



Structural Insights into Substrate Binding and Regulation of E3 Ubiquitin Ligases in the Nedd4 Family using NMR Spectroscopy

Albert Escobedo Pascual

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.

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

TESI DOCTORAL

**Structural Insights into Substrate Binding and
Regulation of E3 Ubiquitin Ligases in the Nedd4
Family using NMR Spectroscopy**

Albert Escobedo Pascual, 2014

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

PROGRAMA DE DOCTORAT EN BIOMEDICINA

**Structural Insights into Substrate Binding and
Regulation of E3 Ubiquitin Ligases in the Nedd4
Family using NMR Spectroscopy**

Memòria presentada per
Albert Escobedo Pascual

Per obtenir el grau de doctor per la Universitat de Barcelona

Revisat per:

Dra. Maria J. Macias Hernández

Directora i Tutora de Tesi

Prof. d'investigació ICREA* i investigadora principal
del grup RMN de proteïnes de l'IRB Barcelona

*(Institució Catalana de Recerca i Estudis Avançats)

Realitzada a l'Institut de Recerca Biomèdica de Barcelona

Parc Científic de Barcelona

Barcelona, 2014

Ah! La ciència sense humanisme
és lletra morta. Viu o mort, un roc
no pot ser filtre del discurs del foc,
oracle que es transmet vora l'abisme.

El darrer mot és sempre l'egoisme;
foguera i olivera semblen poc
als qui no saben els lligams del joc
i fan dels somnis l'únic realisme.

Sequaços, regle en mà, mouen l'atzar
i ignoren el sentit de la florida.
Afirmo contradir-me com el mar,

però amb les rels posades a la vida.
Els savis al servei del militar
es perden en un clos sense sortida.

Joan Brossa, *Antologia de poemes de revolta*, 1979

Acknowledgements

Perhaps the most remarkable thing that I have learned during my years as a PhD student is that individuals don't do science, teams do science. The work presented in this thesis is not only mine; many people have contributed to it in one way or another.

Dr **Maria J Macias** has directed this thesis to what it is now, giving me the guidelines of the project and converting my raw, crude data into a nice piece of science. Maria, by your side I have learned lots, but maybe the most important is the life lesson you have recently given us all. Your courage and optimism are now in my life backpack.

My colleagues at the Macias' lab have also been essential. **Lidia**, I am probably not mistaken if I say that almost no experiment would have been successful without your help. You have been there to convert my pessimistic views into hope. Thanks for your constant support and advice. **Tiago**, I could always count on you when I needed a hand, and there is direct evidence of that in this thesis. At the end of the day, this is what friends do right? Many thanks to the seniors of the group: **Jimmy, Pau, Toni** and **Eric**. Your experience and advice have been of great value for my experiments to succeed. **Constanze**, I am glad to see you are about to make it too. My best wishes for your success and good luck for your defence. **David** and **Jordi**, I have no doubt you will get there some day too, keep working hard. **Cristina**, you may not strictly be a lab mate, but your permanent good mood has helped too. I don't want to forget about former lab members. **Nina**, part of this work is also yours. **Jordi Mas**, my early days in the lab were probably the funniest with you being around.

Several people from outside the lab have also contributed to this work. Special thanks to Dr **Miriam Royo** and **Gerardo Acosta**, whose help in the peptide synthesis has been priceless. Thanks to our collaborators at the group of Dr **Joan Massagué**. People at UB, PCB and IRB facilities have also helped notably, as well as people in the IRB administration offices.

The IRB is full of wonderful people, and I have met good friends in here during these years. With them I have enjoyed criminally long breaks in the corridors of the institute or having beers at the terrace. **Oscar** and **Abraham**, we have shared our worries and hopes since the beginning. Many thanks for your true friendship. **Maria**, I still miss the mid-afternoon donut that used to solve everything. **Edgar**, have an

early recovery, football needs you back. My best wishes for your defence. Many other people should be mentioned: mates at the labs of Drs Giralt, Salvatella, Stracker, Riera, Royo, Pons and Orozco.

Prof Dr **Hartmut Oschkinat** welcomed me to his lab at the FMP Berlin during my internship and I learned lots under his direction. Thank you Hartmut for those months and the nice beers we shared. I send my greetings to all the people at Hartmut's lab. Berlin felt much more like home by the side of the friends I made there. **Ramon, Lidia** and **Laura**, I wish you all the best.

Several public agencies have given me the opportunity to become a doctor. I thank the IRB Barcelona (PhD fellowship), the Catalan Government agency AGAUR (FI fellowship) and the Spanish Government Ministerio de Educación, Cultura y Deporte (FPU fellowship) for the funding along the PhD. My internship in Berlin was possible thanks to the German Government agency DAAD.

Even though it can sound odd to some of us, there is a world outside science. I will switch to Catalan to greet my family and friends.

Una abraçada molt forta a tots els meus col·legues de Lleida. **Jordi**, no sé si és bo o no que et quedi algú de quan encara et menjaves els mocs, però aquí estem i és tot plegat molt salvatge. **Edu**, et desitjo el millor en la teva recta final i molta sort en la defensa. Una abraçada revolucionària per tots dos, com diria Pau Llonch. **Pol**, ets admirable en mil aspectes. **Ori**, sens dubte la trajectòria dels alegres estudiants està gaudint d'un ascens meteòric. **Ernest**, tot devoció i disciplina soviètica, conseqüent amb el que un pensa i sent. A tots tres us desitjo molta sort amb el grup. Petons i abraçades a tota la resta: **Sergi, Vilà, Rose, Uxue** i tots els altres.

No em puc oblidar dels meus "flipatletes" preferits: l'**Isaac**, l'**Abel** i l'**Agus**. Per fi deixaré de donar-vos la vara amb la tesi! És tot un plaer compartir entrenament, reptes i festes majors de Faió amb vosaltres. I ja que parlem de Faió, una abraçada molt forta per als meus amics i quintos **Dani** i **Mireia**.

I per fi arriba el torn de la família, puntal vertader del que sóc. Hi ha qui diu que a la vida no et regalen res. Dec ser molt afortunat: a mi la meua família m'ho ha donat tot.

Tot el meu amor per la **Sandra** i l'**Agus** (l'únic que surt dos cops). Sandra, la inexplicable devoció que sents per mi no té preu. Tots dos heu estat sempre molt importants, però us heu tornat fonamentals en aquests anys que he estat a Barcelona. L'**Alèxia** no podria tenir uns pares millors.

Una abraçada molt forta al meu padrí **Javi** i a l'**Anna Carmen**. A vosaltres us tinc més lluny però sou igualment molt importants. Gràcies per fer-me l'orgullós padrí de la vostra preciosa **Arlet**.

I ja que estem amb els nouvinguts, a més de les ja esmentades, molt d'amor per a la **Júlia**, la **Martina** (encara no-nata!), l'**Iker**, la **Míriam**, l'**Ainara** i la **Ingrid**. Sou l'orgull i l'alegria de Ca Carbassa i Ca Caramelero. A tots us prometo que no descansaré amb l'objectiu de deixar-vos a vosaltres el millor dels mons possibles. Molts d'ànims i abraçades als seus pares respectius.

Petons a la meva preciosa cosina germana **Anna**, m'alegro que les coses et vagin tan bé. Una abraçada fortíssima al meu cosí germà **Adrià**, de qui no dubto que sabrà trobar el seu camí. Tot el meu agraïment als tiets que m'han criat: **Dani**, **Amparo**, **Tere** i **Oscar**. Molts petons també per a la legió de tiets, cosins, cosins segons i tota mena de parentescs a qui no puc citar perquè si no això no s'acabaria mai però a qui aprecio molt.

Les meves àvies han estat segones mares. Citant al gran Ovidi Montllor, "*Com un record d'infantesa, sempre et recordaré a tu, Teresa*". T'estimo, iaia-mamá. Iaia **Margarita**, encontrar la formulita no es cosa fàcil però estamos en ello. Un beso enorme. Molts petons també pel meu avi **Ramón**, que sempre m'ha consentit i estimat, i també per la seva companya **Maria**. Y un recuerdo para el abuelo **José**.

La meva germana **Natàlia** és una gran lluitadora. De vegades sembla que la vida és especialment injusta amb algunes persones particularment bones, però tu sempre li respons amb tempra i bona cara. Hem crescut junts i hem arribat on som, i espero que segueixi així per sempre. A tu i al **Josep** us estimo i us desitjo la millor de les sorts en aquesta nova etapa que acabeu d'encetar.

Per acabar, tot el meu amor, respecte i gratitud pels meus pares **Pepita** i **Miquel**. Ells van decidir dedicar la seva vida a la de la meva germana i a la meva, i l'agraïment que sento no es pot expressar en paraules. Mama, gràcies pel suport constant, diari i fonamental que em dediques. Papa, gràcies pel teu esperit lluitador i inconformista, i per ensenyar-me a calcular sempre la tercera jugada. Aquesta tesi és tant meva com vostra.

To all of you, my most sincere gratitude
A tots vosaltres, la meva gratitud més sincera

Declarations

1. This work has been carried out under the supervision of Maria J. Macias, Ph.D. (ICREA research professor and PI of the Structural Characterization of Macromolecular Assemblies 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. References are given as footnotes at the beginning of the corresponding results chapters.
4. Financial support was obtained from the IRB Barcelona (PhD fellowship), the Generalitat de Catalunya agency AGAUR (FI fellowship), the Ministerio de Educación, Cultura y Deporte, Gobierno de España (FPU fellowship) and the Bundesregierung agency DAAD (internship fellowship). This work was supported in part by the Ministerio de Economía y Competitividad, Gobierno de España (BFU2008-02795 and SAF2011-25119 grants).



DAAD

Deutscher Akademischer Austausch Dienst
Servicio Alemán de Intercambio Académico

Table of Contents

PART I. INTRODUCTION

CHAPTER 1. INTRODUCTION	19
1.1. Ubiquitination	19
1.2. Ca ²⁺ signalling	30
1.3. TGF- β signalling	35

PART II. THESIS AIM AND OBJECTIVES

CHAPTER 2. THESIS AIM AND OBJECTIVES	43
2.1. Background and scope	43
2.2. Thesis objectives	44

PART III. MATERIALS AND METHODS

CHAPTER 3. MATERIALS AND METHODS	49
3.1. Molecular biology methods	49
3.2. Solid-phase peptide synthesis	60
3.3. Isothermal Titration Calorimetry (ITC)	67
3.4. NMR Spectroscopy of Biomolecules	70
3.5. Molecular dynamics simulations	87

PART IV. RESULTS

CHAPTER 4. SMAD LINKER PHOSPHORYLATION AFFECTS SUBSTRATE BINDING CAPABILITIES OF NEDD4L AND RELATED PROTEINS	91
4.1. Introduction	91
4.2. Solid-phase peptide synthesis	94
4.3. Standard procedures used in the reported syntheses	102
4.4. Synthesis of a set of Smad1 phosphopeptides	104
4.5. Synthesis of a set of Smad3 phosphopeptides	113
4.6. Native chemical ligation of a phosphopeptide	118
4.7. Summary	123

**CHAPTER 5. ACTIVITY REGULATION OF THE E3 UBIQUITIN LIGASE
NEDD4L: CA²⁺/IP₃ SIGNALLING TRIGGERS THE TRANSITION FROM THE
LATENT TO THE ACTIVE CONFORMATION** **127**

- 5.1. Introduction **127**
- 5.2. Deciphering the contacts that maintain Nedd4L in a latent state and its switch to the active form **130**
- 5.3. Summary **148**

**CHAPTER 6. FULLY FUNCTIONAL NEDD4L MOLECULES TARGET
PARTIALLY UNSTRUCTURED COUNTERPARTS AS A MEANS TO
REGULATE THE TURNOVER RATE OF THE LIGASE** **151**

- 6.1. Introduction **151**
- 6.2. Ubiquitination of the ubiquitin ligase: a bi-molecular reaction **154**
- 6.3. Summary **169**

**CHAPTER 7. PARAMAGNETIC LABELLING OF THE NEDD4L WW3 –
HECT-PY COMPLEX** **173**

- 7.1. Introduction **173**
- 7.2. MTSL derivatization of Nedd4L WW3 cysteine mutants **174**
- 7.3. Summary **179**

PART V. DISCUSSION

CHAPTER 8. DISCUSSION **183**

- 8.1. Antecedents **183**
- 8.2. Smad phosphorylation affects substrate binding in Nedd4L and related proteins **184**
- 8.3. Nedd4L changes from a latent to an active conformation in response to IP₃/Ca²⁺ **187**
- 8.4. A crossed mechanism for Nedd4L ubiquitination **190**
- 8.5. Concluding remarks **193**

PART VI. APPENDIX

CHAPTER 9. APPENDIX **197**

- 9.1. Chromatographic and MS characterization of the peptides **197**
- 9.2. Reconstruction of the β 1- β 2 loop in the crystal structure of the C2 domain **203**
- 9.3. HSQC spectra of the free and HECT-PY bound WW1, WW2 and WW4 domains **204**
- 9.4. Sequence alignment of the HECT domain of Nedd4L in metazoans **206**

PART VII. RESUMEN EN CASTELLANO

CHAPTER 10. RESUMEN EN CASTELLANO	211
10.1. Introducción	211
10.2. Objetivos	218
10.3. Resultados	220
10.4. Discusión	228
REFERENCES	233
ABBREVIATIONS AND UNITS	245
CURRICULUM VITAE	249

Part I.

Introduction

CHAPTER 1. INTRODUCTION	19
1.1. Ubiquitination	19
1.2. Ca²⁺ signalling	30
1.3. TGF-β signalling	35

CHAPTER 1. Introduction

1.1. Ubiquitination

1.1.1. Post-translational modifications (PTMs)

Living cells organize their fundamental biochemical processes in complex networks of effector agents, most of them being proteins. The abundance of each individual protein is primarily controlled by gene transcription into mRNA, alternative RNA splicing and ribosomal translation into polypeptide chains. The complexity of eukaryotic cells allows them to undergo different cell cycle phases, quickly adapt to environmental changes and coordinate with each other in populations, tissues and organisms. The behaviour and fate of a particular protein in a specific situation is generally modulated by post-translational modifications (PTM), which consist in the covalent attachment of one or more chemical groups or polypeptides to the native sequence.

Over 200 different PTMs have been so far identified by mass spectrometry (Mann and Jensen, 2003). Phosphorylation is currently regarded as a universal mechanism to control enzymatic activity and protein-protein interactions through the reversible addition of the small charged phosphate group, which affects the biophysical and structural properties of proteins (Johnson and Lewis, 2001). Other PTMs such as glycosylation or the addition of lipids are crucial for the proper folding and function of newly synthesized proteins. Histone methylation is one of the fundamentals of epigenetics, affecting chromatin structure and gene transcription with direct consequences in the phenotype (Strahl and Allis, 2000).

1.1.2. Ubiquitin labelling determines protein fate

Ubiquitination is a PTM conserved throughout eukaryotic evolution. It consists in the covalent linkage of one or more moieties of the protein ubiquitin to the substrate protein. Ubiquitin is a highly conserved 76 residue-long polypeptide that was named after its ubiquitous localization in all cell types and tissues (**Figure 1.1A**). It is generally attached through an amide bond involving the free carboxyl group in its C-terminal glycine residue and the ϵ amino group in the side-chain of a lysine residue in the target protein.

Ubiquitin labelling of a protein has direct consequences in its fate (**Figure 1.1B**). Substrate labelling with a single ubiquitin moiety (mono-ubiquitination) has a role in a variety of processes such as DNA repair, histone regulation, endosomal sorting and endocytosis (Hoeller et al., 2006). Substrates with multiple lysine residues can be mono-ubiquitinated in several positions, which is also related to endocytosis.

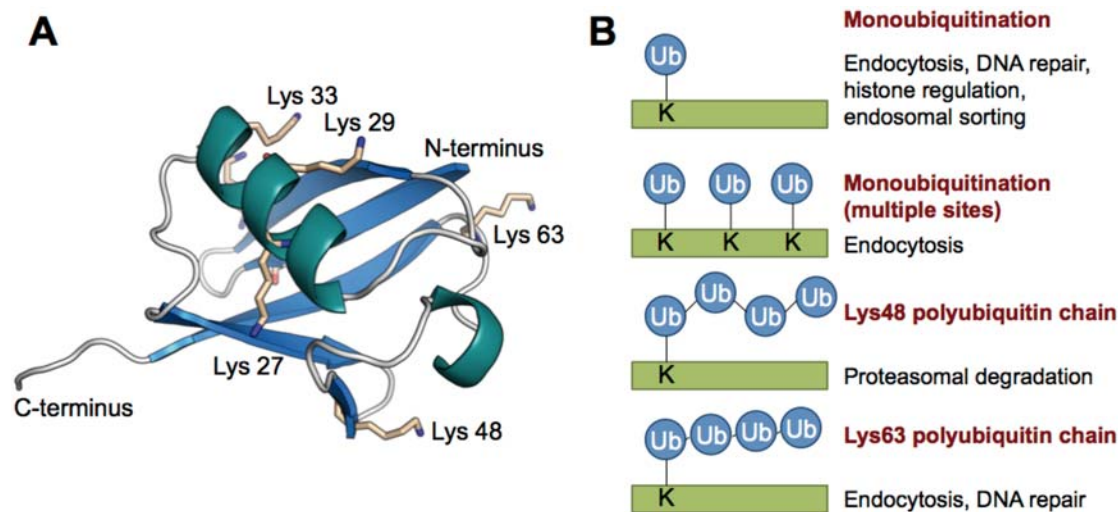


Figure 1.1: Ubiquitination labels proteins for a variety of fates. (A) Crystal structure of ubiquitin (PDB entry 1UBQ). (B) Different ubiquitination patterns have different consequences in the fate of the substrate protein.

Ubiquitin itself has seven lysine residues (namely K6, K11, K27, K29, K33, K48 and K63). Several ubiquitin moieties are attached to each other using the side-chains of these residues and establishing poly-ubiquitin chains, each labelling the substrate for a particular fate. It has recently been shown that ubiquitin moieties are also linked through the backbone amino group of the N-terminal methionine residues. Depending on the linkage, the topology of the chains varies: N-terminal and K68 linkages form extended chains with no contacts between the ubiquitin moieties, whereas K6, 11 and 48 linked chains are closed and present inter-molecular contacts. Ubiquitin binding partners have different types of ubiquitin-binding domains (UBDs), which are able to distinguish the distances between the ubiquitin moieties using spacers of different lengths between the ubiquitin interacting motifs (Komander and Rape, 2012; Sato et al., 2009; Sims and Cohen, 2009).

K48 poly-ubiquitin chains label the substrate protein for its proteolytic processing in the ubiquitin proteasome pathway (UPP). This was first discovered and characterized in the early 1980's by Rose, Hershko and Ciechanover, who were seeking an explanation for ATP dependence of intracellular enzymatically-catalysed proteolysis (Ciechanover et al., 1980; Hershko et al., 1980; Hershko et al., 1979).

Their findings unveiled the major eukaryotic mechanism for the control of protein turnover and disposal, which earned them the Nobel Prize in Chemistry in 2004.

Chains established through linkages at K68 are related to processes such as DNA repair and endocytosis (Miranda and Sorkin, 2007). N-terminal chains are related to NF- κ B signalling (Rahighi et al., 2009). Most of the consequences of alternatively linked chains are still to be elucidated, including mixed K48/K68 and branched chains.

1.1.3. Enzymatic mechanisms of ubiquitin labelling

Ubiquitination is achieved through the coordinate, sequential action of three different enzyme types: E1 (activating), E2 (conjugating) and E3 (ligase). Firstly, the E1 enzyme uses ATP to activate ubiquitin by adenylating its C-terminus. Ubiquitin is then covalently bound to the side-chain of a cysteine residue in the E1 enzyme through a thioester bond (**Figure 1.2**) (Haas et al., 1982). From there it is transferred to another cysteine residue in the E2 enzyme, and finally linked to the target substrate in an E3-dependent manner (Hershko et al., 1983). Humans have two different E1 enzymes, a few tens of E2 and a few hundreds of E3 ligases, which are specific for each particular substrate.

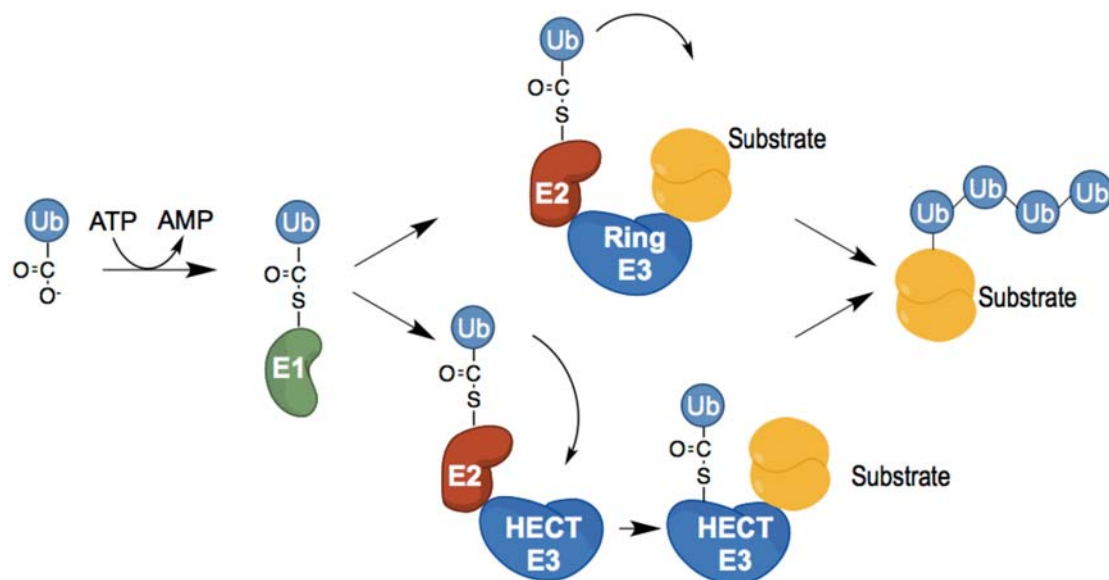


Figure 1.2: Ubiquitination is catalysed by the coordinate action of three enzymes. The E1 activates ubiquitin moieties in an ATP-dependent manner. These are then transferred to the E2, which binds the E3 ubiquitin ligase. Ring-type E3s catalyse the direct transfer of ubiquitin from the E2 to the substrate, whereas HECT-type E3s establish a covalently bound intermediate with ubiquitin before labelling the substrate.

