 pair (Aragón et al., 2011).

WW1-WW2 pair displays a high tendency to form dimers and other higher order aggregates.

Using 3D- and 2D-NOESY experiments we characterized a population of dimers formed via interactions between the WW1 domains of two molecules according to intermolecular NOEs that fit as a beta-clam. To illustrate these interactions, we generated a model using the structure of the Smurf1 WW1 domain (pdb entry: 2laz) and the unambiguously assigned NOEs detected between monomers (Figures 2A and 4A). We have also detected a minor population of dimers formed by interactions between the WW2 domain of one molecule and the pair of prolines present in the linker connecting the WW1-WW2 pair of a second molecule, that can explain the trimeric and tetrameric species identified by IM-MS (Figures 2C and 2D). We observed as well the dimerizing tendency with the WW2 domain of Smurf2

in the WW2-WW3 pair. The beta-clam arrangement in the dimer is similar to that described for the WW2 domain of the mouse Salvador homolog 1 protein (Ohnishi et al., 2007). With the WW domain pairs we detected a broadening of the intermolecular NOEs that defined the peptide in the bound conformation and two sets of NOEs for the Y211 with residues in the WW2/WW3 domains. We interpreted the broadening and the presence of the second set of signals for the Y211 aromatic ring as the result of the peptide bound in several complexes, for instance, the main conformations that correspond to the monomer in complex with one ligand and the symmetric dimer with two bound ligands in equilibrium with an asymmetric dimer bound to a single ligand, schematically represented in Figure 2D. At 298K and in the presence of 10% DMSO both sets of NOEs corresponding to the Y211 collapse to one set that we interpret it as the bound

Table 1. Statistics of the Five NMR Refined Complexes Analyzed with Procheck

	<i>Nedd4L</i> WW2-S7 pdb:2lty BMRB:18501	<i>Smurf1</i> WW2-S7 pdb:2ltx BMRB:18500	<i>Smurf2</i> WW3-S7 pdb:2ltz BMRB:18502	<i>YAP</i> WW2-S7 pdb:2ltw BMRB:18499	<i>YAP</i> WW1-S7 pdb:2ltv BMRB:18498
Restrains used for the calculation < SA > ^a					
Interdomain	51	85	76	74	53
Sequential ($ i - j = 1$)	270	116	150	163	150
Medium range ($1 < i - j \leq 4$)	80	41	52	85	83
Long range ($ i - j > 4$)	339	251	338	295	315
Dihedrals	70	113	91	105	101
Hydrogen bonds	10	10	10	10	10
All restraints (unambiguous)	820	616	717	732	712
Restraint ratio (47 residues)	17.4	13.1	15.2	15.5	15.2
Rmsd (Å) from experimental ^b					
NOE ($\times 10^{-3}$):	2.7 ± 2.8	9.1 ± 0.3	10.5 ± 1.6	3.1 ± 0.4	7.30 ± 0.5
Bonds ($\times 10^{-3}$) (Å)	4.2 ± 1.4	7.0 ± 0.2	7.1 ± 0.3	4.1 ± 0.2	4.9 ± 0.1
Angles (°)	0.63 ± 0.01	0.8 ± 0.02	0.83 ± 0.06	0.5 ± 0.03	0.7 ± 0.03
Coordinate Precision (Å) ^c					
Backbone, all residues in the complex (47 residues)	0.45	0.42	0.47	0.31	0.24
CNS potential energy (kcal mol ⁻¹)					
Total energy ^d	-1,698 ± 34	-1483 ± 47	-1238 ± 88	-1,242 ± 31	-1361 ± 31
Electrostatic	-1,918 ± 38	-1784 ± 48	-1787 ± 56	-1,513 ± 33	-1713 ± 36
Van der Waals	-165.5 ± 10.4	-120.8 ± 14	-57.38 ± 39.4	-135.3 ± 10	-113 ± 10
Bonds	14.44 ± 1.3	27.09 ± 3.5	42.91 ± 10.5	14.2 ± 1.0	19.63 ± 1.4
Angles	89.11 ± 5.4	81.63 ± 6.9	158.9 ± 23	95.18 ± 7.9	107 ± 8.4
Structural quality (% residues) 20 best structures					
In most favored region of Ramachandran plot	84.1	87.1	87.5	90.7	89.8
In additionally allowed region	13.9	12.4	12.2	9.0	11.2

^a< SA > refers to the ensemble of 150 structures with the lowest energy selected from a total of 300 calculated structures.

^bNo distance restraint in any of the structures included in the ensemble was violated by more than 0.3 Å.

^cRmsd between the ensemble of structures < SA > and the lowest energy structure.

^dEL-J is the Lennard-Jones van der Waals energy calculated using the CHARMM-PARMALLH6 parameters. EL-J was not included in the target function during the structure calculation.

monomer. Under these experimental conditions, we did not observe, however, contacts between the domains in the Smurf1 WW1-WW2 pair, or between the WW1 and the Smad7 peptide as described for the Smurf2 WW2-WW3 complex with Smad7 (Chong et al., 2010), or as we previously observed in the complex of Smurf1 WW1-WW2 pair with Smad1 (Aragón et al., 2011).

Based on these observations, we focused the structural work in the WW2 domain of Smurf1 and in the WW3 of Smurf2. As both Smurf complexes are very similar, we will describe them in parallel, with the corresponding residues separated by a slash. The Smurf1WW2- and Smurf2WW3-Smad7 complexes are well defined, based on numerous contacts detected from E205-P215 Smad7 residues with the domains (Figures 4B and 4C and Figure S3; Table 1). In both complexes the Smad7 fragment also forms a turn, centered at positions Y211-S212-R213, but it does not form a long hairpin as in the case of the Nedd4L complex, especially in the Smurf1 complex. A comparison of these two complexes with that of Nedd4L revealed some additional differences; for instance, E205 is interacting with Arg295/Arg312 in the second strand, but no contacts with the peptide

are observed for the residues located in loop 1 of the WW domains, which in these cases are Ser293/Thr310 and not a Lys as in the Nedd4L WW2 domain. As a consequence, E205 and D217 are less defined in the Smurf1 WW2 and Smurf2 WW3 complexes.

A mutation introduced in the Smurf1 WW2 domain (Arg295Glu) reduces the affinity to $48.6 \pm 6.4 \mu\text{M}$, suggesting an active implication of Arg295 in the peptide interaction. The equivalent mutation introduced in Smurf2WW3 also reduces the affinity to $34.2 \pm 1.6 \mu\text{M}$. The complexes here described are similar to the previously characterized Smurf2WW3 and Smad7 (Chong et al., 2006), with the main differences involving the contacts with the N-terminal site of Smad7 (E205), the position of P215, and the absence of intra peptide contacts from residues M216-D217. On the other hand, they differ from the complex between the Smurf1WW1-WW2 pair and Smad1, where both WW domains have a direct role in ligand recognition (Figure 4D) (Alarcón et al., 2009; Aragón et al., 2011) and from the complex between Smurf2WW2-WW3 and Smad7, where the contacts with the C-terminal part of Smad7 that we observe to occur with the

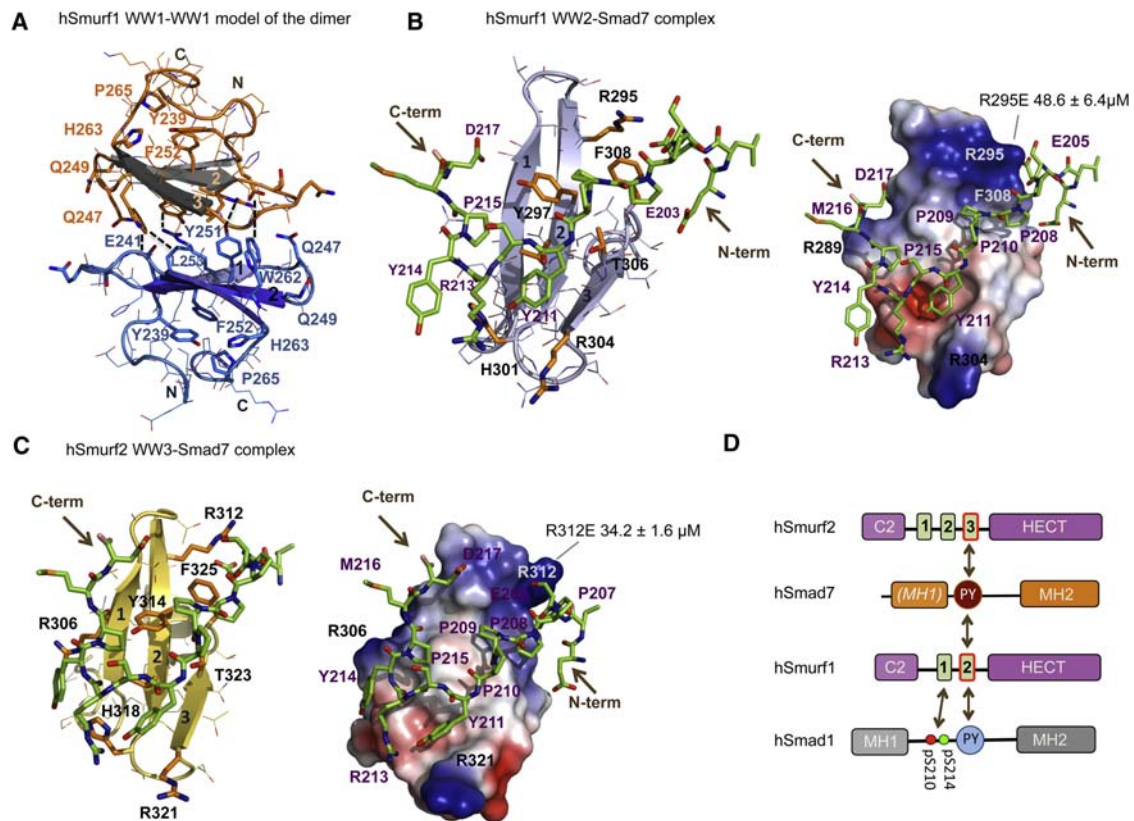


Figure 4. Structure of the Smurf1 WW2 and Smurf2 WW3 Domains Bound to the Smad7 Linker

(A) A detailed view of the binding interface of the dimer between two Smurf1 WW1 domains (residues 233–270). The orientation is rotated by 90 degrees with respect to the schematic representation shown in Figure 2B. The unambiguously assigned NOEs that defined the dimer interface are represented with dotted lines. Monomers are colored in orange-gray (top) and cobalt blue (bottom), with key residues highlighted.

(B) Refined structure of the Smurf1 WW2 domain (residues 277–314, light-blue) bound to the Smad7 peptide (203–217, green). Next to it is the charge distribution on the surface of the Smurf1 WW2 domain in complex with the peptide. The structure with the lowest energy was selected for both representations. The family of 25-calculated structures is shown as Figure S3A. N and C termini of Smad7 peptide are represented with brown arrows. Key residues that participate in the complex interaction are labeled in purple (Smad7) and in black (Smurf1). The gray line indicates the position of the R295E mutation, which decreases the binding affinity of the complex by approximately 8-fold.

(C) Same representations as above for the Smurf2 WW3 domain (297–333, gold) bound to the Smad7 peptide (green). The R312E mutation, which decreases the binding affinity of the complex by approximately 10-fold, is indicated with a gray line. The family of 25-calculated structures is shown as Figure S3B.

(D) Schematic representation of the different binding modes observed for Smurf1 and Smurf2 WW domains with respect to R- and I-Smad linkers. Both Smurf proteins use a unique domain to interact with Smad7 (this work) while Smurf1 uses the WW1-WW2 pair to interact with a diphosphorylated Smad1 PY containing site (Alarcón et al., 2009; Aragón et al., 2011).

WW3 domain were detected with the first loop of the WW2 domain (Chong et al., 2010).

Smad7 Selects YAP WW1 and Not WW2

Both YAP WW domains have been previously shown to recognize PY motifs with different affinities (Aragón et al., 2011; Chen and Sudol, 1995; Macias et al., 1996; Pires et al., 2001). The Smad7 fragment containing the PY motif binds to the YAP WW1-WW2 pair and to its independent WW1 domain with similar affinities suggesting that the WW1 is the preferred binding site. To provide a structure-based interpretation for the different affinities observed between the YAP WW1 and WW2 domains, we have determined the structure of each domain in complex with the Smad7 fragment. As both YAP WW1 and WW2 sequences are very similar (71% similarity), we also describe them in parallel, with the corresponding residues separated by a slash.