E3 ubiquitin ligases are classified in two main groups depending on the mechanism they use to transfer the ubiquitin molecules from the E2 conjugating enzyme to the target substrate (Metzger et al., 2012). The HECT (Homologous E6-AP carboxyl terminus) ligases covalently bind the ubiquitin moiety before it is transferred to the substrate, while the RING (Really interesting new gene) ligases catalyse the direct transfer from the E2 to the target.

1.1.4. The ubiquitin proteasome pathway (UPP)

Substrates labelled with at least four units of K48-linked ubiquitin moieties are degraded in the UPP (Thrower et al., 2000). This is the main mechanism for the cells to control the turnover rates of proteins, which are continuously degraded and newly synthesized at particular rates and in response to defined stimuli as a means to maintain intracellular homeostasis.

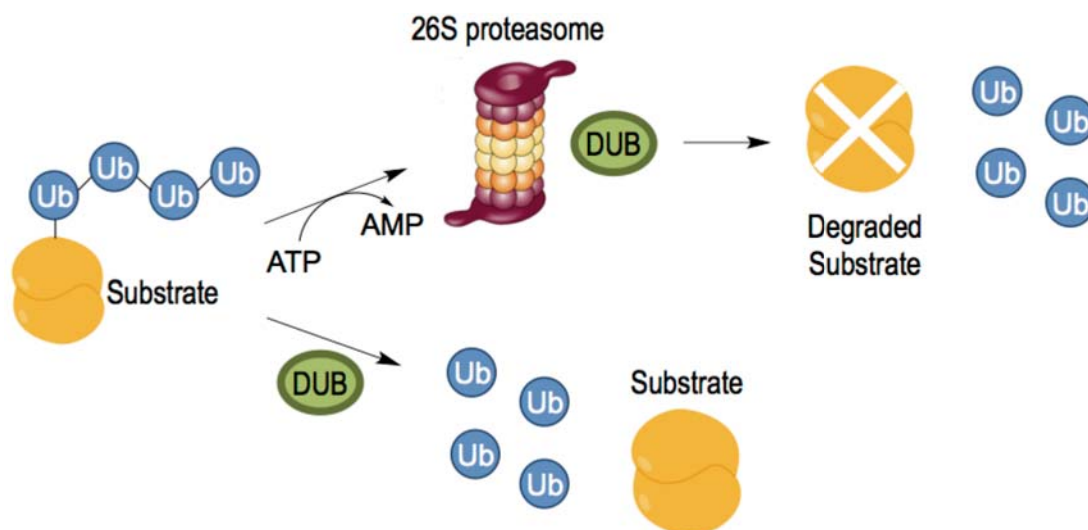


Figure 1.3: Lys48 polyubiquitin chains label substrates for proteasomal degradation in the UPP. The 26S proteasome degrades substrates labelled with at least four ubiquitin moieties, in an ATP dependent manner. DUBs detach ubiquitin from the substrate for further labelling cycles. DUBs can de-ubiquitinate substrates at any stage, rescuing them from degradation.

The labelled substrates are recognized and degraded by the 26S proteasome (Figure 1.3) (Tanaka et al., 1983). This is a macromolecular complex of about 2.5 MDa that is ubiquitous in all eukaryotes, archaea and in some bacteria. It is composed of two 19S and a 20S subunits, named after their Svedberg sedimentation coefficient. The 19S subunits are localized at both sides of the 20S subunit. They are responsible for recognizing the labelled substrate through their ubiquitin-binding motifs and for unfolding them in an ATP-dependent manner. The unfolded substrate proteins enter the 20S subunit, a hollow structure with protease active sites in the inner side (Baumeister et al., 1998). The substrate is decomposed in small peptides

that are further degraded into individual amino acids to be recycled in new protein syntheses.

Ubiquitin chains attached to the substrates are not unfolded and degraded but cleaved and decomposed to individual ubiquitin moieties by deubiquitinating enzymes (DUBs), so they can be recycled in further ubiquitination cycles (**Figure 1.3**). Indeed, DUBs are substrate-specific enzymes that oppose the action of E3 ligases at any stage of the ubiquitin-labelling pathway as a pathway regulatory mechanism (Komander et al., 2009).

1.1.5. The Nedd4 family of HECT-type E3 ubiquitin ligases

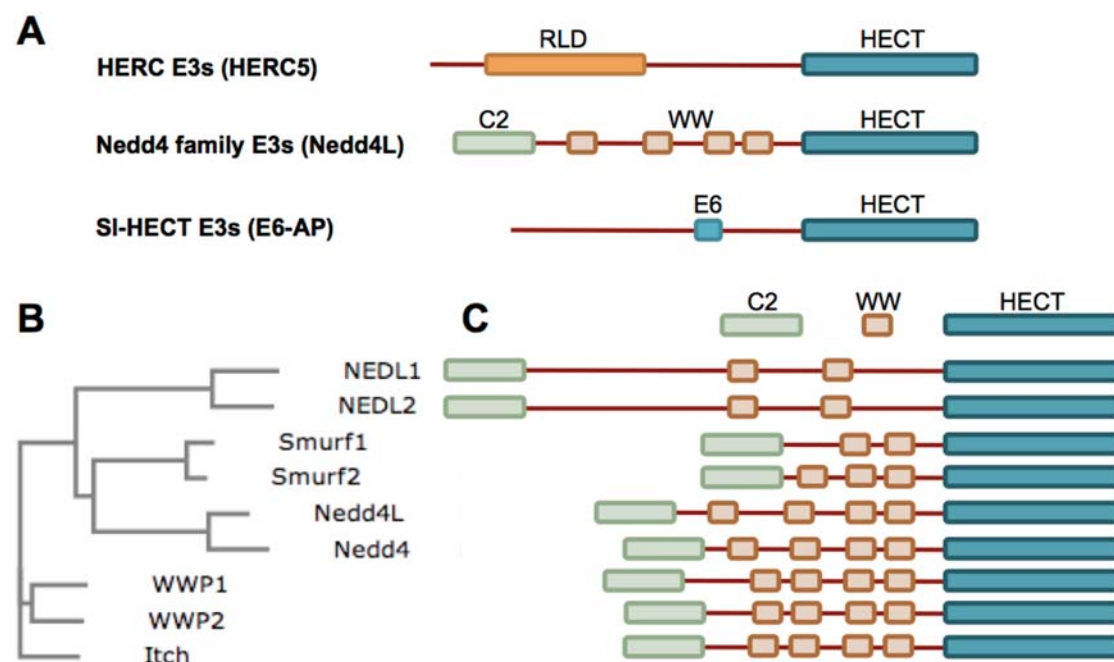


Figure 1.4: Human HECT-type ubiquitin ligases. (A) HECT-type E3 ubiquitin ligases are classified in three classes depending on their modular architecture. (B) Phylogenetic tree displaying the evolutionary relationship among the nine human members of the Nedd4 family of ligases, as obtained by a multiple sequence alignment using the software ClustalW2 (EBI-EMBL) (C) Domain diagrams of these ligases.

Based on genomic analyses, 28 different putative proteins with a HECT domain have been identified in the human genome (Scheffner and Staub, 2007). Those characterized so far are modular proteins with the HECT domain localized at the C-terminus. Depending on the accompanying domains, they are classified in three classes (**Figure 1.4A**). Members of the HERC family (six ligases in humans) contain at least one RLD (RCC1-like) domain, whereas ligases classified in the Nedd4 family (nine in mammals) have one N-terminal C2 domain and a variable number (between 2 and 4) of WW domains. Both RLD and WW domains act as protein-protein

interactions mediating modules and are generally involved in substrate recognition. Ligases without any of these two domains are classified as SI-HECT E3s and include E6-AP, the first HECT-type ligase to be discovered.

The Nedd4 family of ubiquitin ligases is currently the subset of HECT E3 ligases that has been studied with most detail (**Figure 1.4B** and C). There are nine members of the family in mammals, while Rsp5 is the only member in *S. cerevisiae* and its deletion is lethal (Hein et al., 1995). *S. pombe*, *C. elegans* and *D. melanogaster* express three members of the family. These ligases poly-ubiquitinate transmembrane receptors and intracellular proteins for their disposal by the UPP, and regulate trafficking and endocytosis of plasma membrane proteins by mono-ubiquitination.

1.1.5.1. Domain architecture

The **C2 domain** (Rizo and Sudhof, 1998) has a rod-shaped, β -sandwich structure with 8 β -strands and binds phospholipids such as phosphatidylinositol, phosphatidylserine and phosphatidylcholine (Plant et al., 1997). It is responsible for the localization of these ligases at the plasma and other intracellular membranes, where some of the substrates are located. The C2 domains are classified in two main classes, depending on their ability to coordinate Ca^{2+} ions. C2 domains that do not bind Ca^{2+} interact with membrane phospholipids using a positively charged patch located at a variable loop region. Instead, domains similar to the C2 in the protein kinase C- α (PKC- α) such as those found in the human ligases belonging to the Nedd4 family have acquired Ca^{2+} -binding capabilities at these positions able to mediate membrane localization.

The **WW domains** are the smallest independently folded protein domains found so far in nature, with about 35 residues and a canonical fold consisting in three anti-parallel β -sheets. They are named after two highly conserved Trp residues located 20 – 22 positions apart that are relevant in protein fold and substrate recognition. One of these residues, close to the N-terminus, is strictly conserved and establishes hydrophobic interactions with a conserved proline residue close to the C-terminus that define the fold of the domain. The other Trp residue is occasionally substituted with other aromatic residues including Tyr and Phe (Macias et al., 2002).

WW domains recognize a variety of proline-rich regions in their binding partners, and are classified depending on the actual motifs they recognize. Type I WW domains (56%) bind PY motifs (PPxY sequences), whereas type II and III domains (25%) bind PPLP and PR motifs (poly-proline motifs containing Leu and Arg residues

respectively). Type IV domains (4%) bind phospho-S/TP sequences, while all other domains remain unclassified (Kato et al., 2002; Kato et al., 2004).

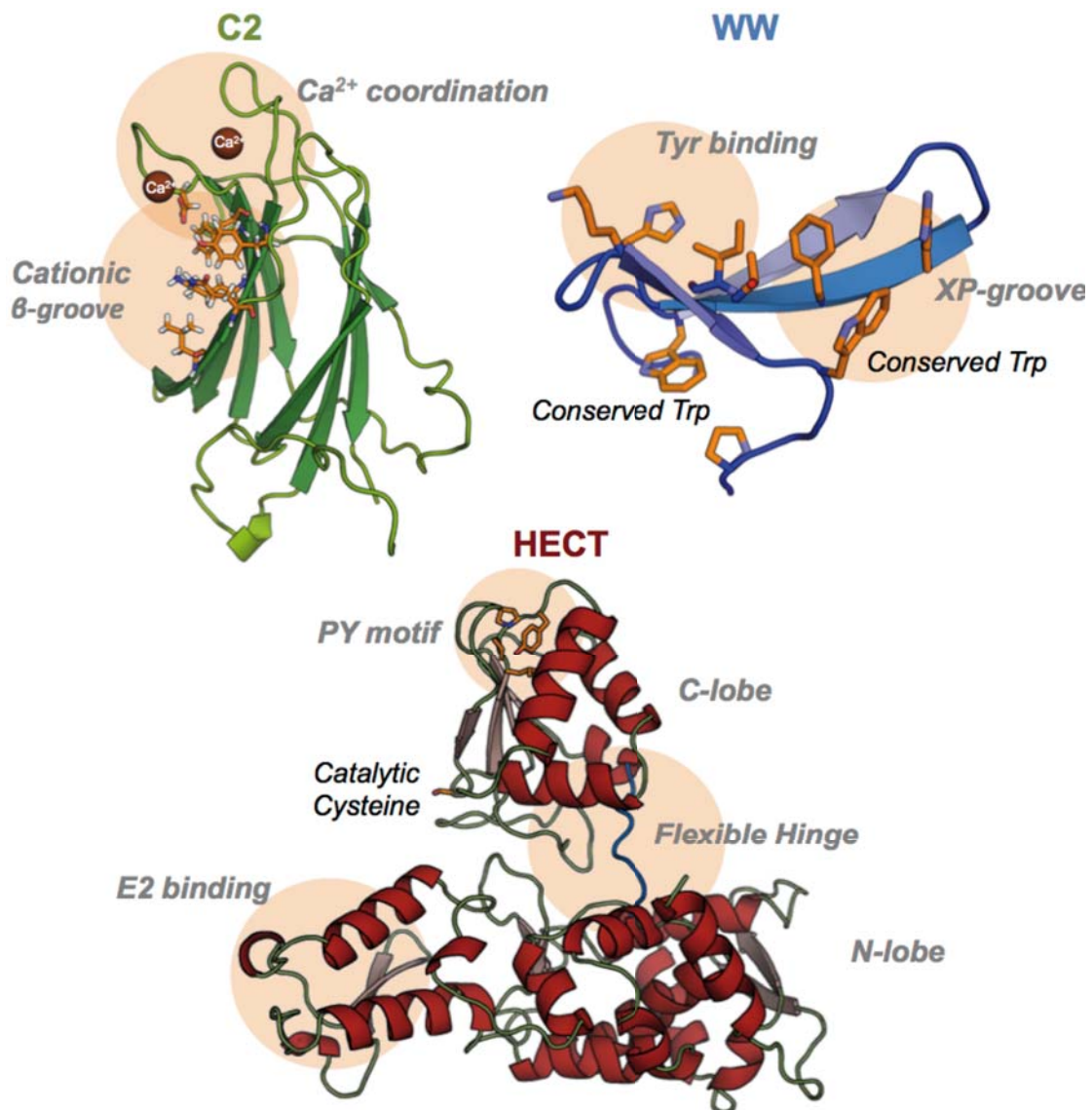


Figure 1.5: Domains integrating the modular architecture of E3 ubiquitin ligases in the Nedd4 family. All structures displayed are domains of Nedd4L. The crystal structure of its C2 domain (2NSQ) is shown in green, with its Ca²⁺-coordinating loops and its cationic β-groove (both implied in lipid binding) highlighted. The NMR structure of its WW3 domain (2MPT), a type I WW domain, is displayed in blue. The side-chains of the two conserved Trp and other relevant residues are shown in orange. Its XP-groove and its Tyr binding pocket are highlighted. The crystal structure of its HECT domain (3JW0) is shown in red. The flexible hinge connecting the C- and the N-terminal lobes is displayed in blue and highlighted, as well as its highly conserved PY motif and the E2 enzyme binding area. The position of the catalytic cysteine (here mutated to serine) is highlighted.

The **HECT** (Homologous to the E6-AP Carboxyl Terminus) is the biggest domain in these proteins (about 350 amino acids) and the one in charge for catalytically transferring the ubiquitin moieties to the substrates. It is structurally divided in two independently folded subdomains, namely the N- and the C-terminal lobes, which

have about 210 and 120 residues respectively and are connected through a flexible hinge. The ubiquitin-loaded E2 enzymes bind on the HECT N-lobe surface, while the highly conserved catalytic cysteine residue is located at the C-lobe. Structural and mutational studies have revealed that the hinge connecting both sub-domains is flexible enough to permit structural rearrangements of the sub-domains relative to each other for the catalytic residue to approach the ubiquitin-loaded E2 enzyme (Kamadurai et al., 2009; Verdecia et al., 2003). Notably, next to the catalytic cysteine and partially buried in the structure of the C-lobe the HECT domains of ligases in the family have a highly conserved I/LPPY motif, which slightly differs from the canonical PPxY sequence of a typical PY motif but that has been shown to bind WW domains when the HECT domain is unfolded (Bruce et al., 2008).

1.1.5.2. Function and regulation

The targets of the HECT E3 ligases of the Nedd4 family include several tumour suppressors, immune system differentiation agents and transmembrane receptors and channels (**Figure 1.6**). Genetic aberrations and altered expression patterns of many of these substrates have been observed in human cancers and immunological diseases (Bernassola et al., 2008). Many of them are recruited by viruses to maintain latency (Epstein-Barr Virus (EBV)) or as essential partners for budding (retroviruses such as Human immunodeficiency virus (HIV) and Human T-lymphotropic virus (HTLV)) (Ingham et al., 2004). Therefore, a tight regulation of the catalytic activity, substrate specificity and cellular localization of these ligases is required for the cell to maintain its homeostasis. Such regulation is achieved at different levels through PTMs (CHAPTER 4), the interaction with accessory proteins and inter-domain contacts that establish closed and inactive latent conformations able to transition to open and active ones in response to specific triggers (CHAPTER 5). Additionally, these ligases are able to regulate their own turnover rates by auto-ubiquitination as a means to modulate their abundance at each particular stage of cell cycle and differentiation (De Bie and Ciechanover, 2011; Mund and Pelham, 2009) (CHAPTER 6).

AIP4 is the human homologue of the mouse protein **Itch** and its main substrate is JunB, a T-cell differentiation factor. Defects in Itch are the cause of the *itchy* phenotype in mice, characterized by permanent skin irritation and multi-systemic inflammation (Perry et al., 1998). A closed conformation has been identified mediated by HECT contacts with the central region of the ligase that are released by JNK-1 phosphorylation of a proline rich region (PRR) next to WW1 (Gallagher et al., 2006). The transition to the active conformation is also mediated by Ndfip1/2 (Nedd4 family-interacting proteins 1 and 2) binding. Additionally, the

phosphorylation of Tyr371 next to WW3 by FYN impairs Itch recognition of many of its substrates including JunB, Notch and Smad2 (Morales et al., 2007; Yang et al., 2006). The ligase auto-ubiquitinates with K63 chains not leading to proteasomal degradation (Scialpi et al., 2008).

Smurf1 and **2** are best known for participating in the regulation of several R-Smads (1/5/8 and 2/3 respectively) and the I-Smads (6/7), the intracellular messenger proteins of the BMP and TGF- β signalling pathways (Lin et al., 2000; Zhu et al., 1999). Additionally, Smad7 recruits Smurf2 to form a stable complex with TGF- β receptors, resulting in its ubiquitin-mediated lysosomal degradation (Kavsak et al., 2000). The substrate recognition capabilities of Smurf1 are enhanced by the adaptor protein CKIP-1 (Casein Kinase 2-interacting protein 1) (Lu et al., 2008), while substrate preferences of this ligase are switched by PKA phosphorylation (Cheng et al., 2011). In both cases, inactive conformations have been described involving contacts between the HECT and the C2 domains, although these are intra-molecular in the case of Smurf2 and inter-molecular in the case of Smurf1 (Wan et al., 2011; Wiesner et al., 2007). Binding to their respective Smad7 and Cdh1 substrates triggers the transition to the active form. Both ligases control their turnover rates through auto-ubiquitination.

Both **WWP1** and **2** are also involved in TGF- β signalling regulation. WWP1 targets the TGF- β receptor using Smad7 as a scaffold, while WWP2 targets Smad2/3 and Smad7 (An et al., 2014). In addition, WWP1 targets LATS1 in the Hippo/LATS tumour suppressor signalling pathway that controls organ size (Yeung et al., 2013), while WWP2 targets PTEN that is one of the most affected tumour suppressors in a variety of cancer types (Maddika et al., 2011). Several isoforms of both ligases are expressed that display different substrate specificities (Flasza et al., 2002; Soond and Chantry, 2011). A closed conformation of WWP2 established through C2-HECT contacts has been described that is released by competition with incomplete isoforms. Both of them display auto-ubiquitinating activity.

NedL1 and **2** have been related with the stabilization of the tumour suppressors p53 and p73 respectively (Li et al., 2008; Miyazaki et al., 2003). NedL1 also targets DVL1 (dishelved-1), involved in neuroblast specification (Miyazaki et al., 2004). Their regulatory mechanisms remain unclear.

Nedd4-1, or simply **Nedd4**, is a positive regulator of IGF-1 (insulin-like growth factor-1) signalling by down-regulating c-Cbl, a RING ubiquitin ligase that targets the receptor (Cao et al., 2008). However, Nedd4 directly targets the IGF-1 receptor in neurons under stress and may be involved in neurodegeneration (Kwak et al., 2012).

It has also been related to PTEN down-regulation, although this is questioned (Fouladkou et al., 2008; Wang et al., 2007). A closed conformation mediated by C2-HECT contacts keeps the ligase latent and is released by Ca^{2+} and NDFIP1/2 binding (Wang et al., 2010).

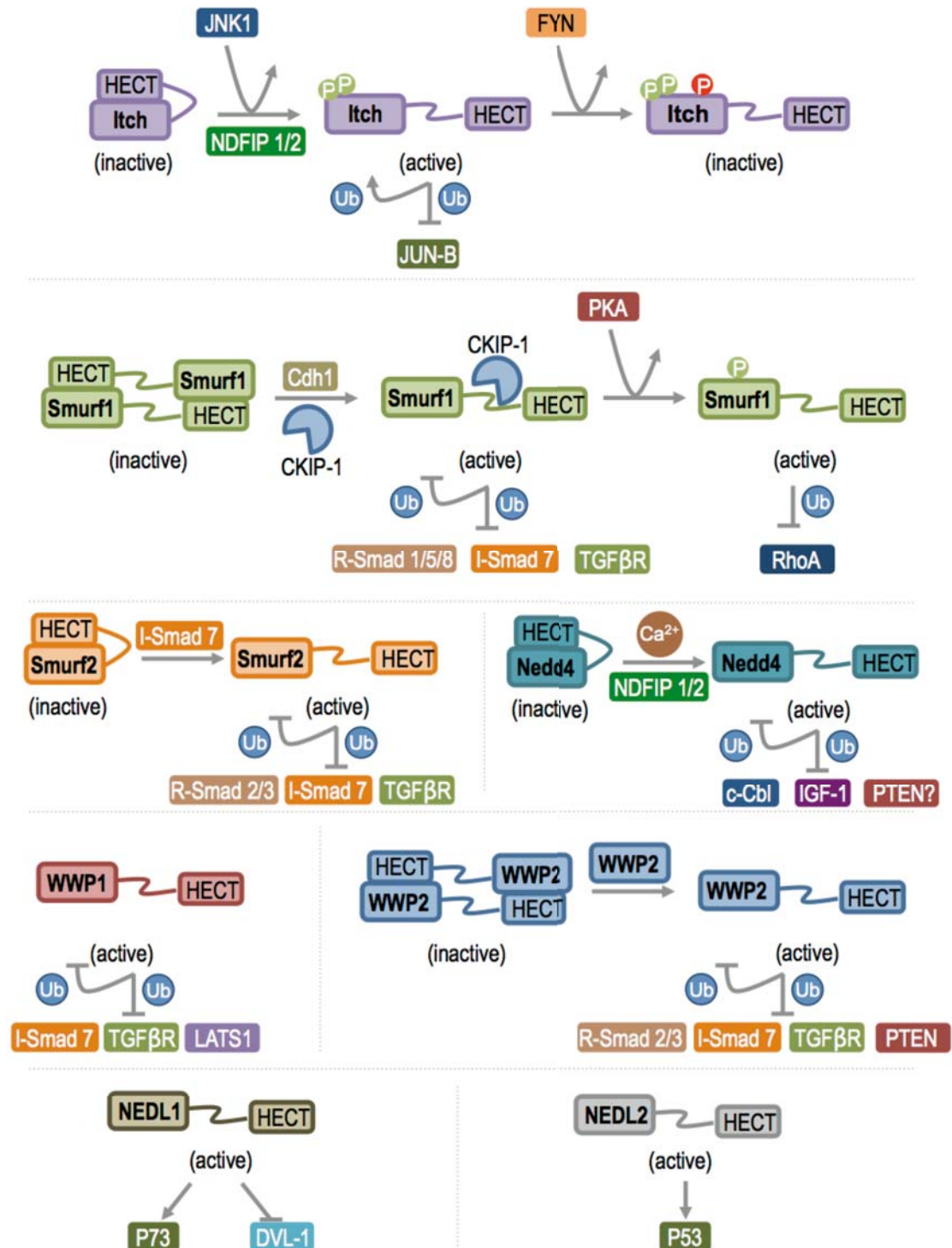


Figure 1.6: Function and regulation of E3 ubiquitin ligases in the Nedd4 family. Major targets and regulatory mechanisms for each ligase are summarized in a schematic manner. Arrows indicate activation, enhancement or stabilization. Flat lines indicate degradation.

1.1.5.3. Nedd4L

Nedd4L (or Nedd4-2) was primarily discovered to ubiquitinate a variety of membrane ion channels (**Figure 1.7**) including ENaC (Na⁺, lung and kidney), hERG (K⁺, heart and brain (Albesa et al., 2011)) and CFTR- Δ F508 (mutant version of CFTR found in cystic fibrosis, Cl⁻, lung (Caohuy et al., 2009)). The epithelial sodium channel ENaC plays a fundamental role in blood, epithelial and epithelial fluid homeostasis through the reabsorption of Na⁺ ions in the kidneys, lungs and colon. Mutant versions of ENaC affecting its cytoplasmic PPxY motifs impair recognition by Nedd4L and lead to the rare heritable form of hypertension known as Liddle's syndrome (Harvey et al., 2001; Kamynina et al., 2001; Staub et al., 1996). The ligase has also been related to cystic fibrosis, where it targets the mutant version of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR- Δ F508) which is found in 70-90% of the cases and whose down-regulation provokes the over-expression of ENaC by an unknown mechanism, contributing to lung inflammation through the dehydration of the airway. Treatment of CFTR- Δ F508 mutant CFPAC-1 cells with the glucocorticoid dexamethasone results in the activation of SGK1 (glucocorticoid-induced kinase 1), which down-regulates Nedd4L and restores the normal levels in the plasma membrane of CFTR and ENaC (Caohuy et al., 2009).

More recently, Nedd4L has been discovered to bind and ubiquitinate the TGF- β signalling mediator proteins Smad2 and 3 (Gao et al., 2009), as well as the inhibitor Smad7 that also acts as a scaffold for the ligase to target the TGF- β receptor (Aragón et al., 2012; Kuratomi et al., 2005). The structural bases of Smad recognition by Nedd4L as well as Smurf1 have been recently characterized in our laboratory (CHAPTER 4) (Aragón et al., 2011).

Nedd4L is regulated at different levels. Eight different isoforms of the ligase are expressed in cells, with two of them lacking the C2 domain (affecting sub-cellular localization (Garrone et al., 2009)) and another two lacking WW2. Phosphorylation by SGK1 (serum/glucocorticoid-regulated kinase 1) and other kinases at positions S221, T246 and S327 (in the WW domains linker regions) is accompanied by binding of the accessory protein 14-3-3 that inactivates the ligase (Ichimura et al., 2005). Differently, phosphorylation by JNK-1 stimulates the ligase's specificity for DVL-2 (dishelved 2), involved in WNT signalling (An et al., 2014; Ding et al., 2013).

Similar to other ligases in the family, a closed conformation has been identified involving HECT and C2 domain contacts (Wang et al., 2010). The interaction is released by Ca²⁺, which also plays a role in relocating Nedd4L from the cytoplasm to intracellular membranes (where some of its substrates are located) through C2

interactions with phospholipids such as PIP_2 (phosphatidylinositol 4,5-bisphosphate) (Garrone et al., 2009), a common membrane anchor for the lipid binding C2 and PH domains (Section 1.2.2). As part of this thesis, we have explored the structural bases behind this regulatory mechanism (CHAPTER 5).

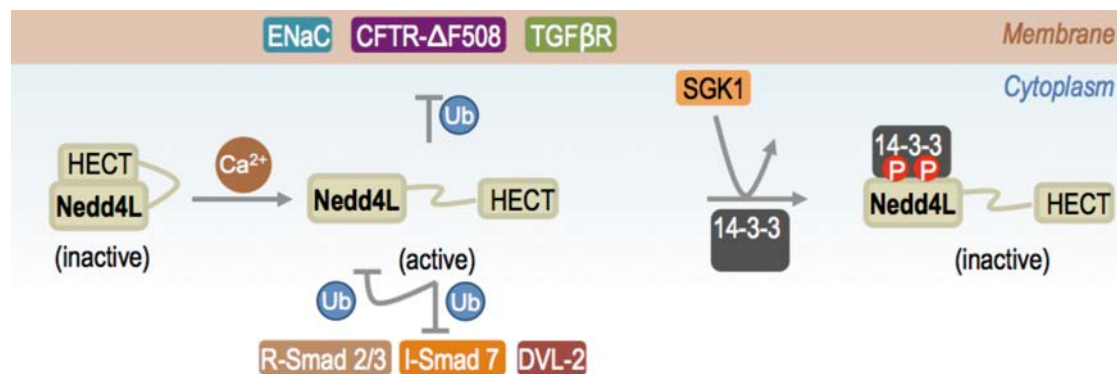


Figure 1.7: Function and regulation of Nedd4L. The closed and inactive conformation of Nedd4L is established through HECT-C2 contacts. Ca^{2+} activates the ligase and facilitates its relocation to the plasma membrane, where it targets several ion channels and the TGF- β cytokine receptor. In the cytoplasm, the active ligase targets the Smad proteins and auto-ubiquitinates. SGK-1 phosphorylation and 14-3-3 binding inactivate the ligase.

Nedd4L also undergoes auto-ubiquitination, which labels it for proteasomal degradation (Mund and Pelham, 2009; Wiesner et al., 2007). Similar to all other ligases of the family, it has a LPxY motif that is partially buried in the HECT C-lobe structure. Its WW domains have been shown to recognize this motif requiring the previous unfolding of the HECT domain (Bruce et al., 2008). We have found that the exposure of this motif can occur at the early stages of protein destabilization, and that WW3 recognizes it with low-micromolar affinity (CHAPTER 6). Our data support the existence of a bi-molecular crossed ubiquitination mechanism where a fully folded Nedd4L molecule targets a damaged counterpart.

1.2. Ca^{2+} signalling

1.2.1. Mechanisms and biological impact

Calcium ions (Ca^{2+}) play a fundamental role in cell biology as key signalling agents regulating a wide variety of processes. Cells invest considerable amounts of energy on keeping cytosolic Ca^{2+} levels low, with basal values normally found around 100 nM. Extracellular Ca^{2+} levels are often as high as 1 mM (Clapham, 2007). High cytoplasmic concentrations of Ca^{2+} are toxic and can lead to cell death both by necrosis and apoptosis. Therefore, intracellular Ca^{2+} levels are tightly controlled by several mechanisms such as chelation, extrusion outside the cell and ATPase-

mediated compartmentalization (mainly in the endoplasmic reticulum (ER) and to a lesser extent in the mitochondria).

1.2.1.1. Signal initiation and cascade

Transient increases of cytoplasmic Ca²⁺ levels are initiated in response to external stimuli (**Figure 1.8**). Many receptors on the cell surface that are responsive to a wide variety of ligands are coupled to the phospholipase C (PLC) family (Berridge et al., 2000). This enzyme is responsible for the cytoplasmic delivery of inositol 1,4,5-trisphosphate (IP₃), which is able to diffuse and mobilize the internal Ca²⁺ stores in the ER (Burgess et al., 1984; Streb et al., 1983). IP₃ is the polar, soluble moiety of the membrane phospholipid phosphatidylinositol 4,5-bisphosphate (PIP₂) whose hydrolysis also yields diacylglycerol (DAG). These receptors are classified in two main groups, depending on whether they are coupled to G protein (GPCRs) or linked to tyrosine kinases (RTKs). They vary extensively depending on the cell type, with GPCRs including the histamine and the oxytocin receptors and RTKs including the epidermal growth factor and the T and B-cell receptors.

Other Ca²⁺-mobilizing secondary messenger molecules exist although they have been characterized to a lesser extent, including cyclic ADP ribose (cADPR) generated by ADP ribosyl cyclase that is also activated by surface receptors (Clapper et al., 1987).

Receptors for these secondary messengers are found on the surface of the ER. Two main classes of such receptors have been characterized: the IP₃ receptors (IP₃R) and the ryanodine receptors (RyR) (Berridge, 1993). Both of them act as Ca²⁺ channels gating its delivery from the ER to the cytoplasm. In both cases, Ca²⁺ itself has a role in activating such delivery through a mechanism that is central in Ca²⁺ signalling and that is known as **Ca²⁺-induced-Ca²⁺-release**. Both IP₃ and Ca²⁺ are required to activate IP₃R (Finch et al., 1991), whereas RyR are activated by Ca²⁺ and modulated by cADP. High cytoplasmic concentrations of Ca²⁺ block the channels.

IP₃R and RyR are organized in small clusters and act in an inter-dependent and coordinate manner. The coordinate opening of the channels in a cluster is known as *puff* for IP₃R and *spark* for RyR (Cheng et al., 1993; Yao et al., 1995). This results in the delivery of small amounts of Ca²⁺ into the cytoplasm, which initiates a cascade by activating the receptors in neighbouring clusters. Therefore, IP₃, cADPR and Ca²⁺ signalling contribute in an interrelated manner to the cytoplasmic release of Ca²⁺.

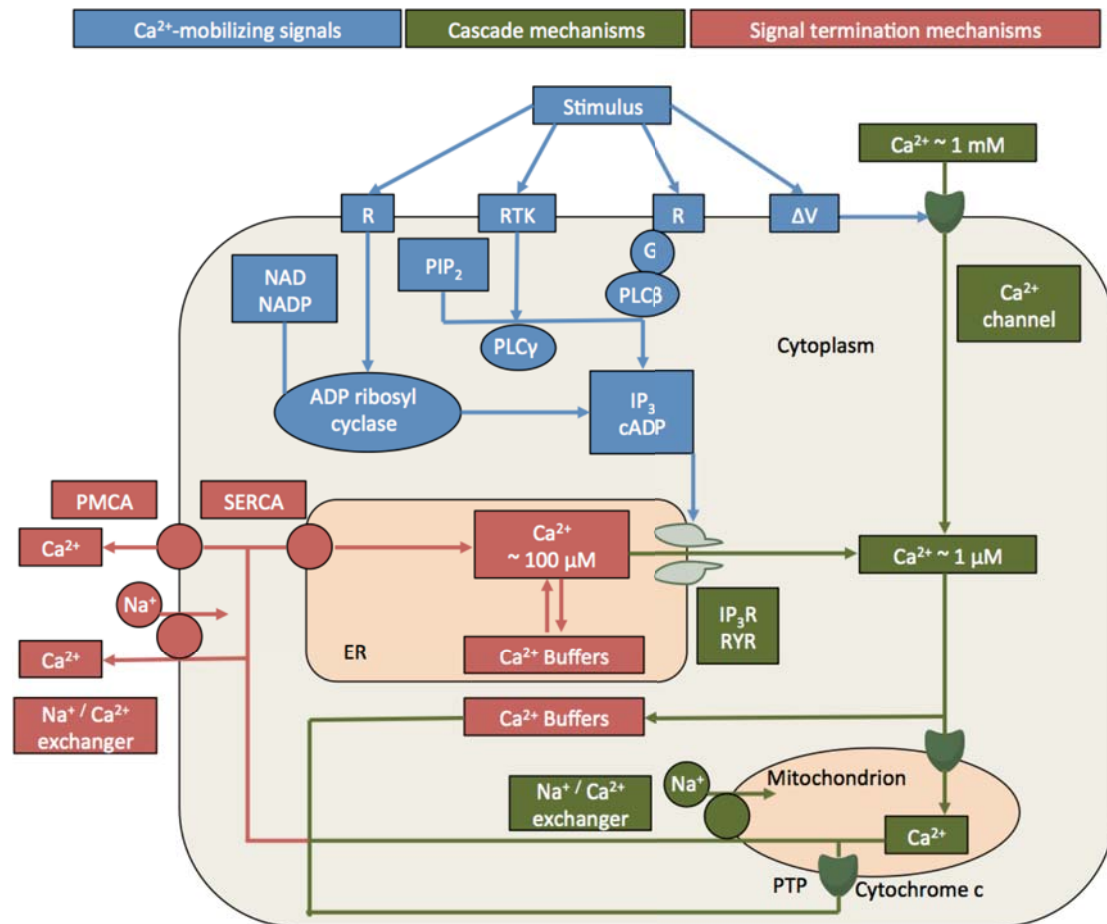


Figure 1.8: Major mechanisms involved in Ca^{2+} signalling. Adapted from (Berridge et al., 2000). External stimuli activate the Ca^{2+} -mobilizing signals (in blue), which are the seeding events for the signalling cascade (green). After that, cytoplasmic Ca^{2+} concentration is raised to about $1 \mu\text{M}$. The signal termination mechanisms (red) restore the basal Ca^{2+} concentration (about 100 nM).

After Ca^{2+} mobilizing signals have been generated, a cascade of events creates a transient peak of intracellular Ca^{2+} concentration. Apart from Ca^{2+} release from the ER, extra-cellular Ca^{2+} enters the cell through channels in the plasma membrane. These include voltage-operated channels (VOC) responsive to membrane depolarization, receptor-operated channels (ROC) that open in response to neurotransmitters such as ATP and acetylcholine, and store-operated channels (SOC) that are activated by internal store depletion (Berridge et al., 2000).

As a global result, cytoplasmic Ca^{2+} levels are transiently raised up to $1 \mu\text{M}$, which is rapidly lowered by several means.

1.2.1.2. Signal termination

Ca^{2+} -buffering proteins such as calbindins and parvalbumin chelate most of the Ca^{2+} that enters the cytoplasm. The sarco-endoplasmic reticulum ATPase (SERCA)

returns Ca²⁺ to the ER. In the plasma membrane, the Na⁺/Ca²⁺ exchanger (NCX) and the plasma membrane Ca²⁺-ATPase (PMCA) extrude Ca²⁺ outside of the cell (Blaustein and Lederer, 1999; Pozzan et al., 1994).

Mitochondria play an important role after Ca²⁺ signal initiation by rapidly up-taking Ca²⁺ after its cytoplasmic release and slowly delivering it back when basal levels have been restored (Budd and Nicholls, 1996; Duchen, 1999). Ca²⁺ uptake is coupled to the generation of the H⁺ gradient that will permit the synthesis of ATP, which is further enhanced by Ca²⁺ activation of mitochondrial enzymes including those responsible for ATP synthesis. A Na⁺/Ca²⁺ exchanger mediates the slow return of Ca²⁺ to the cytoplasm. Ca²⁺ overload of the mitochondria results in the cytoplasmic delivery of cytochrome c, which activates the caspase-mediated apoptotic pathway.

1.2.1.3. Signal modulation and Ca²⁺-sensitive processes

Ca²⁺ signals are modulated in amplitude (amount of Ca²⁺ delivered) but also in frequency (Berridge et al., 2000). Some cellular responses such as the contraction of skeletal muscle require individual Ca²⁺ concentration spikes, but more prolonged signalling is achieved through consecutive spikes at different frequencies ranging from 1-60 sec (as in the liver and the pancreas) to 24 hours (in mitosis initiation). Ca²⁺/calmodulin-dependent protein kinase II (CAMKII) and protein kinase C are the main proteins decoding these frequency-modulated signals (De Koninck and Schulman, 1998; Oancea and Meyer, 1998). Varying the frequency of these spikes permits the cell to adapt to different stimuli intensities.

The complexity of Ca²⁺ signalling is the basis for its involvement in fundamental life processes including fertilization, embryonic pattern formation, cell differentiation and proliferation, exocytosis in leukocytes, aggregation of blood platelets, neuronal synapse, fluid secretion in the gut, pancreas and salivary gland, contraction of the smooth, skeletal and cardiac muscle and liver metabolism (Berridge, 2009). Many of them are mediated by calmodulin, a ubiquitous protein that exposes a hydrophobic surface upon Ca²⁺ binding able to interact and activate a wide variety of effectors including protein kinases, transcription factors, ion channels and many others. Others are mediated by the membrane phospholipid PIP₂, which is also ubiquitous and abundant and whose negatively charged cytoplasmic head is able to recruit positively charged and Ca²⁺-binding proteins to the plasma membrane (Clapham, 2007).

1.2.2. Ca²⁺-mediated membrane relocation of C2 domains

Many proteins are localized to the plasma membrane through phospholipid-interacting modules such as the PH (pleckstrin homology) and C2 domains (Cho and Stahelin, 2006; Lemmon, 2008). The PH domains have high specificity for PIP₂, which they bind in a Ca²⁺-independent manner and whose cleavage in the Ca²⁺ signalling pathway moves proteins with PH domains away from the membrane. A positively charged patch in the domain mediates the interaction with the cytoplasm-exposed moiety of PIP₂, which is negatively charged and chemically identical to IP₃ (**Figure 1.9A** and B) (Hyvonen et al., 1995).

The C2 domains bind a wider range of phospholipids including phosphatidylserine, phosphatidylcholine and PIP₂. Some of them are able to do it constitutively in a Ca²⁺-independent manner also using a positively charged pocket in their variable loop region. This is the case of the C2 domain of Rsp5, the *S. cerevisiae* orthologous for human Nedd4 family ligases, which uses five lysine residues in the loop regions to bind PIP₂ (Dunn et al., 2004).

However, the main family of C2 domains, i. e. those similar to the C2 domain of protein kinase C (PKC), are able to bind two or three Ca²⁺ ions in their variable loop regions that are fundamental in the interaction with the membrane phospholipids (**Figure 1.9C**) (Murray and Honig, 2002; Verdaguer et al., 1999). This is the case of the C2 domains found in human members of the Nedd4 family of ubiquitin ligases (Garrone et al., 2009; Plant et al., 1997). The Ca²⁺-coordinating residues form a negatively charged pocket where Ca²⁺-independent C2 domains have their positively charged patch, avoiding their constitutive binding to the membrane (Lemmon, 2008). In these cases, Ca²⁺ binding following a Ca²⁺ concentration spike in the cytoplasm acts as an effective switch creating a positively charged patch on the C2 domain surface that permits its interaction with membrane phospholipids including PIP₂.