In each complex the WW1 and WW2 domains adopt the canonical WW fold and bind to the Smad7 PY core in a similar manner (Figures 5A and 5B and Figure S4; Table 1). The Y211 ring is accommodated in its respective tyrosine binding cavities formed by Leu190/Ile249, His192/His251, Gln195/Lys254, and Thr197/Thr256, while the pyrrolidine rings of Smad7 P208 and P209 are, respectively, accommodated in the XP cavities formed by Tyr188/Tyr247 and Trp199/Trp258. The main differences between both complexes are the contacts observed between Smad7 Y214 and Tyr247, Ile249, and Glu237 in the WW2 complex and more significantly, the absence of interactions between the residues located in loop 1 and the residues preceding or following the PY site, which are observed in the complex with WW1, and in the complexes corresponding to the three E3 ubiquitin ligases. The absence of these interactions could be interpreted on the basis of the charge distribution of the YAP WW2

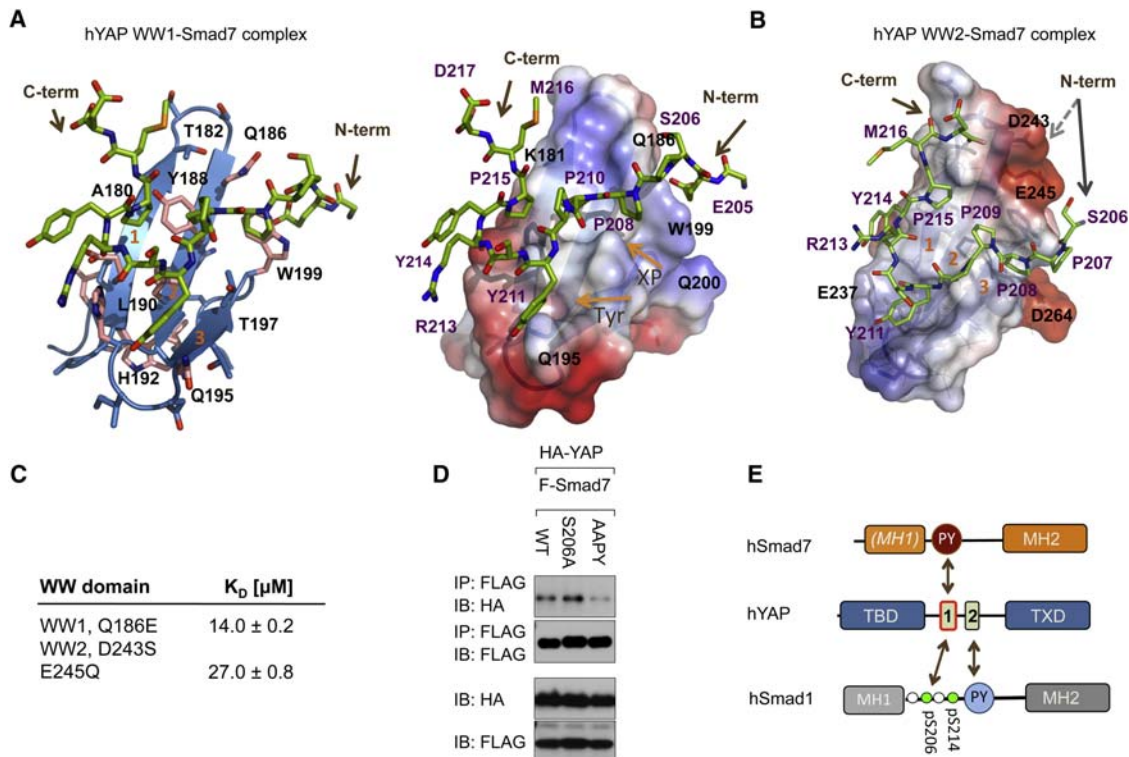


Figure 5. Structure of the YAP WW1 and WW2 Domains Bound to the Smad7 Linker

(A) Detailed view of the interaction of the YAP WW1 domain (residues 163–206, blue) with the Smad7 [PY] peptide (green) and next to it, the charge distribution of the domain in the complex, shown as a semitransparent surface representation. The structure with the lowest energy was selected for both representations. The family of 25 calculated structures is shown as Figures S4A and SB. Critical residues involved in the interaction are labeled in purple (Smad7) and in black (YAP WW1). Both N and C termini of Smad7 peptide are represented with arrows.

(B) A semitransparent surface representation showing the charge distribution of the YAP WW2 domain (residues 227–266, deep blue) bound to the Smad7 peptide represented as sticks (green). The family of 25 calculated structures is shown as Figure S4B. The peptide's N terminus (Ser206) is shifted from the loop1 with respect to the orientation in the complex with WW1. The different positions of the peptide in both complexes are shown with two arrows, with a straight gray line (WW2 complex) and with a dotted gray line (the position in the WW1 complex).

(C) Effect of point mutations in the YAP WW1 and WW2 domains. The Q186E change in the WW1 domain reduces the affinity by 2-fold, while a double change introduced in the WW2 loop—to mimic the sequence of the first WW domain—improves the affinity by a factor of two. Mutated residues are indicated in (A) and (B).

(D) HEK293T cells expressing the indicated constructs were analyzed as described in Figure 3D.

(E) Schematic representation of the binding modes of YAP with Smad7 (top) and with Smad1 (bottom). The interaction with the PY site of Smad7 requires only the first WW domain, while binding to the pS and PY sites of Smad1 requires both WW domains.

domain, which differs from that of the WW1 domain. The presence of negatively charged residues in loop 1 (Asp243 and Glu245) repels both the accommodation of the N-terminal part of Smad7 as well as the negatively charged residues located in the C-terminal extension of the PY motif. Point mutations introduced in Asp243 and Glu245 residues to glutamine resulted in an improved affinity by a factor of two (Figure 5C).

The YAP WW1-Smad7 structure does not support a role (or room) for a phosphate group on S206 (S206 is tightly bound by Trp199), suggesting that Ser206 phosphorylation would not enhance the interaction with YAP *in vivo*. To explore this possibility, we made use of IP experiments with either full-length protein Smad7 (wild-type) or with the S206A and AAPY variants transfected into the HEK293T cell line and the HA-YAP construct. Our experiments showed that binding is observed with both the wild-type and the S206A variant, while the AAPY variant cannot precipitate HA-YAP (Figure 5D) suggesting that the Ser206 or its potential modification would play a minor role in

the function. In summary, the first WW domain of YAP, which is conserved in both YAP splicing variants (Sudol, 1994), is the preferred binding site for the Smad7 PY motif. The discrimination between both domains is based on the more extensive protein-peptide contacts observed with the WW1 compared with the WW2 domain. Again, the interaction of YAP with Smad7 is different from that with Smad1, where both WW domains participate in the complex and where WW2 is responsible for the interaction with the Smad1 PY site while the WW1 recognizes the phosphorylated pSerPro motif (Figure 5E).

DISCUSSION

The rules governing target recognition by HECT type E3 ubiquitin ligases are open questions, and the work to date reflects a more complex scenario than originally expected. Some family members recognize PY motifs using a single WW domain, as is the case of Itch binding to the Epstein Barr virus protein

LMP2A (Morales et al., 2007) and Nedd4 binding to the voltage gated sodium channel (Kanelis et al., 2001) and to Commissureless (Kanelis et al., 2006). In other cases, such as the binding of Smurf1 and Nedd4L to R-Smads, the proteins use a pair of WW domains to expand the binding interface with a composite binding site that includes pSer/pThr-Pro elements in addition to a canonical PY motif, a combination that allows regulation of the interaction by input-driven protein kinases (Alarcón et al., 2009; Aragón et al., 2011; Gao et al., 2009).

The data here presented show a further versatility of WW domains depending on the target protein in that a given HECT E3 ubiquitin ligase can use WW domains singly (this work) or in a combinatorial manner (Alarcón et al., 2009; Aragón et al., 2011; Gao et al., 2009), depending on the target. In the case of Nedd4L, the interaction with Smad7 involves a single WW domain (WW2 preferentially). The high affinity of the Nedd4L WW2 domain for the Smad7 PY region is based on aggregate contacts with a canonical PY motif, a C-terminal extension of this motif that aligns on the first strand of the WW2 domain, and an electrostatic balance between a glutamic and an aspartic acids (E205 and D217) in Smad7 and two positively charged residues located in loop 1 of the WW2 domain. In Smad2/3 a phosphothreonine N-terminal to the PY motif (pT179) makes a critical contribution in the binding to a Nedd4L WW domain (Gao et al., 2009). In the case of Smad7 a corresponding serine residue (S206) does not need to be phosphorylated for high affinity binding to Nedd4L; instead, an acidic residue, E205, plays the part of pT179 in Smad3. Furthermore, Nedd4L binding to Smad2/3 involves a second WW domain, WW3, for contacts with a separate, diphosphorylated region downstream of the PY motif (Aragón et al., 2011). Thus Nedd4L binds Smad7 using a single unregulated WW domain interaction even though it binds Smad2/3 using two WW domains that require multiple phosphorylations of the target region by different protein kinases.

Similar principles govern the interactions of Smurf1 and Smurf2 ubiquitin ligases with Smad7. Binding involves a unique WW domain, whereas Smurf1 binds Smad1 using two WW domains to recognize a PY motif and a phosphorylated motif in the linker region. We observed neither the contacts between the two WW domains nor the contacts between the first WW domain and Smad7 that were described in a recent report (Chong et al., 2010). We observed instead that the Smurf1 and Smurf2 WW-WW pairs have a high tendency to form homodimers via the WW1 domain in case of Smurf1 and via the WW2 domain in Smurf2 (Figure 4A). The presence of Smad7 peptide did not prevent these dimerizations since the WW-WW domain pairs bind the Smad7 peptide mainly through contacts with the WW2 domain of Smurf1 and with the WW3 domain of Smurf2. It has recently been reported that full-length Smurf1 forms homodimers and oligomers in vitro and in vivo through intermolecular contacts mapped to a fragment containing the C2 and the WW domains of one molecule and the HECT domain of the partner (Wan et al., 2011). Furthermore, intramolecular contacts between the C2 domain and the HECT domain of Smurf2 have also been characterized (Wiesner et al., 2007). In both Smurf1/2 ligases the close conformation inhibits the mechanism of protein self-ubiquitination (Wan et al., 2011). It is possible that in addition to these reported interactions between the C2 and HECT domains, the inter WW-WW contacts detected

in our work could also contribute to the formation of the dimers and oligomers in vivo, and to the stabilization of the close conformation of Smurf1 and 2. In the presence of two Smad7 equivalents, a reverse reaction may occur with the MH1 and PY sites of the Smad7 protein competing for the Smurf1 HECT and the WW2 domains, respectively, pulling apart the WW1-WW1 dimer. The result of this reaction would be the generation of two activated Smurf1-Smad7 complexes, starting from the close and inactive full-length dimer. A similar mechanism may occur with the Smurf1 and Smad1 interaction. These possibilities notwithstanding, the WW-WW homodimers here detected could have resulted from the use of recombinant protein fragments.

Our work also shows that YAP requires only the WW1 domain for binding to the Smad7 PY region. Notably, in the interaction with Smad1, YAP uses its WW1 domain for binding a phosphoserine motif, and instead uses the WW2 domain for binding the Smad1 PY motif (Aragón et al., 2011). By solving the structures of both YAP WW1 and WW2 complexes with Smad7 we discerned the reasons for these differences with respect to the interaction with Smad1. The WW2 domain contains negatively charged residues in the area where the E205 side chain is normally accommodated, destabilizing the interaction. The affinity of the YAP WW1-WW2 pair for the composed pSP-PY site of Smad1 is 8x higher than that of the WW1 domain for the PY site of Smad7. However, the concentration of Smad7 in the nucleus is high and it could compete in vivo with Smad1 for YAP binding, providing a scenario for the inhibitory role of Smad7.

We propose that WW-WW pairs in these Smad regulators form functional units that evolved to recognize PY containing regions of variable length and complexity, including composite PY/phospho-Ser/Thr motifs in R-Smads and simple PY motifs in Smad7. These features expand the functional versatility of E3 ubiquitin ligases by optimizing the interacting surface depending on the needs. With Smad7, Nedd4L and Smurf1/2 act as partners in targeting TGF- β receptors for ubiquitination. Smad7 may also act as a constitutive YAP sequestration or reservoir protein. In contrast, R-Smads are direct targets of the ubiquitin ligases and functional partners of YAP only in specific stages of the Smad signaling cycle (Aragón et al., 2011). The absence of a requirement for phosphorylation in the interaction with Smad7 argues that YAP, Nedd4L and Smurf1/2 are constitutive partners of Smad7 whereas they are conditional, phosphorylation dependent regulators of R-Smads in TGF- β and BMP signal transduction. The features of Smad7 defined here provide a structural basis for its central role as a hub for negative feedback and crosstalk regulation in TGF- β signaling.