Many C2 domains have a second Ca²⁺-independent basic binding pocket in their concave surface known as the cationic β-groove, which is able to bind phospholipids as well (Guerrero-Valero et al., 2009) (**Figure 1.5**).

In the case of Nedd4L, Ca²⁺ binding has been reported to activate the ubiquitin-ligase activity of the enzyme by triggering its transition from the closed to the open conformation (Wang et al., 2010). This event is accompanied by the re-localization of the ligase from the cytoplasm to the plasma membrane. We hypothesized that Ca²⁺ binding switches the binding preferences of the C2 domain of Nedd4L from contacting the HECT domain (as in the closed conformation) to binding IP₃, the

cytoplasm-exposed moiety of PIP₂, which plays itself a central role in the initiation of Ca²⁺ signals. Structural evidences supporting this hypothesis are reported in CHAPTER 5.

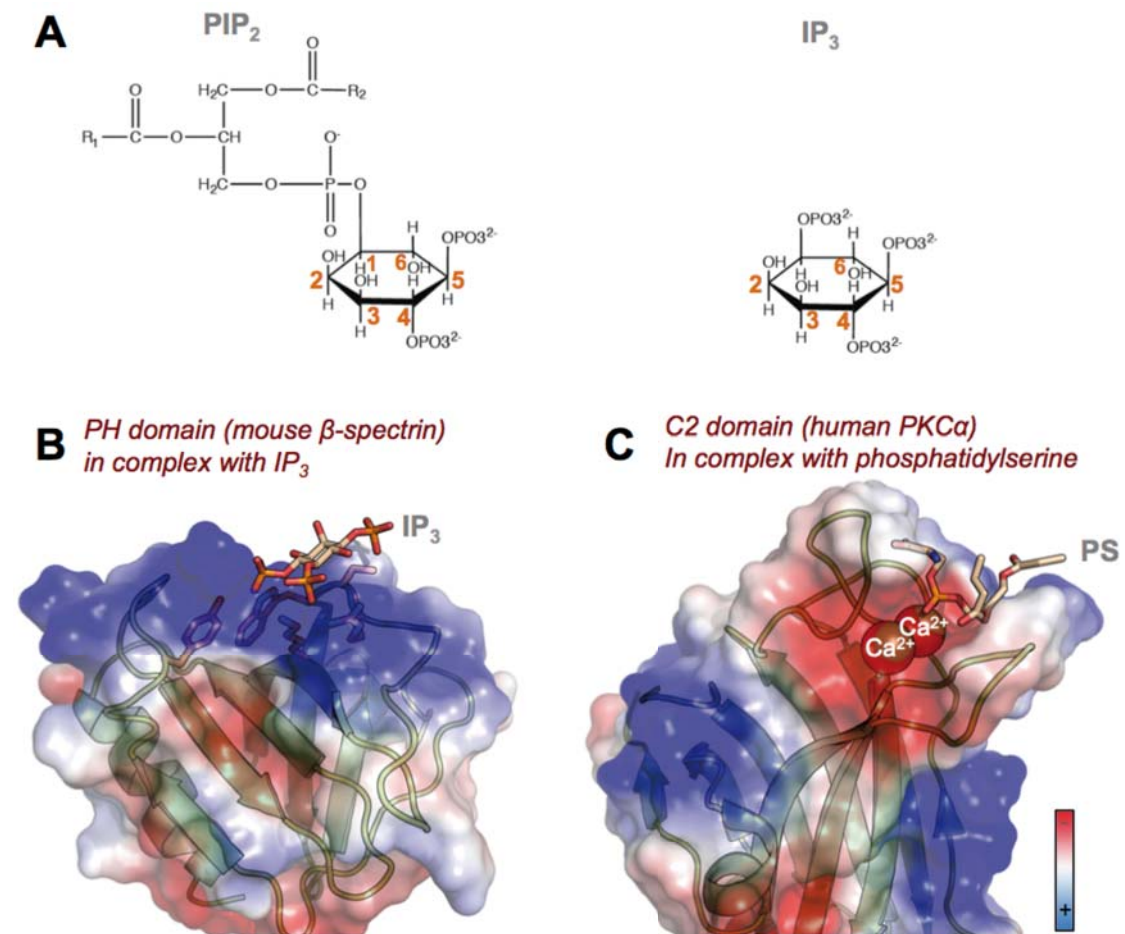


Figure 1.9: Ca²⁺ bridges the interaction with membrane phospholipids in PKC α -like C2 domains. (A) Chemical formula of the membrane phospholipid PIP₂ and the soluble IP₃ resulting from its cleavage by PLC. PH and C2 domains interact with PIP₂ through its inositol moiety as well as with IP₃, which are chemically identical. (B) Structure of the PH domain of mouse β -spectrin bound to IP₃ (1BTN) (Hyvonen et al., 1995). PH domains constitutively bind PIP₂ and IP₃ through a positively charged patch on their surface. (C) Structure of the C2 domain from human PKC α bound to phosphatidylserine (1DSY) (Verdaguer et al., 1999). PKC α -like domains have a negatively charged patch able to coordinate two or three Ca²⁺ ions, which bridge the interaction with membrane phospholipids.

1.3. TGF- β signalling

1.3.1. Biological implications

Transforming growth factor- β (TGF- β) is a family of cytokines that emerged with the first animal species as a means to control metazoan multicellular life (Huminiacki et al., 2009). TGF- β signalling plays a central role in processes such as

cell proliferation and differentiation as well as tissue and organ morphogenesis, homeostasis and regeneration (Massagué, 2000). These cellular events are largely involved in systemically fundamental processes including the differentiation of stem cells, immunity, inflammation, tumour suppression and also haematopoietic, vascular, bone, adipose and muscle tissue homeostasis (Shi and Massagué, 2003). Thus, malfunction of components involved in TGF- β signalling has been widely related to many severe diseases, specially cancer and metastasis.

Up to 42 proteins have been identified as TGF- β cytokines in humans, whereas 9 of them have been found in *D. melanogaster* and 6 in *C. elegans* (Lander et al., 2001). Depending on sequence similarity and the specific pathway they activate, they are classified in two main subfamilies, namely the TGF- β /Activin/Nodal and the BMP/GDF/MIS (bone morphogenic protein/growth and differentiation factor/Muellerian inhibiting substance) subfamilies.

1.3.2. Mechanism

TGF- β cytokines determine cell fate by directly regulating gene expression, which can be enhanced or repressed in function of the target gene and the cellular context (Massagué, 2012). The number of regulated genes varies depending on the cell type and developmental state, ranging from a few ones in pluripotent stem cells to several hundreds in differentiated ones (Mullen et al., 2011; Trompouki et al., 2011). Moreover, different members of the family elicit different cell responses.

Notwithstanding that, all family members share a common signal transduction scheme that presents slight variations in each particular case (**Figure 1.10**). TGF- β cytokines form homo-dimers bound through hydrophobic interactions and a characteristic disulphide bridge (Shi and Massagué, 2003). Plasma membrane receptors for these cytokines have cytoplasmic serine/threonine kinase domains and are divided in two types (I and II) that form hetero-dimers upon ligand binding. Two type I and two type II receptors are recruited at a time to bind the ligand dimer. Cytokines in the TGF- β /Activin/Nodal subfamily bind only type II receptors, and type I molecules are recruited after binding (Massagué, 1998). Differently, ligands of the BMP/GDF/MIS subfamily can bind both receptor types in their isolated forms, although the affinity for type I receptors is higher. Upon ligand binding and receptor docking, type II receptor phosphorylates the N-terminal GS domain (SGSGSG sequence) in type I receptor, which activates the adjacent kinase domain.

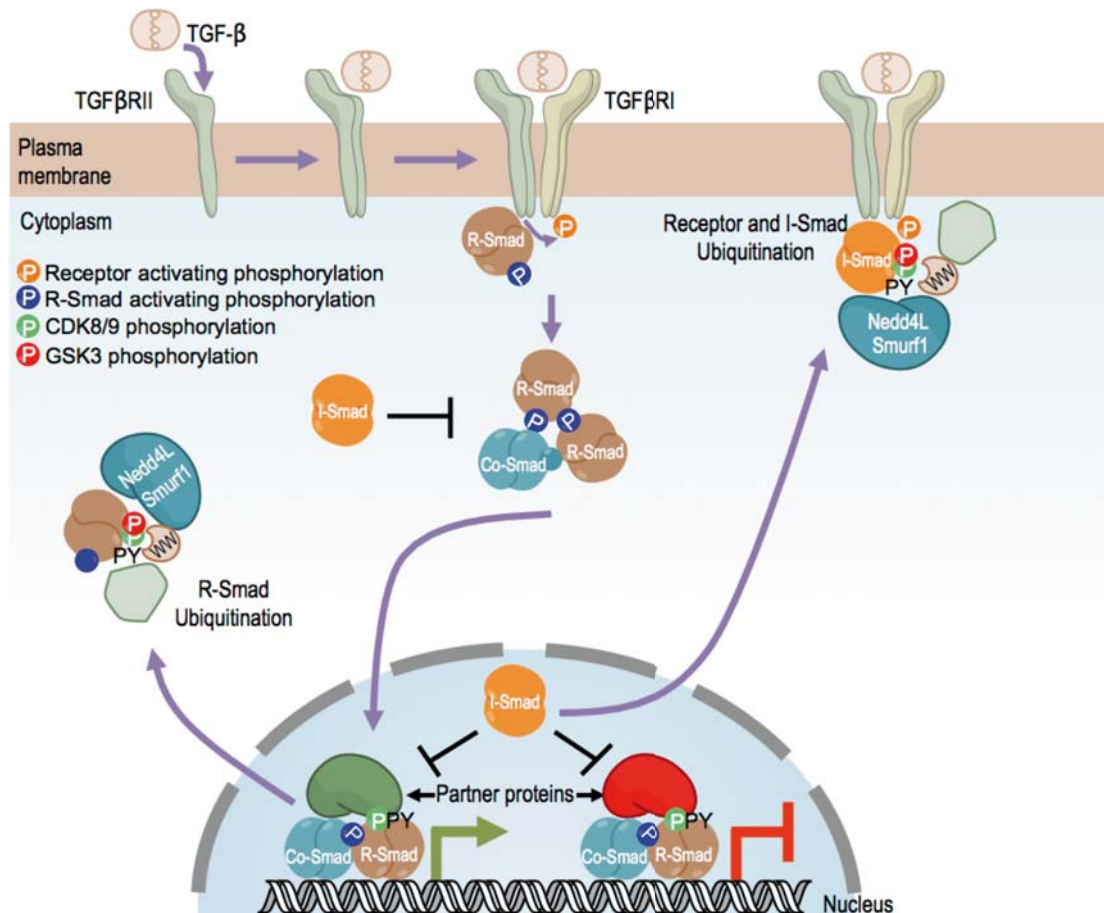


Figure 1.10: General mechanism in TGF- β signalling. The binding of TGF- β /BMP cytokines to their receptor triggers the activation of the pathway. The signal is propagated in the cell through the action of the R-Smad proteins in association with the Co-Smad. In the nucleus, they associate with partner proteins to promote or repress gene expression. The I-Smads provide negative regulatory feedback on the pathway. The signal is terminated through the action of HECT-type E3 ubiquitin ligases including Nedd4L and Smurf1 and 2.

The signal is then propagated within the cell by the recruitment and C-terminal phosphorylation of the intracellular messenger Smad (small mothers against decapentaplegic) proteins, which can be reverted by certain phosphatases that provide a negative feedback control mechanism of the pathway (Lin et al., 2006). Eight different Smad proteins have been identified that are classified into receptor-activated (R-Smads) common-partner (Co-Smad) and inhibitor (I-Smads) depending on their function. Only the R-Smads, which include Smads 1, 2, 3, 5 and 8, have a Ser-X-Ser motif (where X is any amino acid) suitable for type I receptor phosphorylation. Smads 1, 5 and 8 are activated by receptors of the BMP/GDF/MIS subfamily of cytokines, whereas Smads 2 and 3 bind to receptors of the TGF- β /Activin/Nodal subfamily. Once activated, the R-Smads form trimers with a Co-Smad moiety (Smad4) and are shuttled into the nucleus (Inman et al., 2002; Xu et al., 2002), where they associate with accessory proteins and modulate gene expression. The I-Smads provide a negative feedback by opposing TGF- β signalling in different

manners: Smad6 resides in the cytoplasm and competes with Smad1 for the recruitment of Smad4 (Hata et al., 1998), whereas Smad7 sequesters some of the R-Smads nuclear partners such as YAP (Aragón et al., 2012). Smad7 can also re-localize from the nucleus to the cytoplasm upon TGF- β activation, bind the activated type I receptors and recruit the HECT ubiquitin ligases Smurf1, Smurf2 and Nedd4L that label both the receptor and Smad7 for proteasomal degradation (Ebisawa et al., 2001; Gao et al., 2009; Itoh et al., 1998).

1.3.3. Linker phosphorylation couples Smad action and disposal

The Smads are modular proteins with R- and Co-Smads presenting two MH (MAD-homology) domains connected through a linker region (**Figure 1.11**) (Shi and Massagué, 2003). In most cases, the MH1 domain binds DNA in a sequence specific manner. The MH2 domain is highly conserved and mediates receptor interaction, trimerization and nuclear translocation. The inter-domain linker region has a cluster of phosphorylation S/TP sites and an adjacent PPxY motif that enables the interaction with WW domains in partner proteins.

Phosphorylation of the S/TP cluster in the linker region of R-Smads creates docking sites for both co-activators and repressors of TGF- β signalling. In the cytoplasm and in response to stress, mitogen-activated protein kinases (MAPKs) deplete the pool of inactive R-Smads by creating a suitable binding site for E3 ubiquitin ligases (Massagué, 2003).

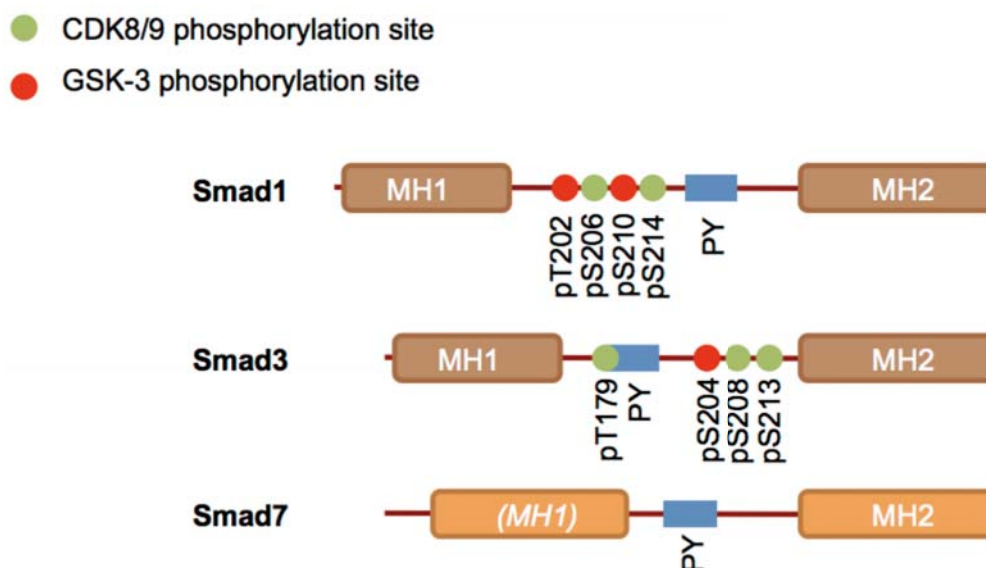


Figure 1.11: Modular architecture of the Smad proteins. Smad proteins consist of two MH domains connected by a linker region. The MH1 domain in Smad7 displays low homology with all others and is shown in italics and in brackets. All Smads have a PY box in their linker region. R-Smads have phosphorylation sites for CDK8/9 and GSK3- β kinases.

Activated R-Smads in the nucleus receive a second round of phosphorylation events by the cyclin-dependent kinases 8 and 9 (CDK8/9), members of the RNA polymerase II regulatory complex (Alarcón et al., 2009; Matsuura et al., 2004; Matsuzaki et al., 2009). This is a priming event for glycogen synthase kinase 3- β (GSK3- β) phosphorylation (Fuentealba et al., 2007; Sapkota et al., 2007). Four different proteins have been described to specifically bind the phosphorylated Smad linker, all of them using WW domains (**Figure 1.12**). Two of them act as TGF- β signalling co-activators: the peptidyl-prolyl *cis/trans* isomerase Pin1 stimulates cancer cell migration in association with Smad2/3 (Matsuura et al., 2010), whereas the Hippo pathway transcription factor YAP (Yes-associated protein) associates with Smad1 in mouse embryonic stem cells to suppress neural differentiation (Alarcón et al., 2009). The other two proteins are the HECT E3 ubiquitin ligases Smurf1, which binds Smad1/5 (Sapkota et al., 2007) and Nedd4L, which binds Smad2/3 (Gao et al., 2009). Both of them act as TGF- β signal terminators by labelling the Smad proteins for proteasomal degradation.

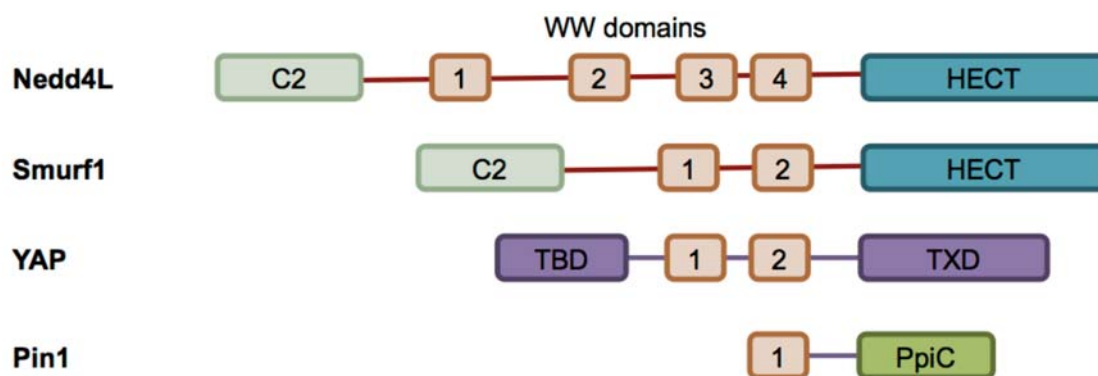


Figure 1.12: Modular architecture of the Smad binding partners. All of them bind the phosphorylated linker region in R-Smads using their WW domains. YAP has a TEAD-binding domain (TBD), two WW domains and a transactivator domain (TXD). Pin1 has a WW domain and a peptidyl-prolyl *cis/trans* isomerase domain.

Thus, nuclear phosphorylation events give rise to a scenario where transcriptional factors and ubiquitin ligases compete to bind the Smads and govern their fate, raising the question on the mechanism permitting Smad action before disposal. In collaboration with the group of Dr Joan Massagué*, we provided biological and structural evidence supporting a timely hierarchical mechanism where CDK8/9 phosphorylation events favour Smad interactions with TGF- β signal co-activators, whereas the subsequent GSK3- β phosphorylation switches binding preferences favouring the ubiquitin ligases (Aragón et al., 2011). Such a mechanism represents a breakthrough in the substrate selectivity mechanisms of Smurf1 and

* Dr Joan Massagué, Cell Biology Program and Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, 1,275 York Avenue, New York 10021.

Nedd4L as well as YAP, as it relies on the synchronized recognition by WW tandems of the PY site and the phospho-S/TP cluster, an arrangement never reported before involving type I WW domains that were not expected to bind phosphorylated sequences.

Part II.

Aim and Objectives

CHAPTER 2. THESIS AIM AND OBJECTIVES	43
2.1. Background and scope	43
2.2. Thesis objectives	44

CHAPTER 2. Thesis aim and objectives

2.1. Background and scope

In humans, the HECT-type E3 ubiquitin ligase Nedd4L regulates the turnover rates of key proteins involved in basic processes with systemic implications. The ligase was first discovered to target the epithelial sodium channel ENaC, which is fundamental in Na⁺ ions homeostasis in blood, epithelial fluid and epithelia. Impaired recognition of ENaC caused by a genetic mutation affecting its cytoplasmic PY motifs leads to the rare form of hereditary hypertension known as **Liddle's syndrome** (Alarcón et al., 2009; Staub et al., 1996). The ligase has also been related to **cystic fibrosis**.

More recently, Nedd4L has been found to be involved in the regulation of TGF- β components, including the down-regulation of TGF- β receptors and the intracellular signal mediator proteins Smad2, 3 and 7 (Aragón et al., 2012; Gao et al., 2009). The signalling pathway activated by TGF- β cytokines governs many aspects of cell and tissue differentiation, morphogenesis, homeostasis and regeneration both during development as well as in adults. Malfunctions in this pathway have been related to several types of **cancer** and **metastasis**, among others (Massagué, 2012).

The important biomedical implications of Nedd4L encouraged us to pursue the study of its regulation and substrate selectivity. Similar to other closely related proteins, the ligase adopts a closed and inactive conformation mediated by inter-domain contacts that protects its substrates from unspecific ubiquitination. An increase of intracellular Ca²⁺ levels triggers the transition of the ligase to the open and active form and permits its relocation to the plasma membrane (Wang et al., 2010). When active, the ability of the ligase to bind its TGF- β -related substrate Smad3 and regulate its turnover depends on the target's phosphorylation state (Gao et al., 2009). Like other E3 ubiquitin ligases, active Nedd4L regulates its own turnover rate through auto-ubiquitination as a negative feedback mechanism to modulate its activity (Mund and Pelham, 2009; Wiesner et al., 2007).

Models for the conformational activation, Smad3 binding and ubiquitination of Nedd4L have been inferred from biochemical and molecular biology experiments. The aim of this thesis is to contribute to the understanding of these processes from a structural and biophysical perspective that helps unveiling their mechanistic bases at the atomic resolution.

2.2. Thesis objectives

A multi-disciplinary approach combining structural biology, organic chemistry and molecular biology has been used in the experimental part of this work. To achieve our aim we have divided the work in four objectives:

Objective 1 – Synthesis of a set of R-Smad linker peptides representing different phosphorylation states to study substrate selectivity in Nedd4L and related proteins.

The ability of partner proteins containing WW domains to bind the R-Smads depends on the phosphorylation state of the linker region connecting their MH domains (Massagué, 2012). When R-Smads are shuttled into the nucleus CDK8/9 phosphorylate this linker region. In this situation, Pin1 and YAP cooperate with Smad3 and Smad1 respectively in signal transduction. CDK8/9 primes the phosphorylation of the linker region by GSK3- β , which switches the R-Smad preferred binding partners in favour of the ubiquitin ligases Nedd4L (Smad3) and Smurf1 (Smad1) resulting in their degradation. In the frame of a study aimed at unveiling the structural and biophysical bases of substrate recognition by these R-Smad associated proteins, my contribution was focused in the synthesis of a set of Smad1 and Smad3 linker phosphopeptides representing different phosphorylation states using solid-phase peptide synthesis (SPPS).

Objective 2 - Structural characterization of the C2-HECT inter-domain contacts in the latent conformation of Nedd4L and the role that IP₃/Ca²⁺ interrelated levels play in the activation of the ligase.

Contacts involving the Ca²⁺-binding C2 domain and the catalytic HECT domain have been shown to maintain Nedd4L in a latent state (Wang et al., 2010). An increase of cytoplasmic Ca²⁺ levels renders the ligase active. This also permits the relocation of the ligase to the plasma membrane, where its C2 domain binds different phospholipids including phosphatidylinositol 4,5-bisphosphate (PIP₂) through interactions with its cytoplasm-exposed inositol moiety (Plant et al., 1997). In many cases these are Ca²⁺-dependent interactions. Indeed, phospholipase C (PLC)-mediated cleavage of PIP₂ and the subsequent cytoplasmic delivery of the resulting inositol 1,4,5-bisphosphate (IP₃) moiety is a central event leading to Ca²⁺ concentration spikes. Using nuclear magnetic resonance (NMR) spectroscopy, we planned to map the binding site of the HECT domain on the surface of the C2 domain in the frame of the closed and latent conformation of Nedd4L, and to evaluate the negative effect of Ca²⁺ in this interaction. Ca²⁺ also binds the surface of

the C2 domain, so we decided to study this binding by NMR and isothermal titration calorimetry (ITC) to find out the structural bases for this negative effect. Furthermore, we hypothesized that Ca^{2+} induces the relocation of the ligase to the plasma membrane by favouring its interaction with IP_3 .

Objective 3 - Structural and biophysical characterization of the recognition of a highly conserved PY motif in the HECT domain by the WW domains in Nedd4L related to auto-ubiquitination.

Similar to other HECT-type ubiquitin ligases, Nedd4L has an LPxY motif in the HECT C-terminal lobe that presents high conservation throughout evolution. This motif is similar to the canonical PPxY sequence recognized by WW domains. The tyrosine residue of the motif, which is fundamental for WW binding, is buried in the crystal structure of the HECT domain. Indeed, unfolding of the domain was required to detect recognition of the LPxY motif by the WW domains of Nedd4L using molecular biology techniques (Bruce et al., 2008). We hypothesized that the increased accessibility of the tyrosine residue in the HECT-PY motif accompanying fold destabilization may initiate the mechanisms involved in the auto-ubiquitination of the ligase. Thus, we planned to evaluate the binding of Nedd4L WW domains to the HECT-PY motif by NMR and ITC, both in the context of the folded HECT C-lobe and when exposed to the solvent as part of a synthetic peptide. Furthermore, we set to study the inner flexibility of the HECT C-lobe by relaxation experiments and molecular dynamics simulations in order to evaluate the accessibility of the HECT-PY tyrosine residue accompanying its destabilization.

Objective 4 – Paramagnetic labelling of a WW domain

Paramagnetic tags enable the measurement of much longer distances when compared to classical NMR observables such as NOEs. Some of these tags give rise to the effect known as paramagnetic relaxation enhancement (PRE), whose measurement and transformation into distance restraints is relatively simple. Taking advantage of the wide structural characterization of the WW3 domain of Nedd4L, we planned to use it to carry out a preliminary study with the objective of exploring the application of this technique to our research interests.

Part III.

Materials and Methods

CHAPTER 3. MATERIALS AND METHODS	49
3.1. Molecular biology methods	49
3.2. Solid-phase peptide synthesis	60
3.3. Isothermal Titration Calorimetry (ITC)	67
3.4. NMR Spectroscopy of Biomolecules	70
3.5. Molecular dynamics simulations	87

CHAPTER 3. Materials and Methods

3.1. Molecular biology methods

3.1.1. Preparation of wild-type and mutant clones

3.1.1.1. Cloning

The human cDNA Nedd4L sequence inserted in a mammal cloning vector was provided by our collaborators (Joan Massagué's laboratory, Memorial Sloan Kettering Cancer Centre, NY). The gene segment codifying each individual construct was amplified from this clone by PCR (Figure 3.1A). DNA samples were analysed using agarose gel electrophoresis (Figure 3.1B).

All the domains of the human E3 ubiquitin ligase Nedd4L were cloned into *E. coli* expression vectors: the C2 (16-154 and 1-154), WW1 (193-228), WW2 (364-402), WW3 (476-519), WW4 (529-563), the HECT (574-949) and the HECT C-terminal lobe (836-949) (Figure 3.1C and Table 3.1). Residue numbering corresponds to isoform 5 of Nedd4L (Uniprot entry Q96PU5-5).

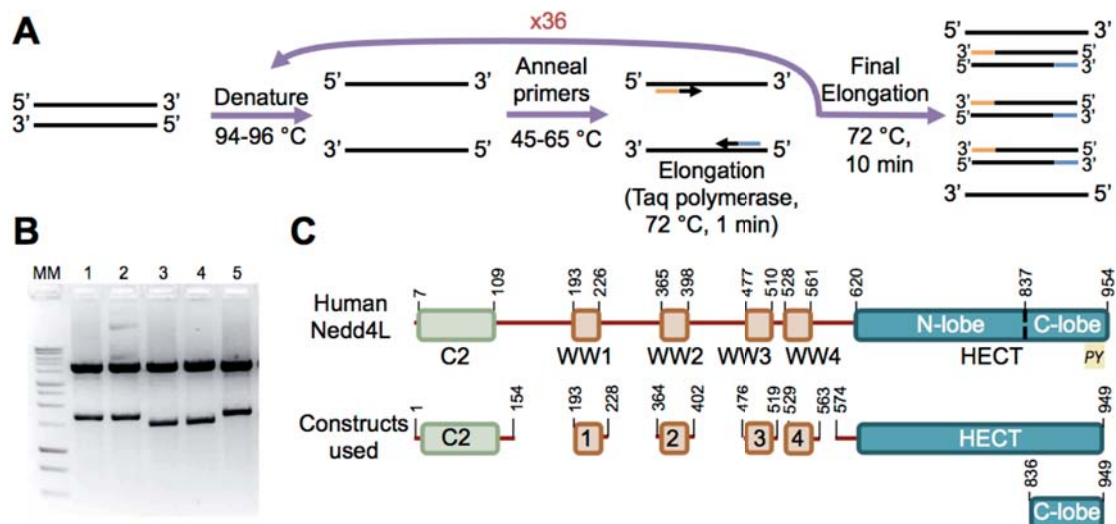


Figure 3.1: Cloning of Nedd4L constructs. (A) Details of the protocol used for the gene amplification by PCR. (B) Agarose gel electrophoresis was used to analyse DNA samples during the cloning procedure. “MM” refers to molecular markers. (C) Domain diagram showing the architecture of Nedd4L and the constructs cloned by PCR and used in this thesis.

Domain	Sequence	$\epsilon_{280 \text{ nm}}$ [mM ⁻¹ cm ⁻¹]	MW [Da]
C2	¹ GMATGLGEPV YGLSEDEGES RILRVKVVSG IDLAKKDIFG ASDPYVKLSL YVADENRELA LVQTKTIKKT LNPKNWEEFY FRVNPNSNHRL LFEVFDENRL TRDDFLGQVD ¹⁵⁴	14.440	17908.5
WW1	¹⁹³ GAMEPPLPPG WEEKVDNLGR TYYVNHNNRT TQWHRPSLMD ²²⁸	13.980	4708.2
WW2	³⁶⁴ GAMETPGLPS GWEERKDAKG RTYYVNHNNR TTTWTRPIMQ LAE ⁴⁰²	13.980	4964.5
WW3	⁴⁷⁶ GAMEQSFLPP GWEMRIAPNG RPPFFIDHNTK TTTWEDPRLK FPVHMRSK ⁵¹⁹	11.000	5656.5
WW4	⁵²⁹ GAMGPLPPGW EERIHLDGRT FYIDHNSKIT QWEDPRLQN ⁵⁶³	12.490	4576.0
HECT	⁵⁷⁴ GAMGSREFKQ KYDYFRKKLK KPADI PNRFE MKLHRNNIFE ESYRRIMSVK RPDVLKARLW IEFESEKGLD YGGVAREWFF LLSKEMFNPY YGLFEYSATD NYTLQINPNS GLCNEDHLSY FTFIGRVAGL AVFHGKLLDG FFIRPFYKMM LGKQITLNDM ESDVSEYYNS LKWILENDPT ELDLMFIDE ENFGQTYQVD LKPNGSEIMV TLENKREYID LVIQWRFVNR VQKQMNAFLE GTELLPIDL IKIFDENELE LLMCGLGDVD VNDWRQHSIY KNGYCPNHPV IQWFWKAVLL MDAEKRIRLL QFVTGTSRVP MNGFAELYGS NGPQLFTIEQ WGSPEKLPRA HTDFNRLDLP PYETFEDLRE KLLMAVENAQ ⁹⁴⁹	71.070	44855.3
HECT C-lobe	⁸³⁶ GAMGDVDVND WRQHSIYKNG YCPNHPVIQW FWKAVLLMDA EKRIPLLQFV TGTSRVPMNG FAELYGSNGP QLFTIEQWGS PEKLPRAHTD FNRLDLPPYE TFEDLREKLL MAVENAQ ⁹⁴⁹	27.960	13567.4

Table 3.1: Sequences of all the Nedd4L constructs after TEV cleavage of the His₆ tag. N-terminal residues highlighted in blue are not part of the wild-type sequence but were required for cloning. The C922D mutation in the HECT domain is highlighted in red.

In all cases, the PCR generated inserts were ligated into kanamycin resistance pET expression vectors for their production in fusion with a His₆ tag alone (pETM-11) or with an additional glutathione-S-transferase (GST) tag (pETM-30) (vectors engineered by G. Stier, EMBL, Heidelberg, Germany). These vectors have a T7/lac promoter and expression can be induced by the addition of both lactose and its analogue IPTG (isopropyl β -D-1-thiogalactopyranoside) to the culture. The expressed constructs contain a cleavable Tobacco Etch Virus (TEV) protease site between the tag and the cloned construct permitting their separation. The NcoI and

HindIII endonuclease cleavage sites were used for both insert and vector digestion. The resulting complementary sticky ends were ligated with T4 DNA ligase (Fermentas).

The C2 and the HECT domains were also cloned in fusion with the solubility tags MBP (pMAL-C2X vector (Addgene), ampicillin resistance) and NusA (pCri1-NusA vector (gift from Dr Miquel Coll, IRB Barcelona), kanamycin resistance). These vectors contain the same inducible expression system reported before.

After the target sequences were ligated into the corresponding vectors, the resulting clones were transformed into *E. coli* DH5 α cells (Invitrogen, Carlsbad, CA, USA) and colonies were screened by spreading the cells on LB-agar plates (**Table 3.3**) containing antibiotic (25 μ g/mL). Typically 10 viable colonies were further screened for insert incorporation using colony-PCR, a modified PCR experiment where the template DNA is substituted with a bacterial colony. Positive clones were amplified by miniPrep (QIAGEN, Hilden, Germany) and checked by DNA sequencing.

Protein expression and solubility trials were performed with all the clones (Section 3.1.2.1). In all cases but the C2 domain, the constructs cloned into the pETM-11 vector gave the highest soluble protein production yields. None of the C2 domain clones (16-154) yielded any soluble protein, and new boundaries were assayed (1-154). The construct cloned into the pET28-MHL vector (analogous to pETM-11: kanamycin resistance, His₆-tag, TEV protease site) was purchased (Addgene plasmid 25615) and yielded soluble protein.

All constructs after TEV protease cleavage of the His₆ tag are summarized in **Table 3.1**. The reported molecular weight values are isotopic average masses. Together with the extinction factors (ϵ , 280 nm) they have been calculated using the software ProtParam (web.expasy.org/protparam).

3.1.1.2. Site-directed mutagenesis

The catalytic cysteine residue in the HECT domain (C922) was mutated into aspartic acid to increase protein stability. Residues P489 and T500 in the WW3 domain were mutated into cysteine to introduce the MTSL/MTS tags for the acquisition of PRE data (CHAPTER 7).

All mutations were introduced using the QuickChange™ site directed mutagenesis kit from Stratagene®. pETM-11 clones of the corresponding genes obtained by miniPrep (Qiagen®) were used as templates. Extensively overlapping (12-15 bases) forward and reverse primers with a sequence mismatch introducing the mutation

were designed (**Figure 3.2**). The whole vector was amplified by PCR using the proofreading Pfu polymerase (Agilent Technologies, CA, USA). After that, the template DNA was digested using DpnI, which targets only methylated sequences (5'-GmeATC-3') that occur only in biosynthesized DNA but not in the *in vitro* generated clones. The mutant clones were purified using the MinElute PCR Purification Kit (QIAGEN, Hilden, Germany), transformed into *E. coli* DH5 α cells and screened by DNA sequencing.

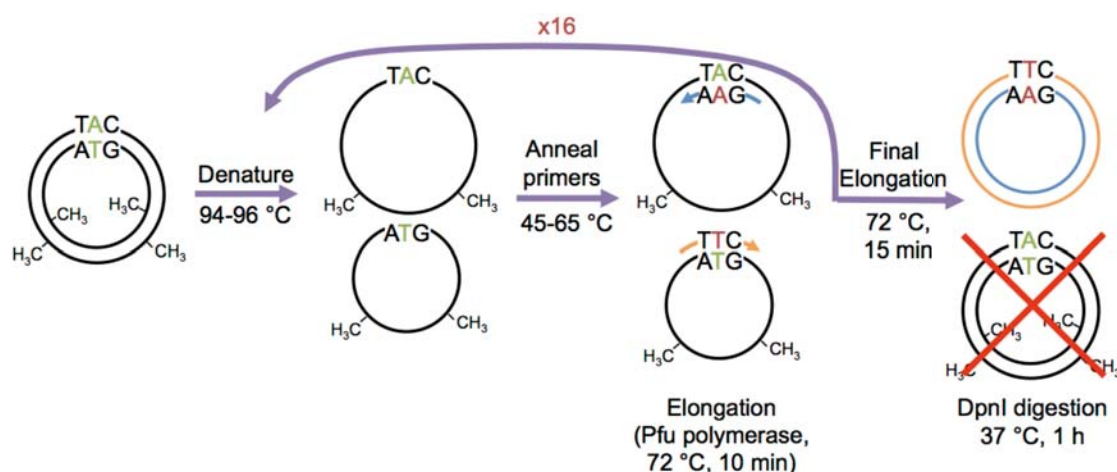


Figure 3.2: Protocol used for site-directed mutagenesis. The whole vector is amplified using primers with a sequence mismatch. The original vector is digested with DpnI, which targets only methylated sequences that are not generated synthetically.

3.1.1.3. Transformation

Similar protocols were used to transform both DH5 α (cloning and site-directed mutagenesis) and BL21 (protein expression, from Novagen, Darmstadt, Germany) *E. coli* strains. Small aliquots of competent cells were slowly thawed in ice from -80 °C. 1 to 5 μ L of the clone were incubated with 80 μ L of competent cells ($OD_{600} = 0.6$) for 20 min in ice. A 45 sec heat shock was applied at 42 °C without agitation, and the cells were transferred back to the ice bucket for 2 additional minutes. 300 μ L of antibiotic-free SOC (Super Optimal broth with Catabolite repression) medium (**Table 3.3**) were added and the cells were grown at 37 °C for 1 hour with 200 rpm agitation. Next, the culture was spread on an LB-agar plate containing the corresponding antibiotic (kanamycin or ampicillin) for the selection of positive cell colonies. Cells were grown on the plate overnight at 37 °C.

3.1.2. Protein production

Figure 3.3 shows the general pipeline that has been used in this thesis to obtain protein samples. Details on each step are provided in this section.

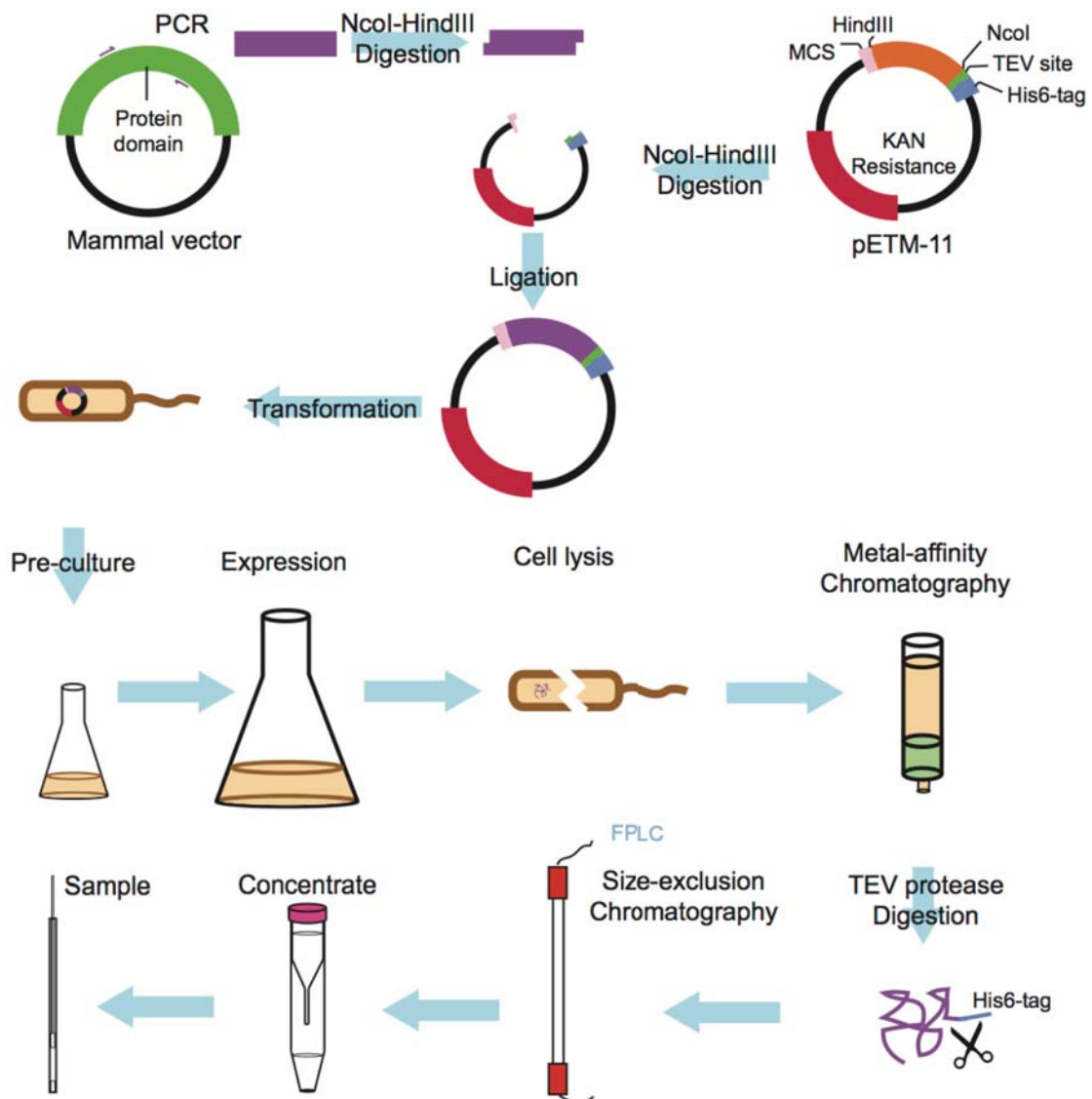


Figure 3.3: Summary of molecular biology methods for protein production.