EXPERIMENTAL PROCEDURES

Cloning

The four independent hNedd4L WW domains, the three Nedd4L WW domain pairs were prepared as described previously (Aragón et al., 2011; Chong et al., 2010). Point mutations were introduced using the QuickChange site directed mutagenesis Kit (Stratagene) with the appropriate complementary mutagenic primers. All wild-type and variants were confirmed by DNA sequencing.

Protein Expression and Purification

Unlabeled, ^{15}N -labeled, ^{13}C , ^{15}N and ^2H , ^{13}C , ^{15}N -labeled proteins were expressed in *Escherichia coli* BL21 (DE3), in LB or in minimal medium (M9),

prepared either in H₂O or in D₂O (99.89%, CortecNet) enriched with ¹⁵NH₄Cl and/or D-[¹³C] glucose as sole sources of carbon and nitrogen, respectively (Marley et al., 2001). Proteins were purified as described (Aragón et al., 2011). For the NMR experiments with Smurf1/2 WW domains we have minimized the aggregation tendency by concentrating the proteins (single WW2 or the WW1-WW2 pair and WW3 or WW2-WW3 for Smurf1 and 2, respectively) in the presence of the ligand.

Transfection, Immunoprecipitation, and Immunoblotting

Transfection of the indicated plasmids was performed as described previously (Gao et al., 2009) HEK293T cells were incubated as indicated with TGF-β1 (100 pM; R&D Systems). Immunoprecipitation and western immunoblotting were done as described (Sapkota et al., 2007).

Smad7 Peptide Synthesis and Purification

The peptide (Ac-ELESPPPPYSRYPM-D-NH₂ (203–217) was synthesized using Fmoc-solid phase peptide synthesis with a rink amide resin (Merck Chemicals), in a CEM Liberty1 microwave synthesizer (0.1 mmol scale). The acetylated peptide was purified by RP-HPLC using a SunFire C18 Sephasil preparative column (Waters) with an ÄKTApurifier10 (GE Healthcare Life Sciences), using a linear gradient of 10%–40% acetonitrile and 0.05% TFA and an elution time of 20 min. The peptide was analyzed by MALDI-TOF mass spectrometry and 2D homonuclear NMR spectroscopy.

NMR Assignment

NMR data were acquired at 285 K/298 K on a Bruker Avance III 600-MHz spectrometer equipped with a z-pulse field gradient unit. Backbone ¹H, ¹⁵N, and ¹³C resonance assignments were obtained by analyzing 3D CBCA(CO)NH and HNCBCA experiments. Side-chain resonance assignments were obtained by analyzing HCCC(CO)NH, ¹⁵N-TOCSY, HCCH-TOCSY and ¹⁵N-, ¹³C NOESY spectra (Sattler et al., 1999). Inter- and intra molecular proton distance restraints were obtained from peaks assigned in 2D-NOESY experiments. All spectra were processed with NMRPipe/NMRDraw software (Delaglio et al., 1995) and were analyzed with CARRA (Bartels et al., 1995). Spectra used for the calculation were integrated using the batch integration method of the XEASY package.

NMR Titration Experiments

¹⁵N-HSQC spectra were acquired using 300 μM ¹⁵N-labeled protein samples to which the unlabeled peptide was added stepwise until saturation was achieved.

Structure Determination and Refinement

Structures were calculated with CNS 1.1 (Brünger et al., 1998), using only unambiguously assigned restraints derived from NOESY experiments, coupling constants ³J(H^N, H^A from HNHA spectra) and hydrogen bonds measured from D₂O exchange experiments. The protocol for the calculation consists of two iterations of 1 and 200 structures, respectively, using 100,000 cooling steps. All calculated structures were water refined and ranked based on minimum values of energy and violations. The water refinement protocol is a modification of the original protocol provided with Aria (Nilges et al., 1997), which uses all experimental restraints during the refinement process. Analysis of the quality of the lowest energy structures was performed using PROCHECK-NMR (Laskowski et al., 1996) and the statistics are shown in Table 1. Images were generated with PyMOL (DeLano, 2002).

Isothermal Titration Calorimetry

ITC experiments were performed using a low volume nano ITC calorimeter (TA instruments) and five different temperatures 5, 15, 20, 25, and 30°C. Details of the experiments are given in the Supplemental Experimental Procedures.

Ion Mobility-Mass Spectrometry

Traveling wave ion mobility mass spectrometry experiments were performed on a Synapt G1 HDMS mass spectrometer (Waters, Manchester, UK). Experimental details are given in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

For the five complexes, the corresponding Protein Data Bank and BioMagResBank codes are, respectively, 2ltv/18498 (YAPWW1_S7), 2ltw/18499 (YAPWW2_S7), 2ltx/18500 (Smurf1WW2_S7), 2lty/18501 (Nedd4LWW2_S7), and 2ltz/18502 (Smurf2WW3_S7). Protein short names are given as in Table 1.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and Supplemental Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.str.2012.07.014>.

ACKNOWLEDGMENTS

We thank the Mass Spectrometry Core Facility at the IRB Barcelona (Dr. M. Vilaseca) for support with the IM-MS experiments. N.G. had an IRB PhD fellowship, and E.A. was financed in part by a Consolider RNAREG (CSD2009-00080) grant. This work was supported by Grant SAF2011-25119 (M.J.M.) and by NIH Grant R37-CA34610 (J.M.). J.M. is an Investigator of the Howard Hughes Medical Institute, and M.J.M. is an ICREA Programme Investigator.

Received: June 15, 2012

Revised: July 27, 2012

Accepted: July 27, 2012

Published online: August 23, 2012

REFERENCES

- Alarcón, C., Zaromytidou, A.I., Xi, Q., Gao, S., Yu, J., Fujisawa, S., Barlas, A., Miller, A.N., Manova-Todorova, K., Macias, M.J., et al. (2009). Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways. *Cell* 139, 757–769.
- Aragón, E., Goerner, N., Zaromytidou, A.I., Xi, Q., Escobedo, A., Massagué, J., and Macias, M.J. (2011). A Smad action turnover switch operated by WW domain readers of a phosphoserine code. *Genes Dev.* 25, 1275–1288.
- Bai, S., and Cao, X. (2002). A nuclear antagonistic mechanism of inhibitory Smads in transforming growth factor-beta signaling. *J. Biol. Chem.* 277, 4176–4182.
- Bartels, C., Xia, T.H., Billeter, M., Güntert, P., and Wüthrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* 5, 1–10.
- Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54, 905–921.
- Chen, H.I., and Sudol, M. (1995). The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc. Natl. Acad. Sci. USA* 92, 7819–7823.
- Chong, P.A., Lin, H., Wrana, J.L., and Forman-Kay, J.D. (2006). An expanded WW domain recognition motif revealed by the interaction between Smad7 and the E3 ubiquitin ligase Smurf2. *J. Biol. Chem.* 281, 17069–17075.
- Chong, P.A., Lin, H., Wrana, J.L., and Forman-Kay, J.D. (2010). Coupling of tandem Smad ubiquitination regulatory factor (Smurf) WW domains modulates target specificity. *Proc. Natl. Acad. Sci. USA* 107, 18404–18409.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293.
- DeLano, W.L. (2002). The PyMOL Molecular Graphics System. <http://www.pymol.org/>.
- Ebisawa, T., Fukuchi, M., Murakami, G., Chiba, T., Tanaka, K., Imamura, T., and Miyazono, K. (2001). Smurf1 interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation. *J. Biol. Chem.* 276, 12477–12480.

- Eichhorn, P.J., Rodón, L., González-Juncà, A., Dirac, A., Gili, M., Martínez-Sáez, E., Aura, C., Barba, I., Peg, V., Prat, A., et al. (2012). USP15 stabilizes TGF- β receptor I and promotes oncogenesis through the activation of TGF- β signaling in glioblastoma. *Nat. Med.* **18**, 429–435.
- Fan, G., Fan, Y., Gupta, N., Matsuura, I., Liu, F., Zhou, X.Z., Lu, K.P., and Gélinas, C. (2009). Peptidyl-prolyl isomerase Pin1 markedly enhances the oncogenic activity of the rel proteins in the nuclear factor-kappaB family. *Cancer Res.* **69**, 4589–4597.
- Ferrigno, O., Lallemand, F., Verrecchia, F., L'Hoste, S., Camonis, J., Atfi, A., and Mauviel, A. (2002). Yes-associated protein (YAP65) interacts with Smad7 and potentiates its inhibitory activity against TGF-beta/Smad signaling. *Oncogene* **21**, 4879–4884.
- Fuentealba, L.C., Eivers, E., Ikeda, A., Hurtado, C., Kuroda, H., Pera, E.M., and De Robertis, E.M. (2007). Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal. *Cell* **131**, 980–993.
- Gao, S., Alarcón, C., Sapkota, G., Rahman, S., Chen, P.Y., Goerner, N., Macias, M.J., Erdjument-Bromage, H., Tempst, P., and Massagué, J. (2009). Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF-beta signaling. *Mol. Cell* **36**, 457–468.
- Hata, A., Lagna, G., Massagué, J., and Hemmati-Brivanlou, A. (1998). Smad6 inhibits BMP/Smad1 signaling by specifically competing with the Smad4 tumor suppressor. *Genes Dev.* **12**, 186–197.
- Hayashi, H., Abdollah, S., Qiu, Y., Cai, J., Xu, Y.Y., Grinnell, B.W., Richardson, M.A., Topper, J.N., Gimbrone, M.A., Jr., Wrana, J.L., and Falb, D. (1997). The MAD-related protein Smad7 associates with the TGFbeta receptor and functions as an antagonist of TGFbeta signaling. *Cell* **89**, 1165–1173.
- He, W., Li, A.G., Wang, D., Han, S., Zheng, B., Goumans, M.J., Ten Dijke, P., and Wang, X.J. (2002). Overexpression of Smad7 results in severe pathological alterations in multiple epithelial tissues. *EMBO J.* **21**, 2580–2590.
- Kanelis, V., Rotin, D., and Forman-Kay, J.D. (2001). Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nat. Struct. Biol.* **8**, 407–412.
- Kanelis, V., Bruce, M.C., Skrynnikov, N.R., Rotin, D., and Forman-Kay, J.D. (2006). Structural determinants for high-affinity binding in a Nedd4 WW3* domain-Comm PY motif complex. *Structure* **14**, 543–553.
- Kavsak, P., Rasmussen, R.K., Causing, C.G., Bonni, S., Zhu, H., Thomsen, G.H., and Wrana, J.L. (2000). Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. *Mol. Cell* **6**, 1365–1375.
- Kuratomi, G., Komuro, A., Goto, K., Shinozaki, M., Miyazawa, K., Miyazono, K., and Imamura, T. (2005). NEDD4-2 negatively regulates TGF-beta signaling by inducing ubiquitin-mediated degradation of Smad2 and TGF-beta type I receptor. *Biochem. J.* **386**, 461–470.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486.
- Macias, M.J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* **382**, 646–649.
- Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–75.
- Massagué, J. (1998). TGF-beta signal transduction. *Annu. Rev. Biochem.* **67**, 753–791.
- Matsuura, I., Chiang, K.N., Lai, C.Y., He, D., Wang, G., Ramkumar, R., Uchida, T., Ryo, A., Lu, K., and Liu, F. (2010). Pin1 promotes transforming growth factor-beta-induced migration and invasion. *J. Biol. Chem.* **285**, 1754–1764.
- Morales, B., Ramirez-Espain, X., Shaw, A.Z., Martin-Malpartida, P., Yraola, F., Sánchez-Tilló, E., Farrera, C., Celada, A., Royo, M., and Macias, M.J. (2007). NMR structural studies of the ItchWW3 domain reveal that phosphorylation at T30 inhibits the interaction with PPxY-containing ligands. *Structure* **15**, 473–483.
- Nakao, A., Afrakhte, M., Morén, A., Nakayama, T., Christian, J.L., Heuchel, R., Itoh, S., Kawabata, M., Heldin, N.E., Heldin, C.H., and ten Dijke, P. (1997). Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signaling. *Nature* **389**, 631–635.
- Nilges, M., Macias, M.J., O'Donoghue, S.I., and Oschkinat, H. (1997). Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J. Mol. Biol.* **269**, 408–422.
- Ohnishi, S., Güntert, P., Koshiba, S., Tomizawa, T., Akasaka, R., Tochio, N., Sato, M., Inoue, M., Harada, T., Watanabe, S., et al. (2007). Solution structure of an atypical WW domain in a novel beta-clam-like dimeric form. *FEBS Lett.* **581**, 462–468.
- Pires, J.R., Taha-Nejad, F., Toepert, F., Ast, T., Hoffmüller, U., Schneider-Mergener, J., Kühne, R., Macias, M.J., and Oschkinat, H. (2001). Solution structures of the YAP65 WW domain and the variant L30 K in complex with the peptides GTPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. *J. Mol. Biol.* **314**, 1147–1156.
- Ramirez-Espain, X., Ruiz, L., Martin-Malpartida, P., Oschkinat, H., and Macias, M.J. (2007). Structural characterization of a new binding motif and a novel binding mode in group 2 WW domains. *J. Mol. Biol.* **373**, 1255–1268.
- Sapkota, G., Alarcón, C., Spagnoli, F.M., Brivanlou, A.H., and Massagué, J. (2007). Balancing BMP signaling through integrated inputs into the Smad1 linker. *Mol. Cell* **25**, 441–454.
- Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158.
- Shi, Y., and Massagué, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685–700.
- Sudol, M. (1994). Yes-associated protein (YAP65) is a proline-rich phosphoprotein that binds to the SH3 domain of the Yes proto-oncogene product. *Oncogene* **9**, 2145–2152.
- Topper, J.N., DiChiara, M.R., Brown, J.D., Williams, A.J., Falb, D., Collins, T., and Gimbrone, M.A., Jr. (1998). CREB binding protein is a required coactivator for Smad-dependent, transforming growth factor beta transcriptional responses in endothelial cells. *Proc. Natl. Acad. Sci. USA* **95**, 9506–9511.
- Ulloa, L., Doody, J., and Massagué, J. (1999). Inhibition of transforming growth factor-beta/SMAD signalling by the interferon-gamma/STAT pathway. *Nature* **397**, 710–713.
- Wan, L., Zou, W., Gao, D., Inuzuka, H., Fukushima, H., Berg, A.H., Drapp, R., Shaik, S., Hu, D., Lester, C., et al. (2011). Cdh1 regulates osteoblast function through an APC/C-independent modulation of Smurf1. *Mol. Cell* **44**, 721–733.
- Wiesner, S., Ogunjimi, A.A., Wang, H.-R., Rotin, D., Sicheri, F., Wrana, J.L., and Forman-Kay, J.D. (2007). Autoinhibition of the HECT-type ubiquitin ligase Smurf2 through its C2 domain. *Cell* **130**, 651–662.
- Yan, X., and Chen, Y.G. (2011). Smad7: not only a regulator, but also a cross-talk mediator of TGF- β signalling. *Biochem. J.* **434**, 1–10.
- Yan, X., Lin, Z., Chen, F., Zhao, X., Chen, H., Ning, Y., and Chen, Y.G. (2009). Human BAMBI cooperates with Smad7 to inhibit transforming growth factor-beta signaling. *J. Biol. Chem.* **284**, 30097–30104.