3.1.2.1. Protein expression and solubility trials

The expression of the different constructs was assessed using small-scale trials. 5 mL pre-cultures of the transformed *E. coli* BL21 cells were grown in antibiotic-supplemented (25 µg/mL) LB medium overnight at 37 °C in 15 mL Falcon™ tubes (BD Biosciences). Three different 24-well plates with a cover allowing air exchange were used to grow 2 mL LB cultures at 20, 30 and 37 °C. 100 µL of the pre-culture were added to each well and cells were grown at 37 °C while monitoring the OD₆₀₀ until it reached 0.6-0.7 Au. After that, each plate was equilibrated at its final temperature for 30 min and IPTG was added at 100, 250 and 500 µM concentrations to induce the expression. 10 µL aliquots from each well were taken after 4 h and overnight expression and mixed with 10 µL 2X SDS-PAGE (sodium dodecyl sulphate

polyacrylamide gel electrophoresis) loading buffer. The expression was evaluated using this technique (**Figure 3.4**).

Next, the solubility of the proteins using optimal expression conditions was evaluated. To reproduce the final production protocol, 1 L LB cultures were prepared as detailed in Section 3.1.2.3. Cells were lysed (Section 3.1.2.4) and small aliquots of the pelleted cell debris and the soluble fraction were evaluated by SDS-PAGE. Proteins expressed in inclusion bodies were recovered in a buffer containing 6M guanidine hydrochloride and, when possible, refolded into buffers with no denaturing agents using different strategies.

The expression and solubilisation conditions that yielded the highest amounts of each domain are summarized in **Table 3.2**. In all cases but the HECT domain optimal production yields were obtained when expressing the proteins in inclusion bodies (37 °C, 0.25-0.5 mM IPTG) and solubilizing them after cell lysis by re-folding. None of the strategies assayed for the re-folding of the HECT domain was successful, and its expression in the soluble fraction required very mild conditions (15 °C, 50 µM IPTG).

Domain	VECTOR			EXPRESSION			SOLUBILITY	
	Vector	Tag	Res. ^a	T (°C)	[IPTG] (mM)	Time ^b	IB/S ^c	Re-folding
C2	pET28-MHL	His ₆	KAN	37	0.5	o/n	IB	✓
WW1	pETM-11	His ₆	KAN	37	0.5	o/n	IB	✓
WW2	pETM-11	His ₆	KAN	37	0.5	o/n	IB	✓
WW3	pETM-11	His ₆	KAN	37	0.5	o/n	IB	✓
WW3								
P489C	pETM-11	His ₆	KAN	37	0.5	o/n	IB	✓
T500C								
WW4	pETM-11	His ₆	KAN	37	0.5	o/n	IB	✓
HECT	pETM-11	His ₆	KAN	15	0.05	o/n	S	✗
HECT C-lobe	pETM-11	His ₆	KAN	37	0.5	o/n	IB	✓

Table 3.2: Optimized protein production conditions obtained from the protein expression and solubility trials. (a) Res. refers to the antibiotic resistance conferred by the vector. (b) o/n is the abbreviation for overnight. (c) IB denotes that the protein is expressed in inclusion bodies. S indicates that it is expressed in solution.

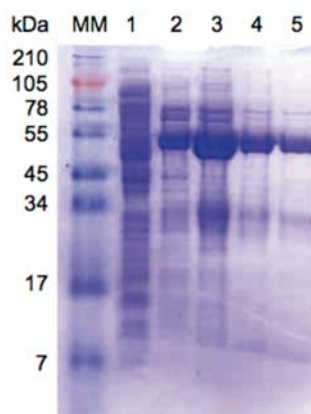


Figure 3.4: SDS-PAGE analysis of protein samples. Sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse protein samples generated during protein expression and purification. “MM” refers to molecular markers.

3.1.2.2. Glycerol stocks

Glycerol stocks of transformed *E. coli* BL21 cell cultures were prepared by laying away 1 mL of the pre-culture and adding sterilized glycerol to a concentration of 15% (v/v). The cultures were cooled down in ice and frozen directly at -80 °C. When required for culture inoculation, the stocks were kept frozen in ice and only a small fraction was taken using a pipette tip.

3.1.2.3. Protein expression

Pre-cultures were prepared by inoculating 10 to 25 mL LB broth (supplemented with 25 mg/mL of antibiotic for plasmid selection) with either freshly transformed *E. coli* BL21 cell colonies in agar plates or from glycerol stocks. Cells were grown overnight at 37 °C in 250 mL crystal Erlenmeyer flasks with 200 rpm agitation.

10 mL of the pre-culture were used to inoculate 1L of LB (25mg/ml of antibiotic) and cells were grown at 37 °C with 200 rpm agitation in 2L Erlenmeyer flasks. The OD_{600} was monitored until it reached 0.6-0.7 Au, typically 3-4 hours after inoculation. After that, cultures set for the production of unlabelled samples were equilibrated at the final expression temperature for 30 min (**Table 3.2**). The expression was then induced by adding IPTG at the optimized concentration and left overnight.

Isotopically enriched proteins for NMR experiments were expressed using the Marley protocol (Marley et al., 2001). In this case the cells were grown in LB (25 mg/mL antibiotic, 37 °C, 200 rpm, OD_{600} = 0.6-0.7 Au) and centrifuged for 20 min at 3000 g and 4 °C. The pellet was washed with PBS (Phosphate buffered saline) and resuspended in 250 mL of M9 minimal media. The appropriate combination of D₂O (99.99%, CortecNet), ¹⁵NH₄Cl (Cambridge Isotopes, MA, USA) and/or D-[¹³C] glucose

(Cambridge Isotopes, MA, USA) were used as sole hydrogen, nitrogen and carbon sources respectively to prepare the labelled samples. The culture was equilibrated at the final expression temperature for 1 h and IPTG was added at the optimized concentration (**Table 3.2**). The expression in isotopically enriched media proceeded overnight with the exception of D₂O cultures that were incubated for 48h.

3.1.2.4. Cell lysis

After the expression, cell cultures were centrifuged (3000 g, 4 °C, 20 min) and resuspended in lysis buffer.

The HECT domain was resuspended in its optimized lysis buffer (specified in the PDB entry 2ONI, Structural Genomics Consortium (SGC)) with added protein fold stabilizers (**Table 3.3**) including 5% glycerol, 1 mM β-mercaptoethanol and 100 μM zinc chloride as well as 0.1 mM PMSF (phenylmethyl sulfonyl fluoride), a protease inhibitor that was essential to avoid the degradation of the protein by *E. coli* proteases during the lysis and Co²⁺ affinity purification steps (Section 3.1.2.5). 30 mL of the buffer were used to resuspend the pellet from 1 L of culture. DNaseI (10 μg/mL) and lysozyme (350 μg/mL) were added to the sample and it was incubated at 4 °C for 20 min with orbital agitation. To avoid the precipitation of the protein, cell lysis was achieved using mild sonication conditions (Sonics® VibraCell™ VCX 750, 20% amplitude, 2 x [15 sec on + 45 sec off, 15 min] cycles). The cultures were cooled down with ice along the lysis procedures.

Cultures expressing the protein in inclusion bodies (**Table 3.2**) were resuspended in a denaturing buffer containing 6 M guanidine hydrochloride at pH 8 (30 mL per L of culture). Cells were lysed using either an EmulsiFlex-C3 (Avestin, equipped with a Peltier temperature controller system designed in-house) cell disrupter with a maximum pressure of 20000 psi or by sonication (Sonics® VibraCell™ VCX 750, maximum amplitude 40%, 2 x [30 sec on + 30 sec off, 6 min] cycles), or a combination of both.

3.1.2.5. Protein purification

After the lysis, the cell debris were removed by ultra-centrifugation at 30000 g, 4 °C for 20 min. The HECT domain was bound to a Co²⁺-affinity resin (Talon® Superflow™, GE Healthcare, Uppsala, Sweden), washed alternatively with optimized wash buffers A and B and eluted with an elution buffer containing 200 mM imidazole (**Table 3.3** and **Figure 3.5A**).

All other domains were bound to a Ni²⁺-affinity resin (ABT Beads, Madrid, Spain) while denatured in the 6 M guanidine hydrochloride buffer pH 8. After 3 washes with this buffer, the WW domains and the HECT C-lobe were re-folded while bound to the resin by washing it with buffer A (x3) (**Figure 3.5B**). They were eluted in their native fold using buffer A with added 50 mM EDTA (ethylenediaminetetraacetic acid).

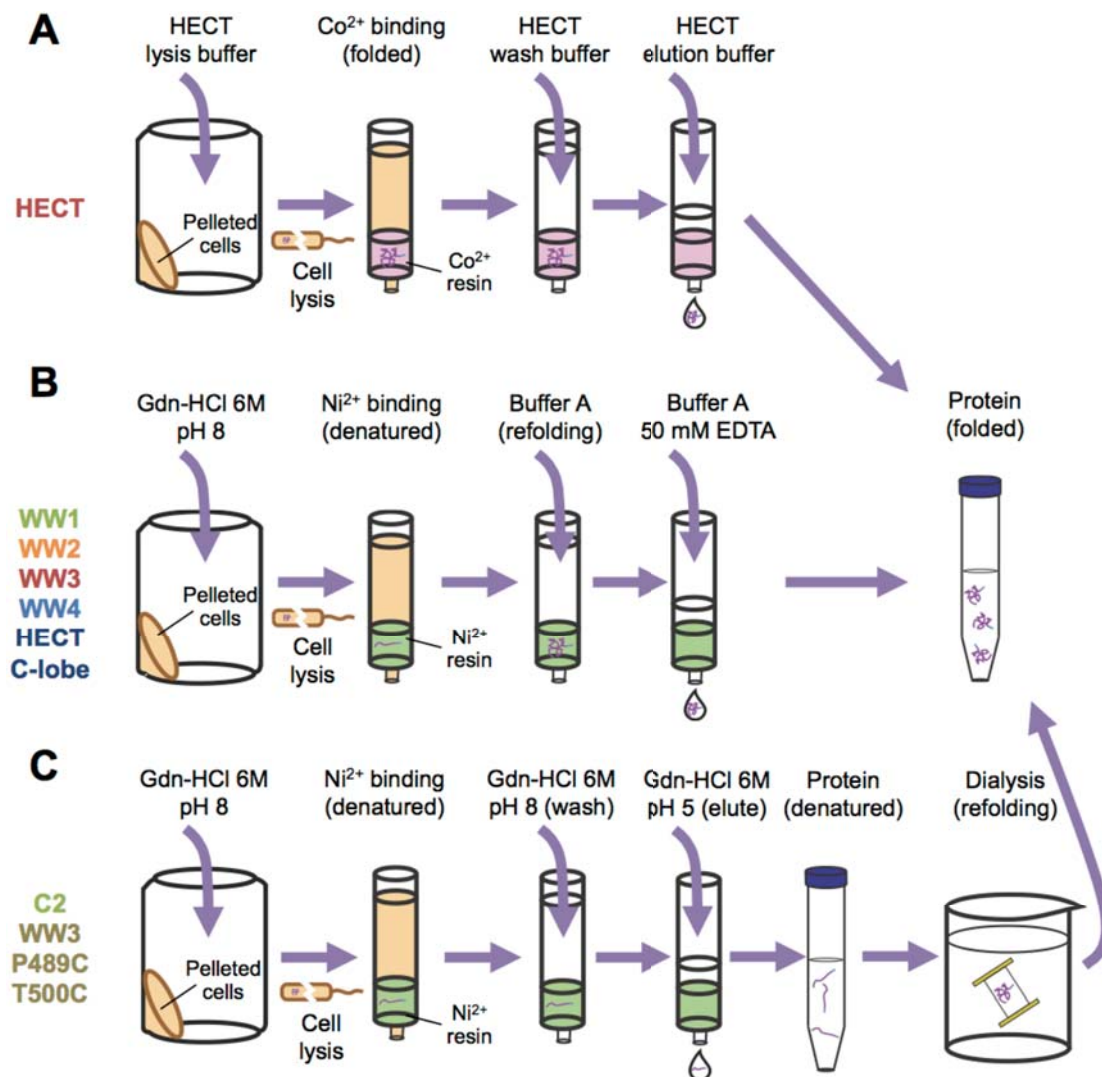


Figure 3.5: Strategies used to obtain the protein constructs in their native fold. (A) The HECT domain was expressed in its native fold and purified in a Co²⁺ resin in denaturant-free conditions. (B) All the WW domains and the HECT C-lobe were expressed in inclusion bodies, recovered with a buffer containing 6 M guanidine hydrochloride pH 8 and refolded while bound to a Ni²⁺ resin. (C) The C2 domain and the WW3 mutants were expressed in inclusion bodies, purified using a Ni²⁺ resin, eluted in a buffer containing 6 M guanidine hydrochloride pH 5 and refolded by dialysis against a denaturant-free buffer.

The C2 domain and the WW3 mutants were eluted from the resin in their denatured states using a 6 M guanidine hydrochloride buffer pH 5 and refolded by dialysis (SpectraPor® Dialysis Membrane MWCO: 1000) (**Figure 3.5C**). In both cases the eluted solution (about 30 mL) was dialyzed straight away against buffers with no

denaturing agents (two steps dialysis at 4 °C, 1 L per step, 1 step overnight plus 1 step 5 h, 250 rpm agitation with a magnetic stirrer). The C2 refolding buffer contained 1 mM EGTA (ethylene-bis(oxyethylenitrilo)tetraacetic acid tetrasodium) to eliminate the residual Ca²⁺ in solution. The domain required high ionic strength to be dissolved (≥250 mM NaCl). Protein precipitate after the dialysis was pelleted by centrifugation (30000 g, 4 °C, 10 min) and the soluble fraction was recovered.

With the domains in their native fold, the His₆ tag was cleaved by TEV (Tobacco Etch Virus) protease digestion overnight at room temperature (4 °C for the HECT domain) with orbital agitation. The proteins were separated from the tag, the protease and remaining impurities by size-exclusion chromatography on an Äkta™ Purifier10/FPLC system (GE Healthcare, Uppsala, Sweden). HiLoad™ Superdex 30, 75 or 200 16/60 pregrade columns were selected depending on the size of the domain. An isocratic elution was used along 1 column volume (120 mL) with the buffer to be used in subsequent experiments, which varied from one protein domain to another (Section 3.1.3). Protein elution was monitored by UV absorption at 220 and 280 nm wavelengths and collected in 2 mL fractions.

Both tag cleavage and size-exclusion chromatography steps were analysed by SDS-PAGE. Fractions containing the purified protein were mixed and concentrated to 350 – 1000 µL final volume using Ultracel® 3 kDa ultrafiltration discs in an Amicon 8050 stirred cell (Millipore, Billerica, MA, USA). Final protein concentration ranged from 100 µM to 1 mM depending on the experiment to be performed afterwards.

3.1.2.6. Determination of protein concentrations

The protein concentration in the purified samples was assessed by UV absorption using the Lambert-Beer law (Equation 3.1).

$$C = \frac{A_{280}}{\epsilon \times l} \quad (3.1)$$

C is the protein concentration (mM), A₂₈₀ is the recorded absorbance at 280 nm (Au), ε is the extinction coefficient (mM⁻¹ cm⁻¹) as calculated with the program ProtParam (**Table 3.1**) and l is the path length (cm).

Protein fluorescence arises from the aromatic rings present in the side chains of some amino acids. Tyr and Trp have an absorbance maximum at 275 and 280 nm respectively and are the major contributors to the value of ε. Because these are commonly located in the hydrophobic core of proteins, the measurement of A₂₈₀

required their exposure by diluting a small aliquot (2-5 μL) of the sample in 6 M guanidine hydrochloride pH 5 (dilution factors ranged from 14 to 40 fold). Measurements were performed using a UV-min 1240 spectrophotometer (Shimadzu, Kyoto, Japan).

3.1.3. Media and buffers

Reagents were purchased to Sigma-Aldrich (St. Louis, MO, USA) or Melford (Ipswich, UK). Bacterial growth media were purchased to Conda (Madrid, Spain).

BACTERIAL GROWTH MEDIA			
Luria Broth (LB) (1 L) (autoclaved)		LB-Agar plates (autoclaved)	
NaCl	10 g	LB Broth	25 g
Tryptone	10 g	Bacteriological Agar	15 g
Yeast extract	5 g	Antibiotic	25 $\mu\text{g}/\text{mL}$
Antibiotic	25 $\mu\text{g}/\text{mL}$		
M9 salts 10x stock (autoclaved)		M9 pellet wash 10x (autoclaved)	
Na_2HPO_4	60 g	Na_2HPO_4	60 g
KH_2PO_4	30 g	KH_2PO_4	30 g
NaCl	5 g	NaCl	5 g
$^{14}\text{NH}_4\text{Cl}/^{15}\text{NH}_4\text{Cl}$	5 g		
Trace Elements 100x (1 L) pH 7.5		M9 medium 1x (1 L) (H_2O or D_2O)	
EDTA	5 g	M9 salts 10x	100 mL
$\text{FeCl}_3 \times 6\text{H}_2\text{O}$	833 mg	Trace Elements 100x	10 mL
ZnCl_2	84 mg	Glucose 20%/ $^{13}\text{C}_6$ -glucose	20 mL/2 g
$\text{CuCl}_2 \times 2 \text{H}_2\text{O}$	13 mg	MgSO_4	1 mL
$\text{CoCl}_2 \times 6 \text{H}_2\text{O}$	10 mg	CaCl_2 [1M]	0.3 mL
H_3BO_3	10 mg	Biotin [1 mg/mL]	1 mL
$\text{MnCl}_2 \times 6 \text{H}_2\text{O}$	1.6 g	Thiamin [1 mg/mL]	1 mL
		Antibiotic	25 $\mu\text{g}/\text{mL}$
SOC medium (0.5 L) (autoclaved)			
Tryptone	10 g	Yeast extract	2.5 g
NaCl [5M]	1 mL	KCl [1M]	1.25 mL
MgCl_2 [1M]	5 mL	MgSO_4 [1M]	5 mL
Glucose [1M]	10 mL		

LYSIS AND METAL AFFINITY PURIFICATION BUFFERS			
Ni²⁺ Buffer A pH 8		Ni²⁺ elution buffer pH 8	
NaCl	150 mM	Ni ²⁺ Buffer A	
Tris-HCl	20 mM	EDTA	50 mM
Imidazole	10 mM		
Gdn-HCl 6M pH 8 (binding) or pH 5 (elution)			
Guanidine hydrochloride	6M	NaH ₂ PO ₄ /Na ₂ HPO ₄	100 mM
HECT lysis pH 8		HECT wash A pH 8	
Tris-HCl	10 mM	Tris-HCl	10 mM
NaCl	0.5 M	NaCl	0.15 M
Glycerol	5%	Glycerol	5%
Imidazole	2 mM	Imidazole	10 mM
β-mercaptoethanol	1 mM	β-mercaptoethanol	1 mM
ZnCl ₂	0.1 mM		
HECT wash B pH 8		HECT elution pH 8	
HECT wash A		HECT wash A	
Tween20	0.05%	Imidazole	200 mM
GEL FILTRATION AND NMR BUFFERS			
Sodium phosphate buffer pH 7 (WW domains and HECT C-lobe)		Tris-HCl buffer pH 7 (C2 and HECT domains)	
NaH ₂ PO ₄ /Na ₂ HPO ₄	20 mM	Tris-HCl	20 mM
NaCl	100 mM	NaCl	250/400 mM
NaN ₃	0.5 mM	EGTA	1 mM
		CaCl ₂ (when appropriate)	20 mM
		NaN ₃	0.5 mM

Table 3.3: Bacterial growth media and buffers used in molecular biology experiments.

3.2. Solid-phase peptide synthesis

3.2.1. Manual and automated Fmoc-SPPS

Either Aminomethyl ChemMatrix® (PCAS BioMatrix Inc., Quebec, Canada) or Rink-Amide (Novabiochem, Merck-Millipore, Billerica, MA, USA) resin were used as solid supports. The synthesis scale varied from 0.2 to 0.25 mmol. Resin washes were

performed using DMF (dimethylformamide) (3-5 x 5 mL, 1 minute agitation each). Resin swelling was achieved with DCM (dichloromethane) (3-5 x 5 mL, 1 minute agitation each). Solutions were drained using vacuum.

Manual syntheses were performed in polypropylene syringes with fitted porous filters. The resin was gently mixed either manually or with orbital agitation at every step. Automated room temperature syntheses were performed using an Applied Biosystems 433A Peptide Synthesizer. Automated microwave enhanced syntheses were carried out with a CEM Liberty1 synthesizer.

3.2.1.1. Resin equilibration

Prior to the start of the synthesis, the resin was equilibrated in mild acidic conditions to drain out residual functional groups not covalently bound to the resin. After initial DMF washing and DCM swelling, 1% TFA (trifluoroacetic acid) in DCM was added to the resin (5 x 5 mL, 1 min each). Next, mild basic conditions were used to neutralize the resin (DIEA (N,N-Diisopropylethylamine) 5% in DCM, 5 x 5 mL, 1 min each).

3.2.1.2. Coupling of the Rink-Amide linker

When using the Aminomethyl ChemMatrix® resin, a Rink-Amide linker was manually attached to the resin to yield peptides with C-terminal amidation. The linker was pre-activated by mixing it (3 eq) with HATU (3 eq), HOAt (3 eq) and DIEA (6 eq) in 1.5 mL DMF. After extensive washing of the resin with DMF and swelling with DCM, the activated linker was added to the resin and the coupling proceeded for 90 min.

The coupling was checked using the Kaiser test for the detection of free primary amines (Section 3.2.1.4). At this point, the Rink-Amide linker in both the Aminomethyl ChemMatrix® and the Rink-Amide resin was deprotected by removing the Fmoc moiety (Section 3.2.1.3).

3.2.1.3. Fmoc deprotection

The deprotection of the initial Rink-Amide linker and the backbone amide groups was achieved with a piperidine treatment. After washing and swelling the resin, a 20% solution of piperidine in DMF was added (2 x 5 mL, 10 min each). The success of the deprotection step was evaluated using either a colorimetric test or by detecting the Fmoc group in the drained solution (Section 3.2.1.4).

Alternatively, a 5% DBU (1,8-Diazabicycloundec-7-ene) solution in DMF was used instead of the piperidine mixture to deprotect phosphorylated residues avoiding β -elimination of the phosphate groups, especially when using microwave-enhanced automated synthesis (Attard et al., 2009).

3.2.1.4. Tests for the detection of free amines

The Kaiser test (Kaiser et al., 1970) was performed to detect primary amines. DMF was extensively drained with DCM (8 x 5 mL, 1 min each) as its interference with the test reagents yields false positive results. A few resin beads were laid away in a small crystal test tube. 3 drops of Kaiser A and 1 drop of Kaiser B solutions (Section 3.2.3) were added and the mixture was heated for 3 min at 100 °C. The solution turned blue in the presence of free primary amines.

The chloranil test (Christensen, 1979) was carried out to detect secondary amines, such as those present in the backbone of proline residues. Similarly, a few beads were transferred to a test tube. 200 μ L of acetone and 50 μ L of a saturated solution of chloranil in toluene were added, and the mixture was shaken for 5 min at room temperature. The beads turned green in the presence of free secondary amines.

The room temperature automatic synthesizer monitored backbone amine deprotection by measuring the conductivity of the drained solution to detect the released Fmoc moiety.

3.2.1.5. Iterative coupling of Fmoc-amino acid derivatives

The Fmoc-amino acid derivatives used for manual couplings were pre-activated in the same way as the Rink-Amide linker (Section 3.2.1.2). Alternative mixtures of activator agents were eventually used, including PyBOP or oxima pure (3 eq) and DIC (3 eq) with DIEA (6 eq) in 1.5mL DMF. The mixture was added to the resin and the coupling proceeded for 90 min.

In room temperature automated synthesis 10 equivalents of the Fmoc amino acid derivatives were used per coupling. These were activated with a mixture of HATU and HOAt (0.45 M each) in DMF in the presence of DIEA. The coupling reaction proceeded for 30 min.

In microwave assisted automated synthesis 0.2 M solutions of the Fmoc amino acid derivatives were prepared and 5 mL of the solution were used in each coupling (4-5 eq). 2 mL of a mixture of HATU and HOAt (0.5 M each, 4-5 eq) together with 1 mL of 2 M DIEA (8-10 eq) were used as coupling activators. The coupling proceeded

for 5 minutes with 25 W and a maximum temperature of 75 °C for non-modified amino acids. Phosphorylated amino acids were coupled for 15 minutes with 25 W at a maximum temperature of 72 °C (Harris et al., 2008).

After the coupling, the resin was extensively washed and swollen with DMF and DCM. The reaction was evaluated using a colorimetric test (Section 3.2.1.4). After a successful coupling, the backbone amide group was deprotected (Section 3.2.1.3). After successful deprotection, the resin was gently washed and swollen and the next Fmoc-amino acid derivative was coupled. The process was iteratively repeated up to the completion of the amino acid chain.

3.2.1.6. Backbone amide acetylation

After the addition of the last amino acid of the sequence, the N-terminal amide group was acetylated. A 1 mL DMF solution containing 50 µL of acetic anhydride and 85 µL of DIEA was added to the resin and the reaction proceeded for 30 min with orbital agitation at room temperature.

3.2.1.7. Resin cleavage, side-chain deprotection and mini-cleavage

Resin cleavage and side-chain deprotection was performed using a 95% TFA solution containing 2.5% H₂O and 2.5% TIS (triisopropylsilyl) as scavengers of the highly reactive species generated. Alternative mixtures were eventually used containing scavenging thiol compounds including EDT (2.5%), DODT (2.5%) or 2-PySH (5 eq).

10 mL of the TFA solution were added to the resin and the reaction proceeded for 1 h. The solution was filtered through the polystyrene porous disc in the synthesis syringe and collected in a crystal round-bottom flask. The process was repeated once.

To avoid oxidation, this step was eventually performed in oxygen-free conditions. The resin was transferred to a round-bottom flask capped with a septum. An argon or nitrogen atmosphere was maintained along the reaction.

After cleavage, TFA was evaporated from the solution using a nitrogen gas flow. The remaining solution was slowly added to 40-50 mL of cold tertmethylbutylether (-20 °C) in a Falcon™ tube resulting in peptide precipitation. The suspension was centrifuged (10000 g, 4 °C, 10 min) and the ether was discarded. The peptide crude was washed three times by resuspending it in cold ether followed by centrifugation. Next, the crude was dried using a nitrogen gas flow, dissolved in deionized H₂O, frozen in liquid nitrogen and lyophilized.

A similar procedure was used for mini-cleavage steps at a smaller scale. Typically 0.5 mL of the TFA solution were added to a few resin beads and 2 mL of the cold ether were used for precipitation. In this case, the crude was directly analysed using LC-MS or MALDI-TOF mass spectrometry (Section 3.2.1.9).

3.2.1.8. RP-HPLC purification of peptides

Lyophilized peptide crudes were dissolved in H₂O (eventually with 5% acetonitrile to increase solubility) and purified by semi-preparative RP-HPLC (reversed-phase high performance liquid chromatography) using a Äkta™ Purifier10 (GE Healthcare, Uppsala, Sweden) controlled with the software Unicorn. Peptide elution was detected by UV at wavelengths 220 nm (peptide bond), 256 nm (Phe aromatic ring) and 280 nm (Trp and Tyr aromatic rings).

A SunFire™ (Waters, Milford, MA, USA) PREP C18 OBD™ column (5 µm particle size, 100 Å pore size, 19x100 mm) was used as the stationary phase. An acetonitrile gradient in H₂O with 0.1% TFA at a flow rate of 10 mL/min was used, typically from 0 to 20% in 5 minutes and from 20 to 30% in 20 minutes. Eventually TFA was substituted with FA (formic acid) enabling the purification of the target phosphopeptide from truncated versions lacking one or more phosphorylated residues.

The eluted solution was collected in 1 mL fractions and these were analysed by LC-MS and/or MALDI-TOF mass spectrometry (Section 3.2.1.9). Fractions containing the pure peptide were mixed together, frozen with liquid nitrogen and lyophilized.

3.2.1.9. LC-MS and MALDI-TOF mass spectrometry

Peptide crudes, purified fractions and pure samples were analysed using coupled liquid chromatography mass spectrometry (LC-MS) with a Waters Alliance coupled to a Waters Micromass ZQ. For LC, a linear acetonitrile gradient (0-100%) in H₂O (0.1% FA) was used with a flow rate of 10 mL/min. 14 min runs were performed using a SunFire™ C18 column (particle size 3.5 µm, pore size 100 Å, 4.6x100 mm) and 6.5 min runs were performed using a XSelect C18 column (particle size 3.5 µm, pore size 100 Å, 4.6x50 mm), both from Waters, Milford, MA, USA. Peptide UV detection was performed at 220 nm. The coupled mass spectrometer used electrospray ionization (ESI) coupled to a quadrupole-time of flight analyser (Q-TOF). Ion detection was achieved using the analyser at the positive mode. UV and TIC chromatograms as well as mass spectra were analysed using the software MassLynx.

Alternatively, MALDI-TOF (Matrix Assisted Laser Desorption Ionisation-Time Of Flight) mass spectrometry was used to analyse both peptide and protein samples. Spectra acquisition was performed with an ABSciex 4800 *Plus* MALDI TOF/TOF or an Applied Biosystems 4700 Proteomics Analyzer. Samples (1 μ L) were co-crystallized with a matrix (1 μ L, Section 3.2.3) by drying the droplets after spotting them on a MALDI plate (Applied Biosystems, Foster City, CA, USA). About 1000 shots were accumulated to obtain a good signal-to-noise ratio. Optimal results for phosphopeptide detection were typically obtained when using DHB (dihydroxybenzoic acid) as the ionization matrix and the analyser in reflector negative mode. Non-phosphorylated peptides were best detected using ACH (α -cyano-4-hydroxycinnamic acid) and the analyser in reflector positive mode. Nedd4L WW3 mutants used in PRE analyses were best detected using DHB and the analyser in linear positive mode. Bigger protein domains were optimally detected using SA (sinapinic acid) and the analyser in linear positive mode.

3.2.1.10. Peptide concentration determination

Peptide samples to be used in NMR and ITC experiments were prepared by dissolving a purified lyophilized aliquot in H₂O and neutralizing the pH of the solution with 1 M sodium hydroxide.

Determination of the peptide concentration in solution was achieved by quantitative amino acid analysis performed by the *Unitat de Tècniques Separatives, Serveis Científic-Tècnics, University of Barcelona*. This method is based on the hydrolysis of the peptide (typically a 5-10 μ L aliquot for expected concentrations between 5 and 10 mM) using a 6 M chlorhydric acid solution (110 °C, 24 h) followed by the chromatographic separation of the amino acids by HPLC and the integration of the peak areas compared to an internal standard.

3.2.2. Native chemical ligation of synthetic peptides

3.2.2.1. Synthesis of the peptide thioester GYPGA-COS-Bzl

The peptide thioester was synthesized using the 4-Sulfamylbutyryl AM resin (0.79 mmol/g load, Novabiochem, Merck-Millipore, Billerica, MA, USA) using a 0.15 mmol scale. Coupling of the C-terminal Ala residue was performed essentially as described (Backes and Ellman, 1999; Huse et al., 2000). Chloroform was filtered in basic alumina and used to dissolve the Fmoc-Ala derivative (3 eq) and DIEA (5 eq). The solution was added to the resin and the mixture was equilibrated for 10 min at room temperature and for 20 min at -20 °C (achieved by immersing the reaction flask in an

isopropanol bath refrigerated with a cryocooler). The coupling activator PyBOP was then added (5 eq) and the coupling proceeded for 8h at -20 °C. The resin was gently washed with chloroform and swollen with DCM. The subsequent couplings were performed as described in Section 3.2.1.5.

After synthesis completion, the peptide was acetylated as described in Section 3.2.1.6 and the resin was activated for cleavage by overnight treatment with a 2 mL mixture of iodoacetonitrile (32 eq) and DIEA (5 eq) in DMF. After extensive washing and swelling with DMF and DCM, the peptide was cleaved by treating the resin with 1 M benzyl mercaptan and 0.25 M DIEA in DCM (43 mL, 6 h). DCM was then evaporated with a nitrogen gas flow, the Tyr side-chain protecting group was removed and the peptide was precipitated and lyophilized as reported in Section 3.2.1.7.

3.2.2.2. Native chemical ligation

Both the peptide thioester GYPGA-COS-Bzl and the phosphopeptide Smad3 pS204-pS208 were dissolved in NCL buffer (Section 3.2.3, (Johnson and Kent, 2006)) to an approximate concentration of 1 mM. The progress of the reaction was monitored by LC-MS (Section 3.2.1.9). 10 µL aliquots of the reaction were diluted in 190 µL of 50% acetonitrile in deionized H₂O and analysed. After the reaction, the ligation product was purified by RP-HPLC, precipitated and lyophilized (Section 3.2.1.8).

3.2.2.3. Cysteine desulfurization

Desulfurization of the cysteine residue in Smad3 pS208 was performed essentially as described (Yan and Dawson, 2001). The peptide was dissolved in 20% acetic acid (1.6 mg/mL) and 10% palladium on activated carbon was added to the solution. The reaction was performed overnight in a H₂ atmosphere. The solution was filtered and an aliquot was analysed using LC-MS (Section 3.2.1.9).

3.2.3. Solutions and buffers used in SPPS

Reagents were purchased to Sigma-Aldrich (St. Louis, MO, USA) or Melford (Ipswich, UK). Fmoc amino acid derivatives were acquired from Novabiochem (Laufelfingen, Switzerland).

KAISER'S TEST SOLUTIONS			
Kaiser's test solution A		Kaiser's test solution B	
Phenol in ethanol	40 g / 10 mL	Ninhydrin	5 g
0.001 M KCN	2 mL	Ethanol absolute	100 mL
Pyridine freshly distilled from ninhydrin	98 mL		
MALDI-TOF MATRICES		NCL BUFFER	
DHB/ACH/SA		Ligation buffer pH 7	
DHB/ACH/SA	20 mg/mL	Guanidine hydrochloride	6 M
ACN in H ₂ O	50%	MPAA	50 mM
TFA (DHB/ACH)	0.1%	Phosphate	20 mM
Ortho-phosphoric acid (SA)	2.5 %	NaH ₂ PO ₄ /Na ₂ HPO ₄	20 mM

Table 3.4: Composition of solutions used in SPPS not detailed in the main text.

3.3. Isothermal Titration Calorimetry (ITC)

3.3.1. Principles and instrumentation

Isothermal Titration Calorimetry (ITC) is a biophysical technique that allows the simultaneous measurement of several thermodynamic parameters associated with any intermolecular interaction, including protein-ligand binding (Ladbury and Chowdhry, 1996). The association constant (K_A , μM^{-1}), the reaction stoichiometry (n) and the binding enthalpy (ΔH , kJ mol^{-1}) are directly measured with no need of labelling or chemical derivatization of neither the protein nor the ligand (Wiseman et al., 1989). Using Equations 3.2 and 3.3, the values of the dissociation constant (K_D , μM , often referred as the affinity constant), the Gibbs energy (ΔG , kJ mol^{-1}) and the entropy (ΔS , kJ K^{-1}) can be deduced.

$$K_A = \frac{1}{K_D} \quad (3.2)$$

$$\Delta G = \Delta H - T\Delta S \quad (3.3)$$

The instrument is composed of two equivalent cells (**Figure 3.6**) made of a highly efficient thermal conducting material, typically gold. Each cell is equipped with a heater. A protecting adiabatic jacket avoids thermal exchange with the outside. The temperature of both cells is measured simultaneously. One of the cells is filled with buffer and used as a reference. Its heater applies a constant power to maintain the

temperature stable along the experiment. The other cell is filled with the protein sample dissolved in the same buffer, and its heater initially applies the same power.

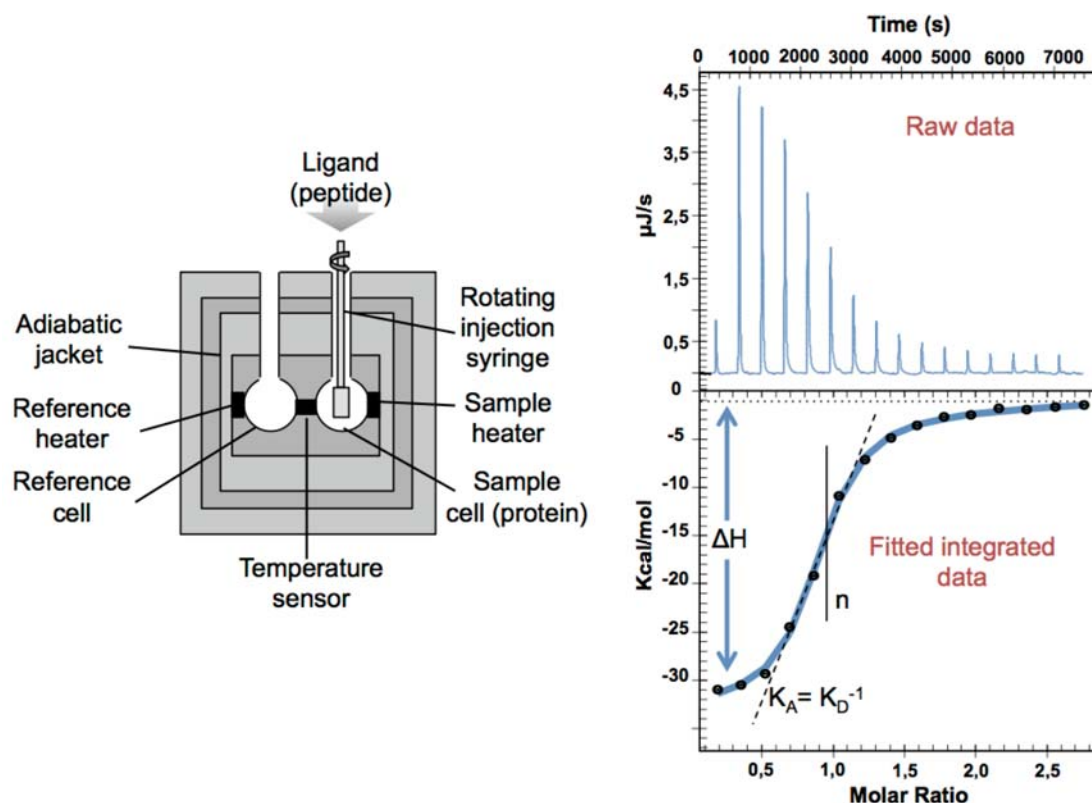


Figure 3.6: ITC instrumentation and data interpretation. On the left, schematic representation of a calorimeter used in ITC. On the right, the outcome data from an experiment is shown. Heat peaks after each ligand injection are integrated and a curve is fitted to the data. Several thermodynamic parameters can be deduced as well as the stoichiometry of the interaction.

The ligand is progressively added to the protein cell in small batch steps using a rotating syringe. At each step, only a fraction of the available protein binding sites is occupied by the ligand. If the reaction is exothermic, the temperature of the cell increases and its heater reduces the applied power to compensate that with respect to the reference cell temperature. If the reaction is endothermic, the opposite occurs. The instrument records the applied power over time ($\mu\text{J sec}^{-1}$). At each ligand injection the recorded power reaches a maximum (or minimum) value and afterwards its initial value is restored, defining a peak. With the consecutive ligand injections, the number of available protein binding sites is progressively decreased which is reflected by a progressive decrease of the released (or absorbed) thermal energy. Eventually the protein sample becomes saturated and peaks display a constant size reflecting the ligand heat of dilution.

The areas under each peak are integrated yielding the total exchanged heat per injection (kcal per mole of injectant), and data is plotted in function of the

ligand/protein molar ratio at each injection point. A mathematical model is fitted to the experimental data. For proteins with a single binding site, a sigmoidal curve is fitted (**Figure 3.6**). The slope at the linear fraction of the curve gives the K_A value. The molar ratio at which the interaction is 50% saturated gives the stoichiometry (n). The difference between the initial and final total exchanged heat gives the enthalpy ΔH .

3.3.2. ITC measurements

Measurements were performed using a nano ITC calorimeter (TA Instruments, New Castle, DE, USA) with an effective cell volume of 190 μL . With an estimate of the dissociation constant (K_D) and the stoichiometry parameter (n), the instrument controller software was used to simulate the experimental curve based on the calculation of the fitting parameter c (Equation 3.4), with optimal values between 1 and 1000.

$$c = K_A n [\text{Protein}] \quad (3.4)$$

Protein concentrations were calculated this way, and ligand concentrations were adjusted related to that. For the Ca^{2+} titration on the C2 domain, a 50 μM solution (300 μL) of the C2 domain was prepared in a buffer containing 20 mM Tris-HCl and 250 mM NaCl at pH 7 and loaded in the sample cell. The syringe was loaded with 50 μL of a 1.17 μM solution (22 fold) of CaCl_2 dissolved in the same buffer. Both solutions were degassed before the experiment, which was performed at 298 K. The CaCl_2 solution was titrated into the sample cell in 50 injections of 1 μL each, with 3 minutes spacing between them and with 200 rpm agitation. The experiment was repeated with no protein in the sample cell to measure the heat of dilution of CaCl_2 , and the measured values were subtracted from the titration. Data were analysed using the NanoAnalyze software (TA instruments) and the multiple binding sites fitting model provided the best fit.

To titrate the HECT-PY peptide on the WW domains, 150 μM solutions (300 μL) of the Nedd4L WW1, WW2, WW3 and WW4 domains as well as 50 μL solutions of the HECT-PY peptide at 10 to 16 fold concentrations (1.5 to 2.4 mM) were prepared, degassed and loaded into the sample cell and the syringe respectively. In both cases a buffer containing 20 mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ and 100 mM NaCl at pH 7 was used. The peptide was titrated in 15 steps of 3 μL each with an inter-injection delay time of 3 min and with 200 rpm agitation. All experiments were performed at 278 K. Blank experiments were performed as reported above. The one binding site model was fitted to the data using the NanoAnalyze software (TA instruments).

3.4. NMR Spectroscopy of Biomolecules

3.4.1. Theoretical and practical aspects

In this section, the basic theoretical and practical concepts behind the NMR experiments that are the basis of the results presented in this thesis are summarized. An extensive theoretical description of NMR can be found elsewhere, both for initiated (Keeler, 2010; Levitt, 2008; Lian and Roberts, 2011) and non-initiated (Doucleff et al., 2011). Practical aspects of NMR are extensively described in the literature (Bieri et al., 2011; Kwan et al., 2011; Lian and Roberts, 2011).