Resumen en Castellano

9. Resumen en castellano:

9.1 Introducción general

El destino final de las células no está abandonado al libre albedrío sino que está controlado por una multitud de señales reguladoras que necesitan funcionar a la perfección para no acarrear consecuencias devastadoras para los organismos vivos. Uno de los responsables clave en esta red de señales lo forman la familia de citoquinas llamadas TGF- β . Estas hormonas desencadenan una inmensa cantidad de respuestas gracias al control que ejercen sobre la familia de factores de transcripción Smad (SMA Mad) y sobre una multitud de proteínas que interaccionan con ellas. De forma simplificada, el modo de actuación de la cascada TGF- β consiste en enviar a los factores de transcripción Smads desde el citoplasma al núcleo, donde participan en numerosos procesos, algunos de los cuales incluyen el mantenimiento y control de la pluripotencia de las células madre, su diferenciación, el desarrollo embrionario, la regeneración de tejidos, y la homeostasis del tejido diferenciado (Massagué, 1998).

Según su función, las proteínas Smad se clasifican en Smads regulados por el receptor (R-Smads), que incluyen Smads 1, 5 y 8 en la versión regulada por BMP, y Smads 2 y 3 en las vías reguladas por TGF- β / nodal / y activina respectivamente. Los R-Smads forman complejos con el co-activador Smad (Co-Smad ó Smad4). La familia Smad también contiene los dos Smads inhibidores (I-Smads), Smad6 y Smad7, que proporcionan la regulación negativa necesaria para el autocontrol de estas vías tan poderosas y ubicuas en casi todos los tipos de células.

Todas las proteínas Smad son modulares (Shi y Massagué, 2003). Los R-Smads y el Co-Smad contienen dos dominios Mad denominados MH1 y MH2 y conectados por una secuencia que funciona de sitio de unión para otras proteínas, las cuales modulan la función que se necesite realizar. Esta secuencia contiene un clúster conservado de sitios de fosforilación adyacentes a un motivo rico en prolinas y que denominamos PY. Los dominios MH1 de R-Smads y Smad4 unen ADN, mientras que el dominio MH2 y la secuencia conectora entre los dos dominios funcionan de sitios de interacción con receptores, proteínas reguladoras, y cofactores. Gracias a la interacción con las otras proteínas que se unen, existe un control sobre el resultado final de la señal que se necesita transmitir (Shi y Massagué, 2003). En comparación con los R-Smads y Co-Smads, los I-Smads tienen baja similitud de secuencia en el dominio MH1 pero conservan en común con los R-Smads un dominio MH2 y el conector que contiene el motivo PY de unión a otras proteínas. La presencia de las regiones comunes facilita la competición de R-Smads y I-Smads por el receptor y también hacia los ligandos facilitando así el papel inhibitorio de los I-Smads (Figura 1). Los I-Smads se expresan además en respuesta a las señales de TGF- β o BMP, para proporcionar la retroalimentación negativa de la vía de forma natural (Bai y Cao, 2002;. Hata et al, 1998; He et al, 2002;. Kavsak et al, 2000;. Nakao et al, 1997;. Yan y Chen, 2011) y en respuesta a otras vías como STAT que se oponen a la señalización de TGF- β (Ulloa et al, 1999). De los dos Smads inhibidores, Smad6 interfiere con la formación de complejos de Smad1 con Smad4 (Hata et al, 1998.) mientras que Smad7 interfiere con la formación de los complejos entre los R-Smads-Smad4 e inhibe además la transferencia de la señal de activación de los receptores de TGF- β y de BMP (Hayashi et al, 1997;. Topper y col ., 1998). La composición modular de las proteínas Smad se muestra en la Figura 1.

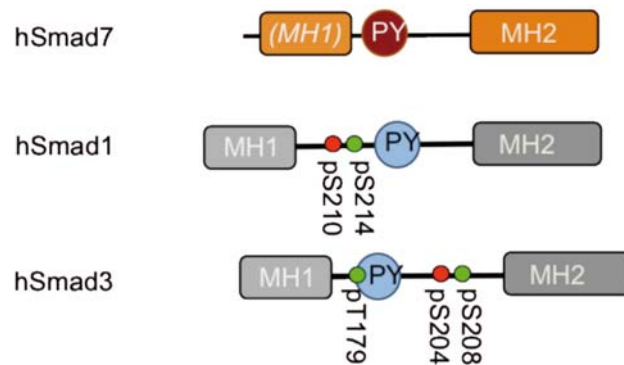


Figura 1: Proteínas Smad y los sitios de fosforilación e interacción en su zona central. La secuencia presente en MH1 Smad7 se muestra en cursiva sólo para remarcar su alto grado de divergencia.

Varias fosforilaciones clave impulsan el proceso de la señalización mediante las proteínas Smads. El mecanismo comienza con la interacción de las citoquinas BMP y TGF- β con sus respectivos receptores que se activan fosforilando primero al receptor y después a los Smads en el extremo C-terminal. El extremo C-terminal fosforilado proporciona un sitio de unión para Smad4, que es un componente esencial en la formación específica de los complejos transcripcionales. Estas fosforilaciones están contrarestandas por la actuación de fosfatasas que limitan la cantidad disponible de moléculas Smad activadas (Inman et al, 2002;. Lin et al, 2006;. Schmierer et al, 2008;.. Xu et al, 2002).

Las proteínas activadas Smads -que se asocian con Smad4 y se unen además a sus genes diana- sufren una segunda serie de fosforilaciones, estas catalizadas por las quinasas CDK8 y CDK9 dependientes de ciclinas (Alarcón et al, 2009; Gao et al, 2009) y después por la quinasa glucógeno sintasa-3 (GSK3) (Alarcón et al, 2009; Fuentealba et al, 2007; Sapkota et al, 2007). Las CDK 8 y 9 son parte de del sistema *mediador transcripcional* y forman complejos de

elongación, respectivamente (Durand et al, 2005; Komarnitsky et al, 2000; Malik & Roeder, 2000). GSK3 es una quinasa dependiente de PI3K y regulada-Wnt (Cohen & Marco, 2001, Wu & Pan). La fosforilación de los R-Smads mediante las quinasas CDK8/9 sirve como nucleación para la siguiente ronda de fosforilación mediante la quinasa GSK3. Todos estos sitios de fosforilación se agrupan en una región de unión entre dominios MH1 y MH2 y así permiten la activación máxima de Smads además de señalar también las proteínas para su poli-ubiquitinación y degradación mediada por el proteasoma (Alarcón et al, 2009; Gao et al, 2009). En un contexto celular diferente, esta región de los Smads es fosforilada por quinasas MAP y CDKs activadas durante la división celular en respuesta a *mitógenos* y *stress* con el objetivo de limitar el alcance de la señalización mediante TGF- β y BMP (Kretschmar et al, 1997;. Kretschmar et al, 1999;. Matsuura et al., 2009).

Se sabe por la bibliografía y por nuestros propios resultados que al menos cuatro proteínas unen específicamente a la region conectora de los R-Smads activados. Entre estas proteínas están la ubiquitina ligasa Smurf1 (Sapkota et al., 2007) y YAP, el efector transcripcional de la vía Hippo, que unen específicamente a los Smad1 / 5 (Alarcón et al., 2009), mientras que la ubiquitín ligasa NEDD4L (Gao et al. , 2009) y la peptidil-prolyl cis / trans isomerasa Pin1 (Matsuura et al., 2009) unen a los Smad2 / 3. YAP coopera con Smad1 para activar genes que suprimen la diferenciación neural en células troncales embrionarias de ratón en respuesta a las señales BMP (Alarcón et al., 2009). Pin1 coopera con Smad2 / 3 para estimular la migración de células cancerosas en respuesta a TGF- β (Matsuura et al., 2009). Por otro lado, Smurf1 y NEDD4L unen a sus targets respectivos con el objeto de marcarlos para su degradación, (Figura 2).

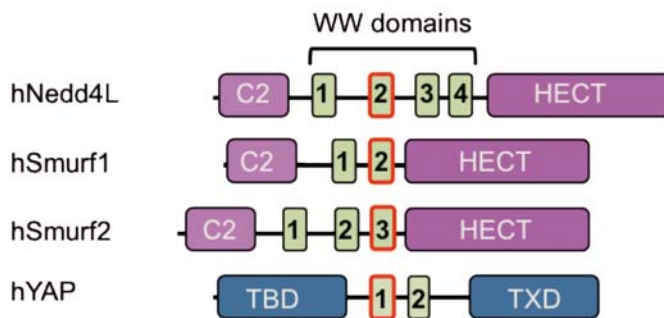


Figura 2: Composición modular de proteínas de unión a Smad

Un análisis detallado de estas proteínas de unión a Smads revela que tienen en común la presencia de dominios WW, un en Pin1, dos en Smurf1 y YAP, y cuatro en NEDD4L. Los dominios WW son unidades estructurales y funcionales de unos 38-40 aminoácidos, que se caracterizan por la presencia de dos triptófanos conservados en su secuencia y por adoptar una estructura de triple hebra antiparalela. Estos dominios unen típicamente secuencias ricas en prolina (PPxY por ejemplo, o "PY") o en el caso del dominio presente en Pin1, secuencias que incluyen residuos fosforilados en Ser o Thr, seguidos de prolina (pS-/pT-P) (Macias et al., 2002). Las proteínas R-Smad contienen un motivo PY situado cerca de los sitios de fosforilación en CDK/GSK3 en la región que conecta los dominios MH1 y MH2. La unión de dominios WW y motivos PY han sido ampliamente descritas en la literatura (Macias et al 2002, (Macias et al, 1996; Pires et al, 2001; Toepert et al, 2001) lo que sugiere que el sitio PY en Smads y los dominios WW presentes en YAP, Pin1 y en las ligasas de ubiquitina pueden ser responsable de las interacciones. Estas líneas de evidencia presentan un escenario en el que las diferentes quinasas fosforilan las Smads en el núcleo para así crear sitios de unión para que cofactores transcripcionales y ligasas de ubiquitina compitan por lograr unir a los Smads. El resultado de estas interacciones gobernará la función final de los Smads, resultando clave para la transducción eficaz de las señales iniciada por las hormonas BMP y TGF- β . Sin embargo, la convergencia de la activación y

degradación en un conjunto agrupado de modificaciones en los Smads plantea interrogantes acerca de cómo consiguen los Smads actuar antes de ser degradados.

La necesidad de reconocer el motivo de pSP de Smad1 que precede al sitio PY por Smurf1 y YAP1 plantea otra cuestión: cómo una secuencia tan larga de Smad1 puede ser reconocida por un único dominio WW?. Una hipótesis atractiva que podría responder a esta cuestión es que en vez de usar un único dominio WW, en este caso funcionen de forma sincronizada el par de dominios WW presentes en cada Smurf1 y YAP1. Una forma sería participando en la interacción utilizando quizás uno para reconocer el sitio PY y actuando el otro dominio para reconocer el sitio de unión a pSP. Aunque tal disposición donde ambos dominios unan a una diana de una manera sincronizada podría explicar los datos experimentales, tal disposición nunca ha sido observada con anterioridad.

Por otro lado, hasta ahora sólo el dominio Pin1WW ha sido descrito como el único dominio WW capaz de reconocer secuencias fosforiladas (con Ser / Thr fosforiladas) seguidas de prolina (Verdecia et al 2000). La comparación de las secuencias del dominio Pin1 y los dominios de YAP y Smurf1 indica que los residuos de Pin1WW descritos como responsables del reconocimiento de los fosfatos no están presentes en ninguna de las secuencias de los dominios WW de YAP, y Smurf1. Por lo tanto, desde este punto de vista, parece que otros dominios WW diferentes de Pin1 han encontrado nuevas soluciones para el reconocimiento de los residuos fosforilados, probablemente una característica crítica para su especificidad *in vivo*. Para terminar de definir la complejidad de este escenario nuestros colaboradores han encontrado que aunque las ubiquitín ligasas Smurf1 / 2 y NEDD4L pertenecen todas a la misma familia de HECT E3 ubiquitín ligasas y además unen R-Smads, su funcionalidad no es indiscriminada. Mientras que Smurf1/2 se han especializado en Smad1 y

Smad5, NEDD4L une a Smad2/3 tanto *in vitro* como *in vivo* (Gao S et al 2009), observación que plantea la pregunta de cómo se consigue esta especificidad.

Además de las cuestiones de cómo YAP, Pin1 y las ubiquitín ligasas del tipo E3 HECT leen este código escrito en los R-Smads fosforilados en la región central, recientes trabajos publicados en la literatura han puesto de manifiesto que el inhibidor Smad7 es un eje central de la regulación negativa de los receptores activados TGF- β y BMP y que lo consigue en parte mediante la interacción con algunas proteínas comunes con los R-Smads (Yan & Chen, 2011). De hecho se conoce que Smad7 recluta las ubiquitín ligasas NEDD4L, Smurf1 y Smurf2 y las traslada a las inmediaciones de los receptores para iniciar el proceso de polyubiquitination del receptor y así comenzar el proceso de endocitosis degradativa (Ebisawa et al, 2001;. Kavsak et al, 2000;.. Kuratomi et al, 2005). Smad7 también se une a YAP (Ferrigno et al., 2002), lo que proporciona un mecanismo para la retención de este mediador y así regular la señalización de BMP (Alarcón et al., 2009). Como con Smad1 / 2 y 3, todas estas interacciones implican la región de Smad7 que contiene un motivo PY y la región con dominios WW de NEDD4L, Smurf1 / 2, y YAP.

Impulsados por estas observaciones nos propusimos caracterizar cómo las interacciones entre las proteínas Smad y sus ligandos ocurren y sobre todo cómo estos dominios WW tan similares en secuencia pueden discernir entre los diferentes Smads y optar por unas u otras en un momento dado de la vida de la célula. Para lograr estos ambiciosos objetivos, el laboratorio en el que he hecho este trabajo estableció una colaboración muy fructífera con el del Dr. Joan Massagué*. En nuestro laboratorio caraterizamos las interacciones entre

*A collaboration with the group of Dr. J. Massagué, Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA.

los pares de dominios WW presentes en YAP y en Smurf1 y varios péptidos de Smad1, y también entre los dominios WW de Pin1 y NEDD4L con Smad3, a nivel atómico mientras que nuestros colaboradores se centraron en descubrir la escenarios biológicos donde estas interacciones ocurren.

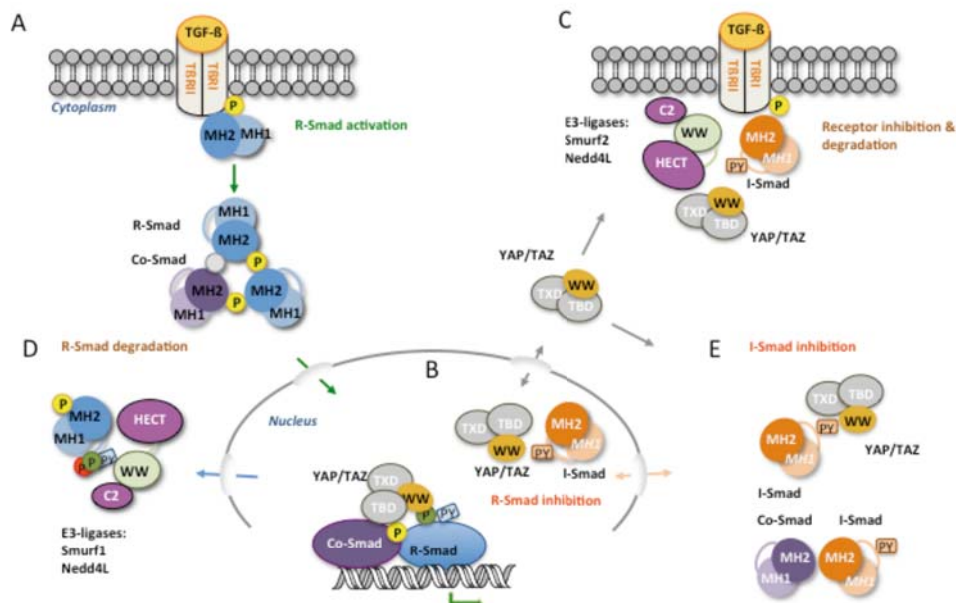
9.2 Objetivos de la presente Tesis

Objetivo 1. En el curso de su actuación en el núcleo, las proteínas Smad sufren procesos de fosforilación que a la vez que les permiten alcanzar el máximo de la actividad transcripcional, los señalan para su destrucción. Estas observaciones presentan a la vez una paradoja, y una oportunidad para controlar el proceso de transmisión de señales pues pudiese existir un mecanismo que garantice la secuencia ordenada de los acontecimientos en este proceso de modo que primero se garantice la posibilidad de favorecer la acción y después ocurra la destrucción de los Smads. Motivados por estas hipótesis y utilizando el potencial de un enfoque funcional y estructural combinado, nos propusimos en este **objetivo 1** *caracterizar el papel de los sitios de fosforilación del Smad1 en la interacción con YAP y Smurf1 y describir con detalle estructural como las proteínas que contienen dominios WW semejantes pueden discernir entre las secuencias objetivo y elegir una en particular.*

Objetivo 2. Smad7 es un inhibidor que actúa como un regulador negativo de la señalización por (TGF- β). La expresión de Smad7 está inducida por TGF- β , a través de la cascada de los R-Smads. Smad7 interactúa con el receptor TGF- β de tipo I (T β R-I), e interfiere con la interacción de los R-Smads inhibiendo su fosforilación. I-Smads además compite con los R-Smads porque contiene un motivo rico en prolina similar al de Smad 1 2 y Smad 3. Con anterioridad a este trabajo ya se había descrito que las ubiquitín ligasas Smurf1 y Smurf2

participan en la degradación de Smad 7 y también en la degradación del receptor de TGF- β . Con el fin de definir las bases estructurales que definen las interacciones hemos planteado en este **objetivo 2** *identificar los dominios WW de YAP, Smurf1 y Smurf2, que participan en el reconocimiento de Smad7, caracterizarlos estructuralmente y comparar las estructuras con las zonas que interaccionan con Smad1 resueltas en el objetivo 1 y así describir la base estructural que gobierna el control de la competencia entre la activación, la inhibición y la degradación de los diferentes Smads.*

Representación resumen del proceso de transmisión de señales desde el receptor al núcleo:



A. Tras la unión de la hormona TGF- β en el receptor, este forma un heterodímero y se activa. La activación es un mecanismo de dos etapas que comienza con la activación del receptor T β II, que es una quinasa que fosforila

a su vez el receptor I. Entonces, los R-Smads se acercan a la proximidad de la membrana y se fosforilan en dos Serinas conservadas en todos los R-Smads en su extremo C-terminal. La presencia de ambos fosfatos crea un sitio de unión para el Smad4. El resultado de este proceso es la formación de un complejo heterotrimérico que está listo para entrar en el núcleo.

B. Una vez en el núcleo, el complejo es fosforilado de nuevo esta vez en el la secuencia que conecta los dominios MH de los R-Smads. Después de la fosforilación este hetero-trímero está listo para unirse a cofactores (como YAP / TAZ) aumentando la especificidad de los Smads por las regiones diana en el ADN y así empezar a transcribir los genes específicos. Entre estos genes está el de Smad7, uno de los inhibidores de la cascada de señalización.

C. Smad7 regula la señalización por TGF- β a través de varios procesos. Entre ellos está el de ocupar el sitio de unión R-Smad en el receptor, lo que impide la activación de R-Smads y su transferencia al núcleo.

D. Después de haber participado en transcripción, los Smads sufren una segunda ronda de fosforilaciones que los marcan para degradación.

E. Smad7 también puede interactuar con Smad4 y así secuestrar esta proteína en el citoplasma. Smad7 también puede interactuar con YAP / TAZ y reducir la cantidad disponible de estas proteínas, lo que afecta a la formación de los diferentes complejos con los factores de transcripción y reduce la eficacia de los Smads como factores de transcripción.

9.3 Resultados

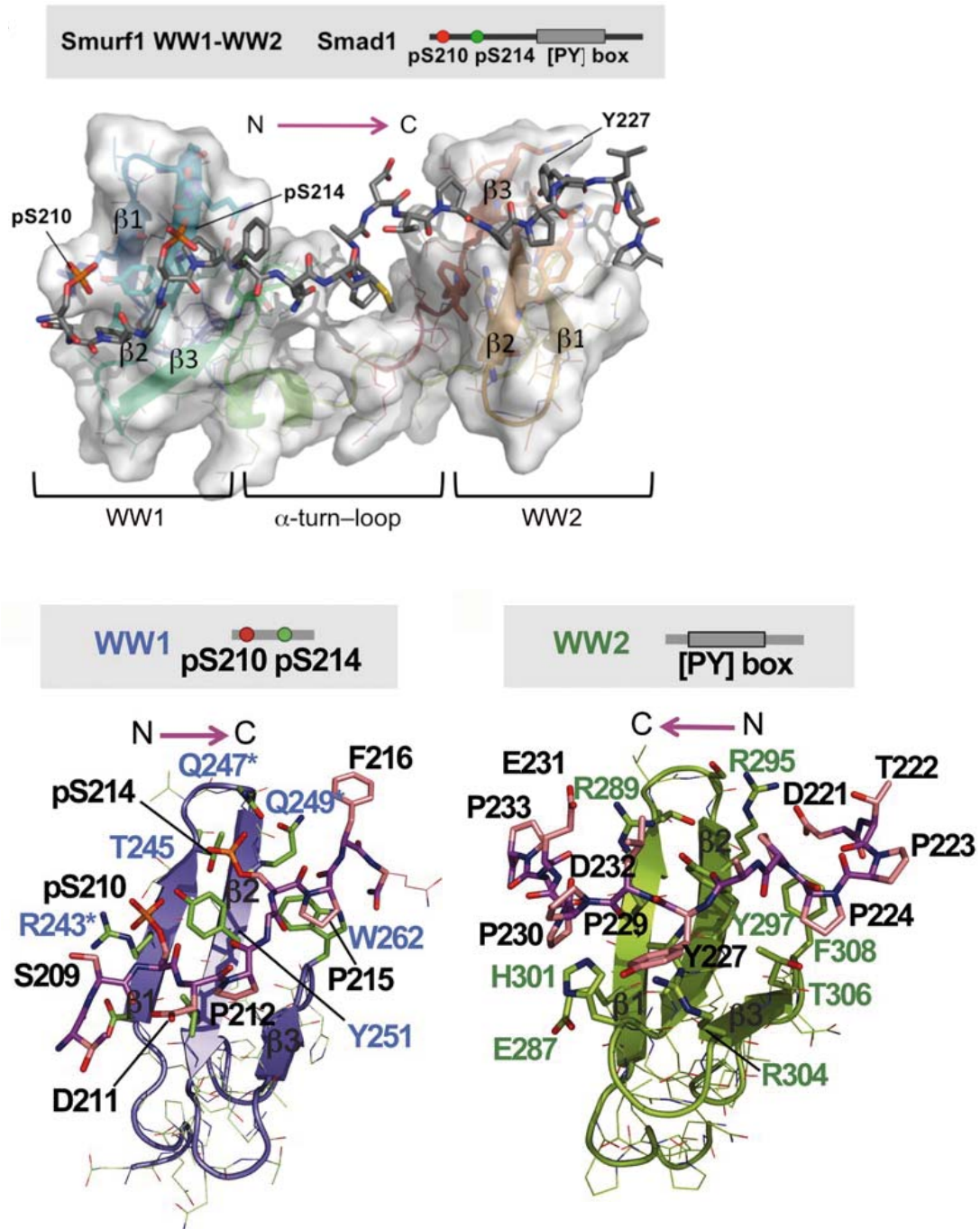
Los resultados de este trabajo se han publicado en dos artículos. En el manuscrito de la Tesis se incluyen copias de cada uno de los manuscritos publicados. Todas las estructuras determinadas así como las asignaciones de resonancia están depositadas en las bases de datos públicas PDB y Bio Mag Res.

Interacción entre el linker de Smad1 y la ubiquitín ligasa humana Smurf1:

Para caracterizar los contactos entre el par de dominios WW₁-WW₂ (residuos 232-314) de la ubiquitín ligasa y el segmento del linker de Smad1 (208-233) di-fosforilado en las posiciones S210 and S214 se utilizaron diferentes muestras de proteína con marcajes específicos (sobrexpresadas en bacteria) y varios péptidos preparados por síntesis. Los análisis de los datos de resonancia magnética nuclear revelaron que el primer dominio WW es el responsable de la interacción con los residuos fosforilados mientras que el segundo dominio WW reconoce el extremo C-terminal del linker de Smad1, que contiene un motivo PY característico de reconocimiento de dominios WW.

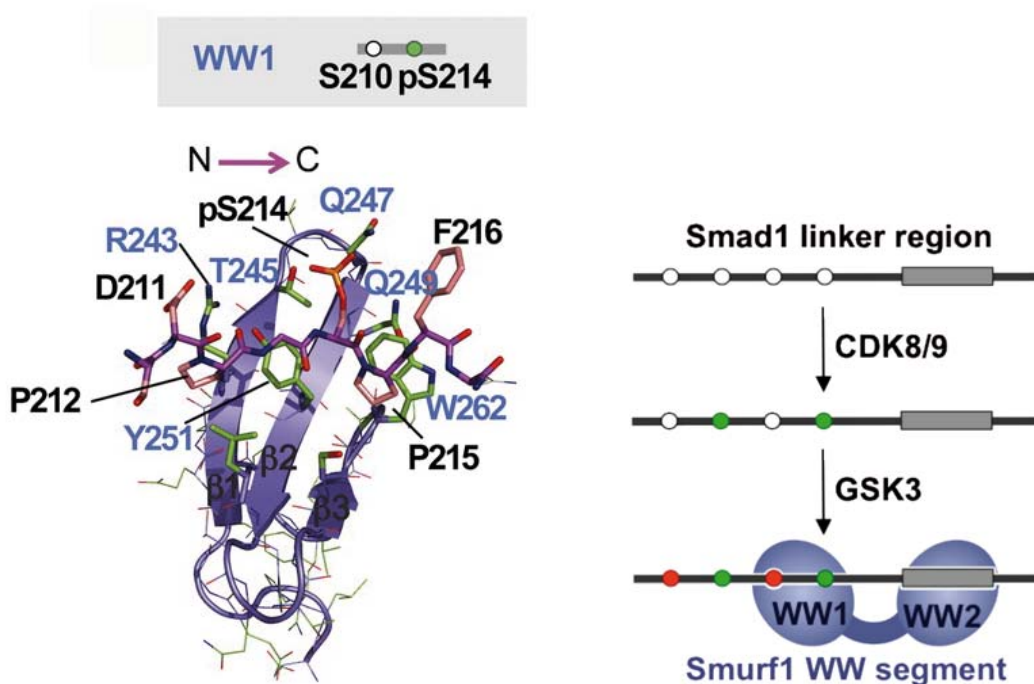
En la figura que corresponde al complejo, la proteína se ha representado como una superficie semitransparente y los elementos de estructura secundaria se han resaltado para que se vean a su través. El péptido se ha representado con todos los aminoácidos en gris. Debajo del complejo global se representa con detalle las interacciones entre el dominio WW₁ (figura de la izquierda) y del WW₂ (a la derecha). En esta representación se destacan los elementos de estructura secundaria de los dos dominios (en azul y verde respectivamente) mientras que los fragmentos del péptido reconocido por cada uno de ellos se han representado el rosa palo. Los diferentes aminoácidos que participan en

las interacciones están representados con cadenas laterales y etiquetados con el número que corresponde a la secuencia de proteína.



En nuestros estudios por calorimetría encontramos que Smurf1 puede reconocer también a Smad1 cuando está únicamente fosforilado en la posición

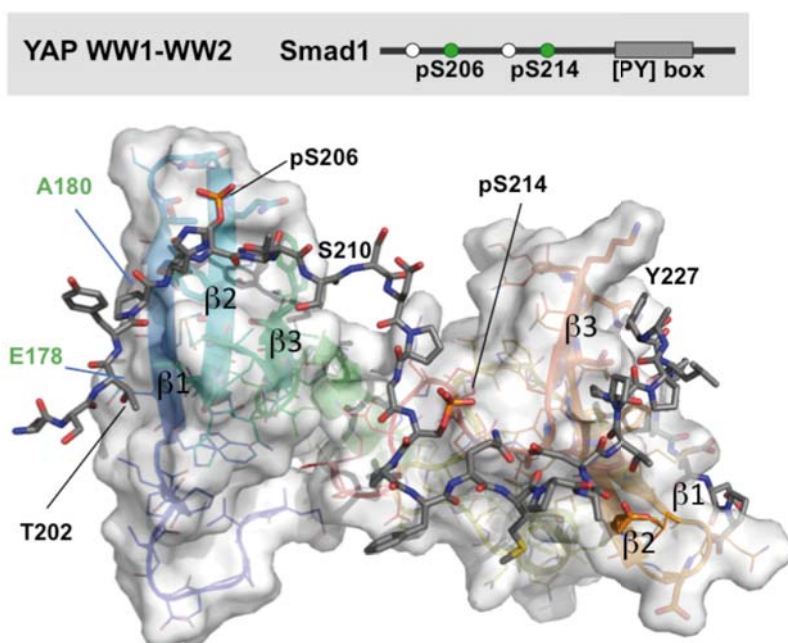
Ser214, aunque la interacción es de menor afinidad que para el motivo con doble fosforilación. Con el objeto de comprender las razones estructurales que explican la menor afinidad detectadas en el fragmento monofosforilado se determinó también la estructura con el fragmento de Smad1 que incluye una única fosforilación. En este complejo (en la figura se destaca solo la parte que reconoce el sitio de fosforilación) se observa que el número de interacciones entre el péptido y la proteína es menor que en el caso de la interacción con la misma secuencia y dos fosforilaciones. Esta disminución de contactos explica la reducción en afinidad. En conclusión, la interacción entre Smurf1 y Smad1 requiere que los sitios de fosforilación generados por las quinasas CDK8/9 y GSK3 estén presentes.



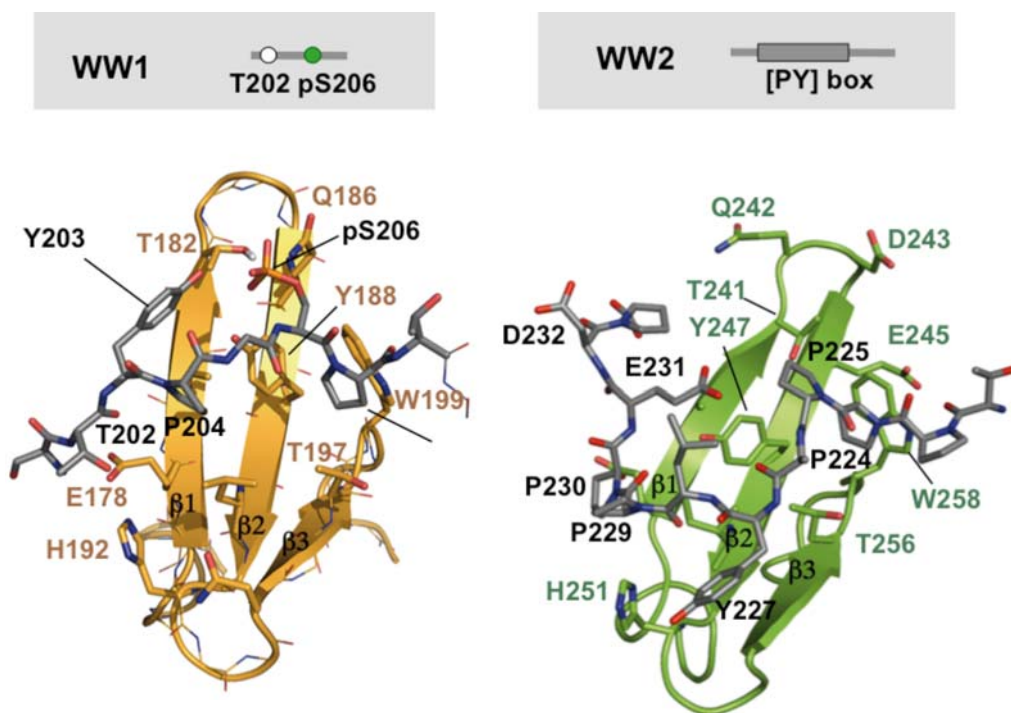
Interacción entre el linker de Smad1 y la el cofactor de transcripción YAP:

Según los datos bioquímicos y celulares obtenidos por nuestros colaboradores y cuantificadas in vitro mediante calorimetría y resonancia magnética nuclear, el coactivador YAP selecciona los sitios de fosforilación generados por las quinasas CDK8/9 en detrimento de las secuencias doblemente fosforiladas preferidas por la ubiquitín ligasa Smurf1. Para entender las razones conformacionales de estas preferencias investigamos la interacción de Smad1 fosforilado en las posiciones 206 y 214 (generadas solo por CDK8/9) que correspondía a la de mayor afinidad obtenida por calorimetría y la construcción de YAP que incluyen el par de dominios WW1-Ww2.

La estructura del complejo revela que el primer dominio WW reconoce la fosforilación en la posición 206 mientras que la fosforilación en 214 induce la presencia de un giro en el ligando, lo que favorece su interacción con el linker que conecta los dos dominios WWs. Finalmente, el segundo dominio WW es el responsable de la interacción con el motivo PY.

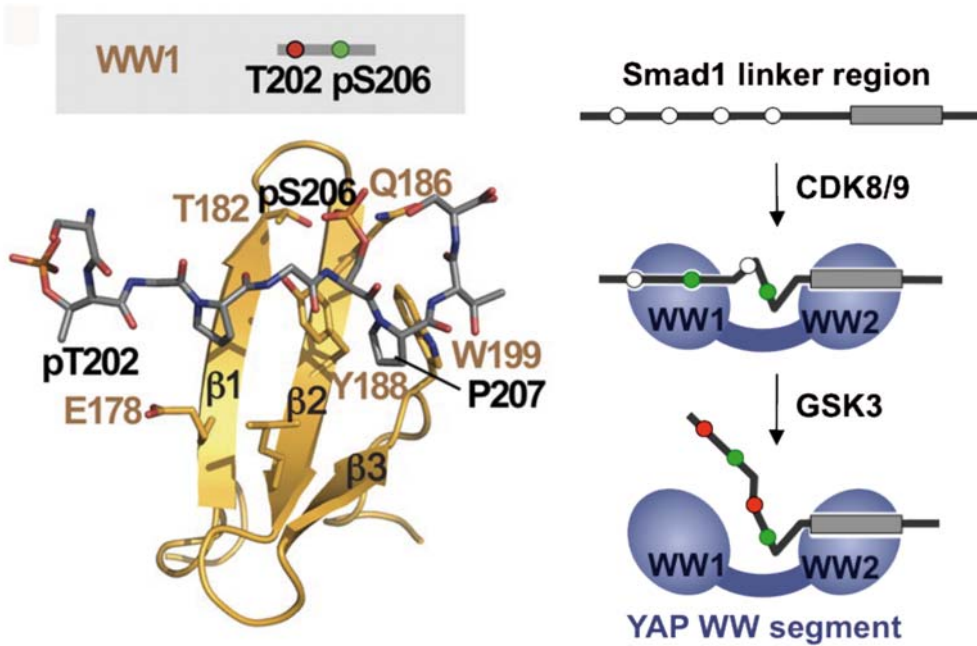


La interacción de YAP con el péptido de Smad1 doble fosforilado se muestra con una representación de superficie semitransparente para la proteína y de enlaces para los aminoácidos del péptido. Para facilitar la identificación de los contactos se muestran en amarillo y verde los dominios WW1 y WW2 respectivamente, con los aminoácidos que participan en la interacción etiquetados.



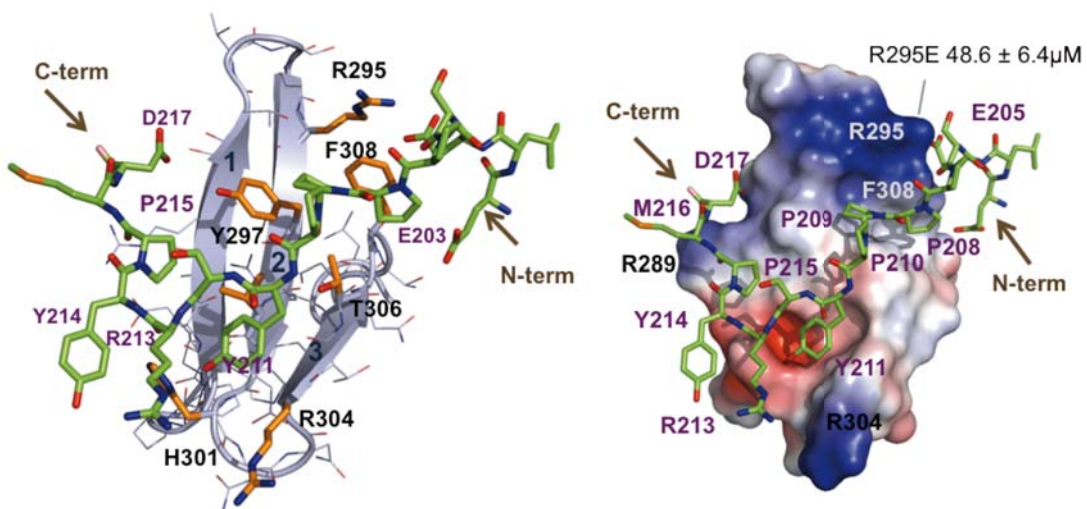
También investigamos la interacción con el motivo triplemente fosforilado pT202, pS206 y pS214, con el objeto de clarificar las razones por las que la fosforilación en la posición T202 reducen la afinidad de la interacción. Tal y como se destaca en la ampliación de la estructura, se puede ver que la presencia del tercer grupo fosfato reduce el número de contactos entre el péptido y el primer dominio, justificando así el papel desfavorable de esta fosforilación.

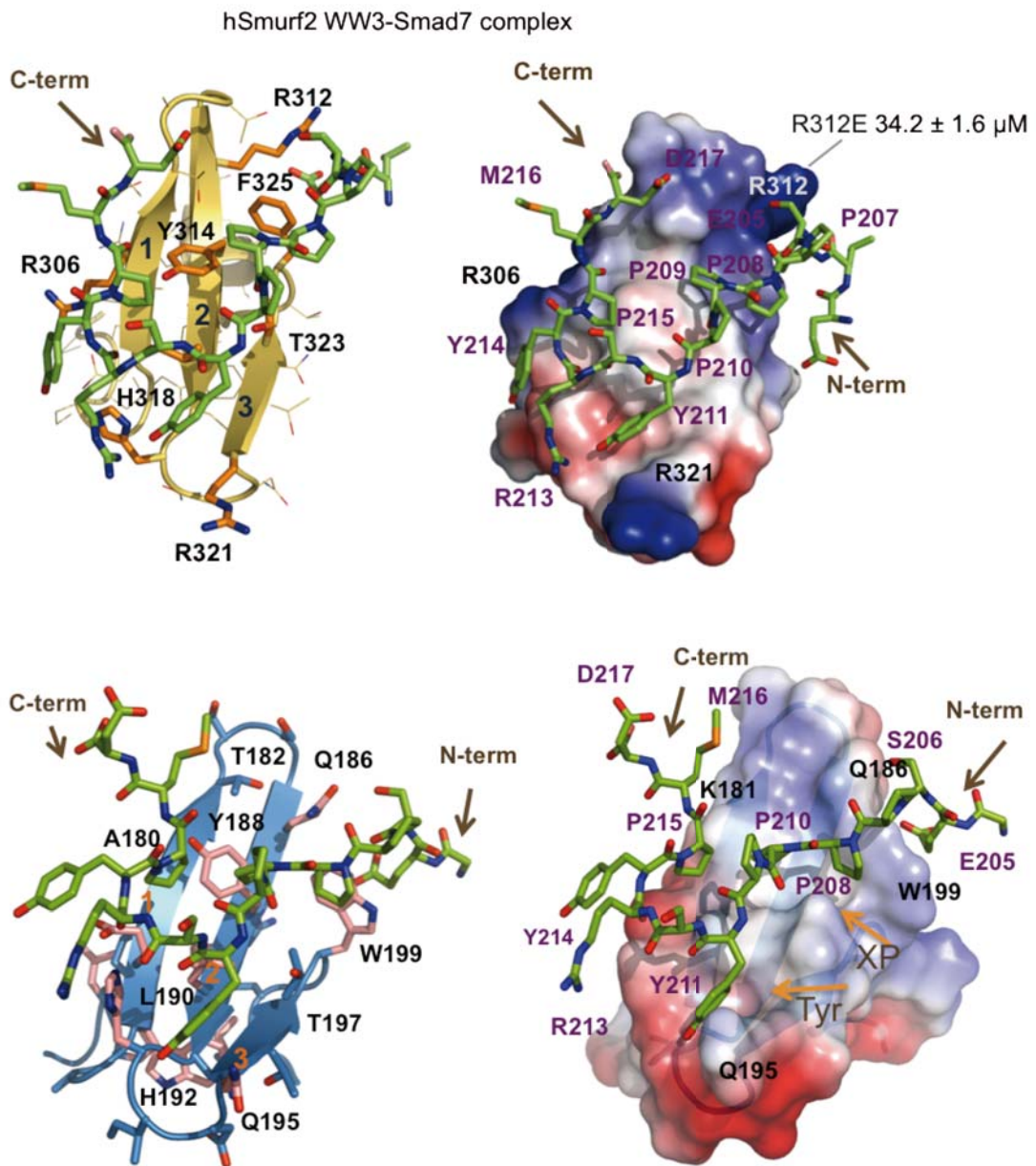
En el esquema de la derecha representamos de forma simplificada la preferencia de los dominios WW de YAP por la secuencia de Smad1 cuando contiene los sitios de fosforilación de CDK8/9 y como al añadir GSK3 nuevas fosforilaciones se dificulta la interacción.



Interacción entre el linker de Smad7 y las ubiquitín ligasas Smurf1 y Smurf2 y YAP.

Nuestros resultados revelan que las interacciones de las ligasas Smurf1 y Smurf2 con Smad7 implican a un único dominio WW (el WW2 y el WW3 respectivamente) y un único sitio de unión que es independiente de fosforilación, como se muestra en las figuras inferiores, donde se han etiquetado los residuos críticos implicados en los contactos. Nuestro trabajo también muestra que YAP requiere principalmente el dominio WW1 para la unión con la región de Smad7 que contiene el motivo PY.





9.4 Discusión

Desde la identificación de la familia de citoquinas TGF- β a finales del siglo pasado, y tras más de cincuenta mil publicaciones aún no se ha podido descifrar en su totalidad el misterio de cómo y porqué las consecuencias de la señalización mediante estas hormonas son tan impredecibles. En efecto, la hormona TGF- β aparentemente puede desencadenar, -con similar capacidad-, una función dada y la contraria (puede inhibir la proliferación celular y promover el crecimiento celular o aumentar la pluripotencia y la diferenciación celular, funciones descritas recientemente en (Massagué, 2012)). Estas paradojas aparentes son causadas por el amplio espectro de señales que se inician por la hormona TGF- β , y que son esenciales para el correcto funcionamiento de los organismos multicelulares. Una de las redes mejor estudiadas implican la familia de factores de transcripción Smads, que actúan como mediadores en la transmisión de la señal creada por la interacción de la hormona TGF- β con su receptor de membrana. Los Smads interactúan directamente con la parte interna del receptor para su activación, con nucleoporinas y con importina- β para la entrada y salida del núcleo, así como con ADN de forma directa y con varios factores de transcripción, coactivadores de transcripción o corepresores que ajustan la función de Smads en cada contexto (descrito en (Massagué, 2012; Massagué et al, 2005)). Así, utilizando una aproximación simple, las proteínas Smad pueden ser interpretadas como un sistema conservado presente en todas las células, que pueden ser fácilmente adaptadas para formar complejos funcionales versátiles. A su vez, cada uno de estos complejos específicos lleva a cabo numerosas tareas que son requeridas por las células animales sanas y así desarrollar sus funciones vitales y de destrucción natural. Con el presente trabajo, y gracias a la combinación de los enfoques de biología celular y estructural hemos investigado algunos aspectos críticos de la red de

señalización TGF- β con el objetivo de iluminar algunas de las caras ocultas de la función de los Smads, específicamente el problema de cómo se regulan los controles que permiten pasar de la activación mediante Smads a su degradación y de cómo se regula la competencia entre R-Smads y I-Smads.

En la primera parte del trabajo se muestra que los pares de dominios WW presentes en las proteínas de unión a los Smads resultan ser lectores especializados del código escrito sobre la secuencia de los Smads. Este código se consigue gracias a la modificación de la proteína Smad mediante la introducción de una fosfoserina en la secuencia de Smad1 y que lo convierte en sitio de interacción con el coactivador YAP y así logran obtener un máximo de acción transcripcional. La presencia de un segundo sitio de fosforilación hace que los Smads se conviertan en el objetivo de las ubiquitin ligasas, y se desencadena su posterior eliminación tras participar en la transcripción. El código de este interruptor de acción-degradación está escrito por las quinasas CDK8/9 y por GSK3, que actúan en la región de unión de los R-Smads. En nuestro modelo, CDK8/9 crean sitios de unión que son preferentemente reconocidos por cofactores transcripcionales. Estas fosforilaciones activan los Smads para su posterior fosforilación por GSK3, lo que crea sitios para la unión de ubiquitin ligasa a expensas de los sitios de unión al cofactor transcripcional. Por lo tanto, GSK3 introduce una fosforilación en la región conectora de los Smads que los modifica desde favorecer la acción de los Smads a favorecer su destrucción.

Para explicar estas observaciones, proponemos que la degradación es un precio que pagar cuando las moléculas R-Smad han participado activamente en la transcripción de genes. CDK8 y 9 son los componentes del complejo mediador CDK8/CyclinC/Med12/Med13 transcripcional y del complejo CDK9/CyclinT P-TEF β que regulan la actividad de la ARN polimerasa II durante la transcripción (Durand et al, 2005; Komarnitsky et al, 2000; Malik

y Roeder, 2000). CDK8 / 9 tienen acceso a las moléculas de Smad en la cromatina pero no para aquellas moléculas de Smads que no se quedan enganchados en los complejos transcripcionales (Alarcón et al., 2009).

Nuestro trabajo proporciona una base estructural y funcional que explica la participación de los tándem de dominios WW en estas interacciones proteína-proteína. Por un lado, los dominios WW de Smurf1 y YAP logran alta especificidad global en el reconocimiento de la diana actuando en pares, extendiéndose la superficie interactuante para reconocer no sólo el sitio canónico PY sino también los motivos adyacentes pS / pTP. Las interacciones con los sitios de pSP fueron inesperadas sobre todo desde las predicciones basadas en la conservación de secuencia, ya que sólo el dominio WW de Pin1 se ha descrito como un motivo pS / pTP de unión, y los residuos del dominio WW de PIN1 implicados en la interacción con el fosfato no están estrictamente conservados en los dominios WW de YAP o de Smurf1.

Por otra parte, puesto que la región de unión de R-Smads contiene un conjunto de sitios de fosforilación de CDK, y un conjunto de sitios GSK3 dependientes de CDK, los motivos claves de reconocimiento de serinas fosforiladas pueden sintonizarse mediante fosforilación para así unirse de manera óptima tanto a cofactores transcripcionales (mono-fosforilado) o ligasas de ubiquitina (di-fosforilado). Los dominios WW2 tanto de Smurf1 como de YAP reconocen el motivo PY canónico mientras que los WW1 respectivos de cada proteína reconocen el sitio fosforilado presente en Smad1. Sin embargo, el dominio WW1 de YAP tiene preferencia por el sitio de mono-fosforilado, mientras que el dominio WW1 atípico de Smurf1 prefiere el motivo di-fosforilado. La capacidad de los dominios WW para reconocer el motivo pS (-4) pS era desconocida hasta ahora, demostrando una vez más la gran adaptabilidad funcional de estos dominios de proteína.

La interacción de estas proteínas YAP y Smurf1 con Smads se determina no sólo por la especificidad de unión de los distintos dominios WW, sino también por la configuración de los dos sitios de unión en los Smads. Por ejemplo, Smurf1 no se puede unir a Smad3 porque la posición N-terminal de la pT motivo [PY] relativa a la agrupación pSer en Smad3 es opuesta a la orientación requerida por los dominios WW1 y WW2 de Smurf1. Los presentes hallazgos amplían la versatilidad estructural y funcional de los dominios WW como módulos de interacción proteína-proteína y sugieren que aún reservan muchas sorpresas por descubrir, sin duda más de las que se propusieron inicialmente (Macias et al., 2002).

El reconocimiento de los distintos códigos de fosforilación en los Smads por diferentes proteínas con dominios WW ofrecen amplias oportunidades para la regulación de la función. Como un adaptador que se une al conector WW1-WW2 de Smurf1 (Lu et al., 2008), CKIP1 puede inducir una orientación óptima de estos dominios WW para facilitar el contacto con Smad1. Es de señalar que existe también una isoforma de Smurf1 que contiene una inserción en la región que conecta los dos dominios WW1 y WW2 para la que quizás se haya optimizado otra forma de reconocimiento con Smad1, que igual requiera la presencia de otro adaptador (Schultz et al., 2000). De forma similar, en células de mamíferos y en drosophila, YAP realiza algunas respuestas BMP, pero no en otras (Alarcón et al., 2009). Análogamente, Pin1 mejora los efectos de ciertas señales de TGF- β en la migración de células de mamífero, pero otras no (Matsuura et al., 2009). Nada nos hace pues evitar pensar que exista un repertorio más amplio de lectores de la secuencia de los R-Smad con variaciones de códigos que actualmente se desconocen.

El mecanismo de control de acción-degradación delineado aquí implica una notable concentración de las funciones de unión a proteínas en una región determinada de las proteínas Smad (Figura 9). Sin embargo la existencia de

mecanismos on-off regulados por fosforilación no es exclusivo de los Smads pues también ocurre en la región reguladora GS de tipo I de TGF- β quinasa del receptor (Huse et al., 2001). La vía de núcleo TGF- β /Smad se caracteriza, pues, por el uso económico que hace de los elementos estructurales clave.

En la segunda parte del trabajo presentado en esta tesis, se ha investigado la interacción del coactivador YAP y de dos ubiquitín ligasas con Smad7, el inhibidor de las cascadas de señales iniciadas por TGF- β y BMP. El interés de esta parte del trabajo residía en intentar recabar información sobre los factores que determinan que las mismas ubiquitín ligasas reconozcan tanto a los R-Smads como a los I-Smads. De hecho, las reglas que rigen el reconocimiento específico de las proteínas diana por parte de las E3 ubiquitín ligasas tipo HECT aún permanecen en la actualidad como una de las grandes incógnitas de su mecanismo de actuación. Lo único cierto hasta la fecha es que los resultados experimentales disponibles reflejan un escenario más complejo de lo que se esperaba originalmente. Algunos miembros de la familia de estas ligasas reconocen motivos PY usando para ello un único dominio WW. Ejemplos de este caso son el reconocimiento de la proteína del virus de Epstein Barr LMP2A por parte de la ubiquitín ligasa Itch (Morales et al., 2007) y de la unión por parte de Nedd4 al canal de sodio (Kanelis et al., 2001) y a Commissureless (Kanelis et al., 2006). En otros casos, tales como la unión de Smurf1 a los R-Smads, las ligasas utilizan un par de dominios WW consecutivos para expandir la interfaz de unión con un sitio de unión compuesto que incluye sitios pSer /pThr-Pro además del motivo canónico PY, una combinación que permite la regulación y optimización de la interacción de las ligasas con sus dianas gracias a la actuación de quinasas específicas que trabajan de forma coordinada (Alarcón et al, 2009; Aragón et al, 2011; Gao et al, 2009).

Los estudios presentados en este trabajo demuestran que los dominios WW poseen una mayor versatilidad funcional de la que se predecía hace algunos años (Macias et al 2002). Parece que al menos algunos de ellos se adaptan a la proteína diana a la que las ligasas tienen que reconocer, siendo capaces de especializarse para funcionar bien de forma independiente (este capítulo) o de una manera concertada (Alarcón et al, 2009; primer capítulo y en Aragón et al, 2011.; Gao et al, 2009), dependiendo de la necesidad particular. Por ejemplo, las interacciones de las ligasas Smurf1 y Smurf2 con Smad7 implican a un único dominio WW y un único sitio de unión, mientras que Smurf1 une a Smad1 utilizando dos dominios WW y dos sitios de unión. En nuestros estudios realizados con Smurf2 y Smad7 no observamos sin embargo ni los contactos descritos entre los dos dominios WW ni los contactos entre el primer dominio WW y Smad7 que se describe en un trabajo reciente (Chong et al., 2011). Hemos observado además que los dominios WW1 y WW2 de Smurf1 y Smurf2 respectivamente tienen una alta tendencia a formar homo-dímeros a través del dominio WW1 en caso de Smurf1 y a través del dominio WW2 en el de Smurf2. Estos homodímeros se han detectado tanto por NMR, como por ITC y mediante la técnica de Ion mobility aplicada a espectrometría de masas. La presencia del péptido de Smad7 usado aquí no impidió la formación de los dímeros, interaccionando principalmente a través de contactos con el dominio de WW2 Smurf1 y con el dominio WW3 de Smurf2, en cuyas interacciones centramos el trabajo estructural. Recientemente se han publicado por otros autores que la proteína entera Smurf1 forma homo-dímeros y oligómeros in vitro e in vivo a través de contactos inter-moleculares en los que se implican el dominio C2 y el dominio WW de una molécula y al dominio HECT de la pareja (Wan et al., 2011). Además, previamente también se habían caracterizado contactos intra-moleculares entre el dominio C2 y el dominio HECT de Smurf2 (Wiesner et al., 2007). En ambas ligasas Smurf1/2 la conformación cerrada inhibe el mecanismo de auto- ubiquitinación de la proteína, que

siempre se observa en todos los experimentos en los que se emplean las proteínas enteras (Wan et al., 2011). Es posible que, además de estas interacciones descritas entre los dominios C2 y HECT existan también, los contactos que hemos observado entre los dominios WW WW detectados en nuestro trabajo y contribuyan a la formación de los dímeros y oligómeros en vivo, y a la estabilización de las conformaciones cerradas de Smurf1 y 2.

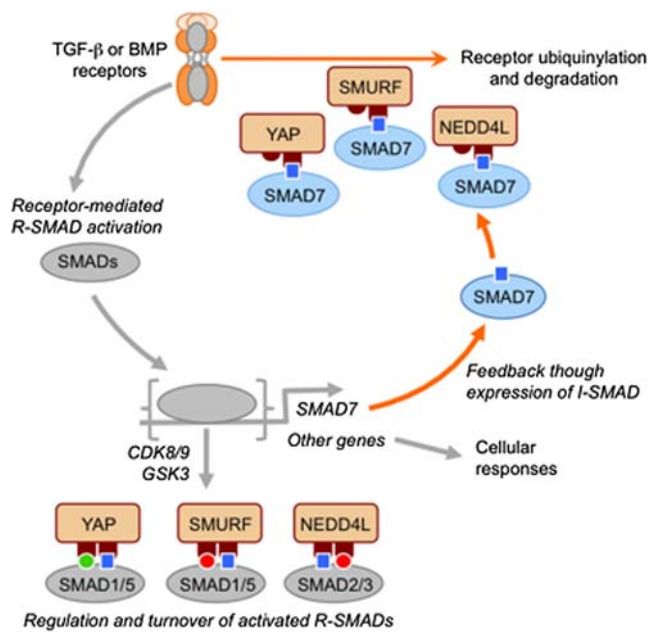
Quizás un posible mecanismo de interacción incluya un proceso de apertura tanto de Smad7 como de las proteínas Smurf1/2. Por ejemplo, en presencia de dos equivalentes de Smad7, podría producirse una reacción de apertura de los contactos entre los sitios MH1 y PY de la proteína Smad7 compitiendo el dominio MH1 de Smad7 por el dominio HECT de Smurf1 y los dominios WW2 interaccionando con el sitio PY, facilitando así la apertura del dímero entre dominios WW1-WW1. El resultado de esta reacción sería la generación de dos equivalente activados de complejos Smurf1-Smad7. Un mecanismo similar podría ocurrir entre Smurf1 y la interacción con Smad1. Por supuesto, otra explicación podría ser que los homodímeros entre los pares de dominios WW-WW aquí detectados pudiesen ser únicamente una consecuencia de las condiciones experimentales utilizadas in vitro.

Nuestro trabajo también muestra que YAP requiere sólo el dominio WW1 para la unión con la región de Smad7 que contiene el motivo PY. Cabe destacar que en la interacción con Smad1, YAP utiliza este dominio WW para la unión de un motivo con fosfoserina, mientras que emplea el dominio WW2 para la unión del motivo Smad1 PY equivalente al de Smad7 (Aragón et al., 2011). Para poder discernir las razones de estas diferencias con respecto a la interacción con Smad1, decidimos resolver las estructuras de los dos complejos posibles de YAP con Smad7, aunque la interacción con el segundo dominio fuese de más baja afinidad. Al comparar las estructuras vimos que el dominio WW2 contiene residuos de carga negativa en la zona donde la cadena

lateral del residuo E205 presente en Smad7 contacta con el dominio, lo que desestabiliza la interacción. La afinidad de la YAP-WW1 WW2 por la región de Smad1 con dos sitios de unión pSP-PY es 8 veces mayor que entre sólo el dominio WW1 y el sitio PY de Smad7. Sin embargo, la concentración de Smad7 en el núcleo es muy alta y en principio podría competir in vivo con Smad1 para secuestrar parte de la proteína YAP localizada en el núcleo, proporcionando un nuevo escenario para el papel inhibidor de Smad7.

Proponemos que los pares de dominios WW-WW presentes en las proteínas reguladores de los R-Smads forman unidades funcionales que han evolucionado para reconocer regiones que contienen motivos PY de longitud y complejidad variable, incluyendo motivos PY compuestos con sitios fosforilados adyacentes además de otros motivos más alejados que también contienen sitios fosforilados en Ser y Thr mientras que prefieren motivos más simples y sin necesidad de modificaciones en el caso del inhibidor Smad7. Estas características amplían la versatilidad funcional de las ubiquitín ligasas E3 mediante la optimización de la superficie de interacción en función de las necesidades. Por otro lado, NEDD4L y Smurf1/2 se asocian con Smad7, para participar en la degradación de los receptores de TGF- β de forma indirecta. Smad7 también puede actuar como secuestrador de YAP en el núcleo y así limitar también la función activadora de los R-Smads. Por otro lado, los R-Smads son dianas directas de las ubiquitín ligasas y colaboradores funcionales de YAP sólo en etapas específicas del ciclo de señalización (Aragón et al., 2011).

La ausencia de un requisito adicional de fosforilación en la interacción con Smad7 implica que tanto YAP, NEDD4L como Smurf1/2 son dianas constitutivas de Smad7 mientras que son reguladores condicionales, - dependiente de fosforilación- de R-Smads en la transducción de señales por TGF- β y BMP (véase el Representación esquemática de los resultados en la Figura).



9.5 Conclusiones

Nuestros datos revelan una sorprendente independencia de la fosforilación en las interacciones de Smad7 mientras que la fosforilación es crítica para la interacción de los R-Smads con YAP, Smurf1, Smurf2 y proteínas NEDD4L. Además, mientras que los pares de dominios WW reconocen a los R-Smads, la unión al inhibidor de Smad7 requiere la presencia de un solo dominio WW. Estas observaciones sugieren una versatilidad funcional de las proteínas que contienen dominios WW y también de los dominios WW como mediadores de las interacciones específicas con proteínas Smads, mucho mayor de la que pensábamos hasta ahora.

Los resultados de este trabajo se destacan las siguientes cuatro conclusiones:

1. E3 ubiquitín ligasas del tipo HECT y el transcripción coactivador YAP pueden utilizar dominios WW solos o asociados en pares para seleccionar y distinguir entre las distintas proteínas Smad.
2. Los sitios de unión pueden ser múltiples, como en el caso de los R-Smads o sitios únicos como en I-Smads.
3. Los sitios de unión en R-Smads dependen de fosforilación mientras que el sitio de reconocimiento para los I-Smad es independiente de la fosforilación.
4. En Smurf1 los homodímeros de los dominios WW-WW detectadas por técnicas de RMN y de IM-MS pueden estabilizar la conformación cerrada de la ligasa, descrita previamente *in vivo*.