3.4.1.1. Fundamental concepts

Nuclear Magnetic Resonance Spectroscopy is a highly versatile technique widely used in a variety of fields for the study of the physicochemical properties of molecules and materials. In organic chemistry and biochemistry it is nowadays used in a widespread manner in structural studies of molecules in a wide range of sizes, from small ligands to proteins. Intra and inter-molecular interactions can be extensively characterized at atomic resolution using this technique. In addition, relaxation measurements permit the study of the dynamic properties of these molecules, which are often key in their biological function.

NMR exploits the magnetic properties of atomic nuclei arising from the spin. The spin is a quantum mechanical property of elementary particles including atomic nuclei. It gives rise to the component of angular momentum known as the nuclear magnetic moment μ , which does not arise from rotation and has no equivalent in classic mechanics. The component of the quantic angular momentum originated by rotation is known as the orbital angular momentum.

The nuclear magnetic moment μ of a nucleus is proportional to the Planck's constant h and is defined by its gyromagnetic ratio γ (rad s⁻¹ T⁻¹) and its spin quantum number I , which are quantum mechanical properties particular of each nucleus (Equation 3.5).

$$\mu = \gamma I \hbar \quad (3.5)$$

NMR active atomic nuclei have an associated spin quantum number I different from 0, which defines $2I + 1$ spin states with magnetic quantum number m ($-I, -I + 1, \dots, I - 1, I$). Therefore, nuclei with $I = \frac{1}{2}$ have only two spin states: $m = \frac{1}{2}$ (α) and $m = -\frac{1}{2}$ (β). These nuclei have a spherical nuclear charge distribution and yield the

simplest NMR spectra. The most common elements in biological molecules have isotopes with $I = \frac{1}{2}$, including ^1H , ^{13}C , ^{15}N , ^{19}F and ^{31}P . The natural abundance of both ^{13}C (1.108%) and ^{15}N (0.37%) is very low, and protein NMR requires isotopic enrichment of the samples. Moreover, their gyromagnetic ratio γ (6.7 and -2.7×10^7 $\text{rad s}^{-1} \text{T}^{-1}$ respectively) is much smaller than that of ^1H (26.7×10^7 $\text{rad s}^{-1} \text{T}^{-1}$). In contrast, ^1H is ubiquitous in biomolecules (99.99% natural abundance) and due to its bigger gyromagnetic ratio γ its sensitivity is higher. Thus, most of the protein NMR experiments are acquired using ^1H for detection.

In absence of an external magnetic field both spin states are degenerate, meaning that the energy of both of them is equivalent. When a magnetic field is applied, spins in the α state tend to align with the field (lower energy), whereas spins in the β state tend to align against the field (higher energy) in a phenomenon known as Zeeman splitting (**Figure 3.7**). The energy of each state depends on the intensity of the applied magnetic field B_0 as well as on the gyromagnetic ratio γ of the nucleus (Equation 3.6).

$$E_m = -m\gamma\hbar B_0 \quad (3.6)$$

The presence of the external magnetic field B_0 causes the stochastic precession of the spins around its application axis (z by convention). This precession occurs at the Larmor frequency ω_0 (rad s^{-1}), which is in the range of radiofrequencies (Equation 3.7).

$$\omega_0 = \gamma B_0 \quad (3.7)$$

The number of spins in the lowest energy level α is slightly higher than in the highest energy level β , and the relative population distribution is described by the Boltzmann equation (Equation 3.8), which is proportional to the energy level difference ΔE and the temperature. This population difference gives rise to a net magnetization vector M pointing towards the B_0 field direction. Because ΔE is small, the net magnetization vector M is small as well making NMR an intrinsically low sensitive technique.

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{kT}} \quad (3.8)$$

The application of a radiofrequency (RF) pulse B_1 on-resonance with the nuclei Larmor frequency forces them to precess in phase. If B_1 is applied in a direction different from B_0 , a fraction of the spin population is forced into the β state and M changes its direction: a 90° RF pulse in the y-axis will bring M to the x-axis,

generating transversal magnetization (name given to M when it is on the x-y plane). After the pulse, M tends to re-align with the field in a process known as relaxation while precessing around the z-axis. This originates a fluctuating electric current whose x-y component is detected giving rise to the Free Induction Decay (FID), a time-domain signal that is converted into the frequency-domain NMR spectrum using the Fourier Transformation (FT). The FT is used to decompose the periodic FID signal into a sum of sine and cosine functions. Several experiments (scans) are accumulated to counteract the low sensitivity of NMR by improving the signal-to-noise ratio, as the signal is proportional to the number of scans N while the noise is only proportional to $N^{1/2}$. In the spectrum, a peak is observed at the frequency corresponding to the Larmor frequency of each particular nucleus. This is the simplest possible NMR spectrum, known as 1D.

In response to B_0 , the electronic trajectories are reorganized creating small local magnetic fields in the opposite direction. Therefore, the actual field effective on each nucleus depends on the electronic density surrounding it. Because this is strongly dependent on the chemical environment, each nucleus in a molecule perceives a slightly different magnetic field and thus resonates at a slightly different frequency. The field-normalized difference of a particular nucleus' resonating frequency and a reference frequency is its chemical shift (ppm, Equation 3.9). Therefore, the NMR spectrum shows one peak at the particular resonating frequency of each nucleus in the molecule.

$$\delta = \frac{\omega_0 - \omega_{0,ref}}{\omega_{0,ref}} \quad (3.9)$$

Magnetization is transferred from one nucleus to another by two major mechanisms. Through-bond transfer occurs between nuclei connected through a limited number of covalent bonds and is known as the scalar coupling. The coupling between two particular nuclei is described by the field-independent coupling constant J (in Hz) that is much smaller than the chemical shift. Time evolution of the J -coupling in the x-y plane generates components of the transversal magnetization vector in both the x and y axes. Because relaxation takes a much longer time (several ms) than RF pulses (μ s), these can be used to select these components and trace magnetization paths that yield multidimensional correlation spectra informing about the covalent atomic bonds in the molecule. The vectorial model (**Figure 3.7**) is insufficient to describe this and product operators are used.

Through-space transfer occurs between nuclei that are spatially close ($<5 \text{ \AA}$) through the Nuclear Overhauser Effect (NOE). This phenomenon is exploited to

study the atomic spatial arrangement and to characterize the three-dimensional structure of the molecules.

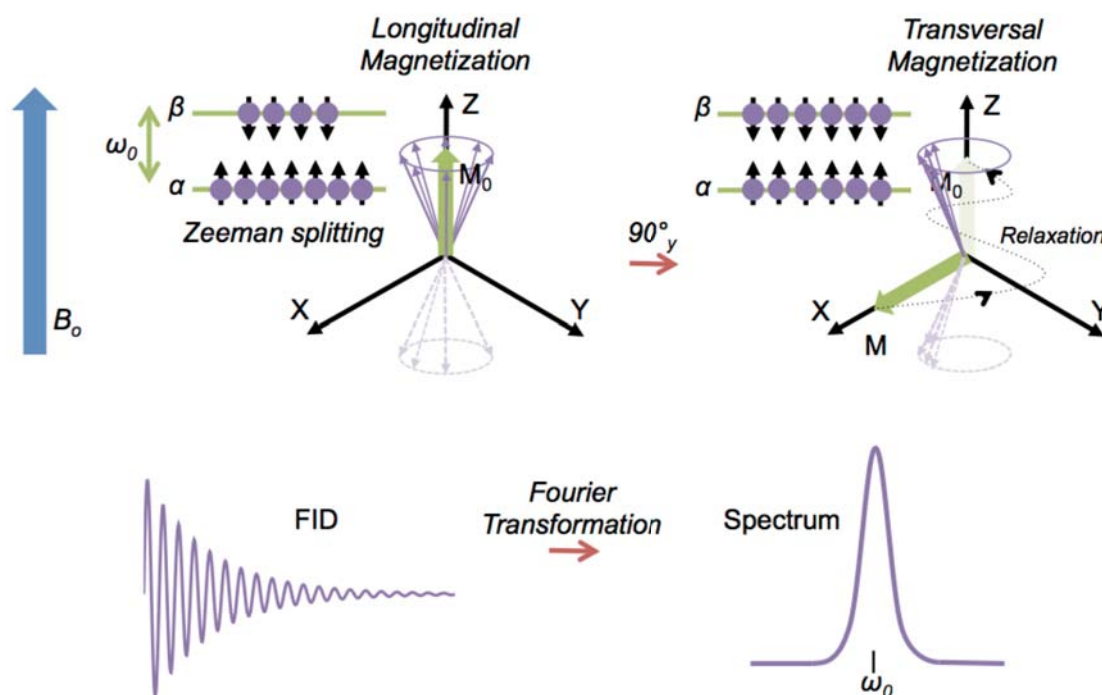


Figure 3.7: Physical basis of RF pulse-based NMR. A detailed explanation is provided in the main text.

3.4.1.2. Basic NMR characterization of a protein: 1D and HSQC

The acquisition of a ^1H **1D** experiment is usually the first step in the characterization of a protein sample. This experiment does not require the isotopic enrichment of the sample and gives information on the protein fold and stability. In a folded protein protons sample a large variety of well-defined chemical environments, so signals in the spectrum are dispersed in a wide range of frequencies ranging from 9 to almost 0 ppm. Unfolded proteins yield 1D spectra with much lower signal dispersion.

However, the number of protons in a protein is large and many signals can not be resolved (more than 300 in a WW domain, the smallest characterized protein domain). Therefore, the complexity of 1D experiments is too high to obtain further information.

Isotopic enrichment of the protein with ^{15}N allows the acquisition of the **HSQC** (heteronuclear single quantum coherence) experiment. This is a 2D experiment displaying correlation peaks between nitrogen atoms covalently bound to protons (**Figure 3.8**). Therefore, a peak is obtained for each amino acid's backbone amide

group (except for Pro), as well as for the amino groups in Lys and the aromatic rings of His and Trp; the amide groups in the side-chains of Asn and Gln; and the guanidinium group in the side-chain of Arg. Consequently, each peak is considered to roughly represent the chemical environment of each amino acid in the sequence. Similar to 1D experiments, the HSQC spectrum of a folded protein displays well-dispersed sharp signals in the 6 – 10 ppm range in the ^1H dimension and in the 135 – 105 ppm range in the ^{15}N dimension.

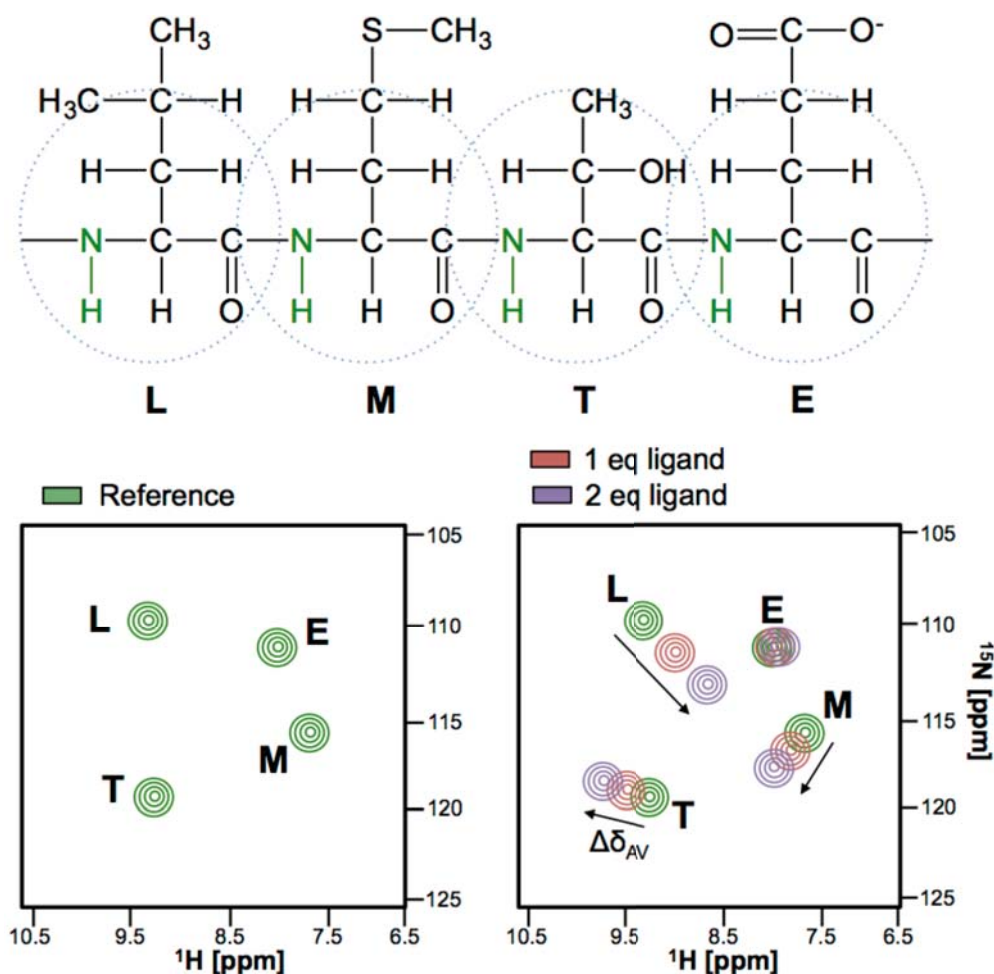


Figure 3.8: Illustration of an HSQC experiment and a chemical shift perturbation (CSP) assay. On top, the backbone of the peptide sequence *LQTE* is represented, with amide bonds and atoms coloured in green. Below, in the left panel, the HSQC spectrum of the sequence shows one peak per residue in the sequence. The addition of a ligand affects the chemical shift of some of the resonances in a different manner, which is evaluated by calculating the dimension-weighted averaged chemical shift differential ($\Delta\delta_{AV}$). In this example, the interaction occurs in the fast exchange regime.

HSQC experiments are very useful to characterize the interaction of a protein with another one, a peptide or a small ligand. In general, a series of HSQC experiments are acquired to monitor the protein resonances while titrating progressively increasing amounts of the interacting partner (Figure 3.8). Ligand binding will

cause changes in the electronic environment of residues directly involved in the interaction and also those suffering structural rearrangements. These changes will be reflected in the spectra as chemical shift perturbations (CSP). The nature of these perturbations depends on the exchange rate (k_{ex}) between the complex (k_{on}) and the free protein (k_{off}) (Equation 3.10).

$$k_{ex} = k_{on} + k_{off} \quad (3.10)$$

Three main exchange regimes can be distinguished depending on the relative value of k_{ex} with respect to the difference of the resonating frequencies in the bound and free states ($\Delta\omega$):

- **Fast exchange:** $k_{ex} \gg \Delta\omega$. This is the case for interactions that exist only during a fraction of the acquisition time. The observed chemical shift at each titration point is the weighted average of the chemical shifts in the free and bound forms. With the progressive addition of ligand, the peaks are displaced towards the value of the chemical shift in the bound form.
- **Intermediate exchange:** $k_{ex} \approx \Delta\omega$. In this situation, peak broadening and intensity loss are observed with the progressive addition of ligand.
- **Slow exchange:** $k_{ex} \ll \Delta\omega$. This is the case for stable interactions during the acquisition time, where protein molecules are either in the bound or the free state during the whole experiment. With the progressive addition of ligand, new resonances will appear at the chemical shift of the bound form, whereas the resonances of the free state will progressively disappear.

The HSQC titration experiment has been extensively used in this thesis to characterize bindings.

3.4.1.3. Sequential assignment of backbone resonances

The interpretation of the HSQC titration experiment requires the assignment of each resonance in the spectrum to a particular residue in the protein sequence. This is usually achieved using a 3D pair of experiments of the type of **HNCACB/HNCA(CO)CB** that allows the assignment of backbone ^{13}C and ^{15}N resonances.

These spectra display correlation peaks between the backbone amide proton, the amide nitrogen, C_α and C_β . For interpretation of the spectra, the nitrogen dimension is placed in the z-axis and planes displaying ^1H - ^{13}C correlations (x and y axes respectively) are obtained. Planes are decomposed in ^1H - ^{13}C strips, each corresponding to one resonance in the HSQC spectrum (**Figure 3.9**).

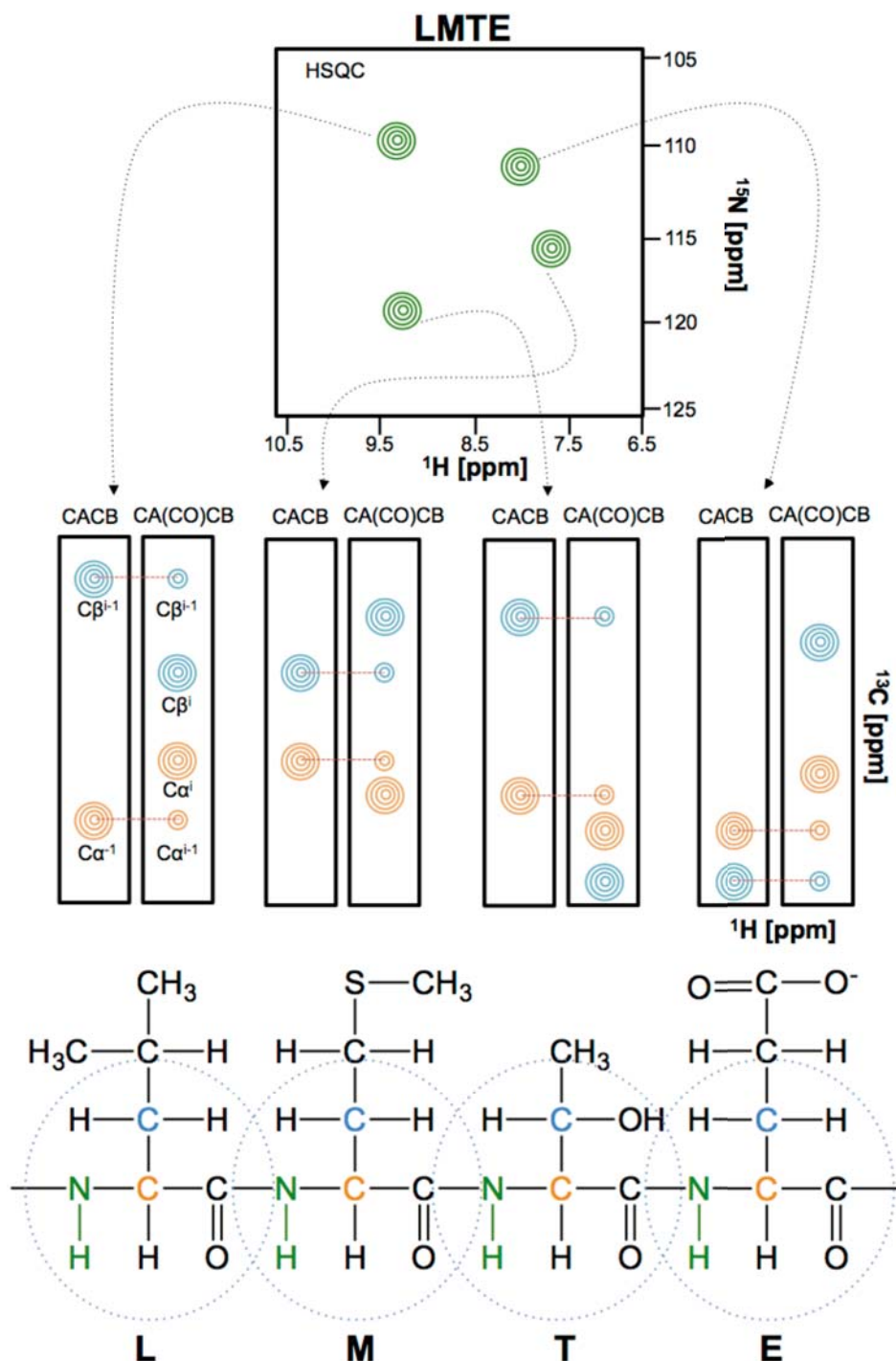


Figure 3.9: Sequential backbone resonance assignment using a HNCACB/HNCA(CO)CB pair of 3D experiments. Initially the protein sequence is known, but the assignment of HSQC peaks is not. A strip is generated from both 3D experiments displaying C_{α} and C_{β} resonances of residue $i-1$ (HNCACB) and i (HNCA(CO)CB), permitting the sequential assignment of the strips to residues in the protein sequence. C_{α} and C_{β} chemical shifts are characteristic of the amino acid type.

In the HNCACB experiment correlation peaks between a particular residue i backbone amide nitrogen and proton atoms with the C_{α} and C_{β} atoms of residue $i-1$ are selected, whereas in the HNCA(CO)CB experiment carbon signals from residue i

are observed as well. Notably, in the HNCA(CO)CB spectrum the intensity of the signals corresponding to residue $i - 1$ is lower than to residue i . Therefore, strips can be connected by searching for one particular residue's C_α and C_β resonances in the next strip.

The C_α and C_β chemical shifts are strongly dependent on the amino acid type. The most common values for each amino acid type are extensively characterized, and they are used to identify particular sequence segments and match them to strip sequences. Therefore, a particular ^1H - ^{15}N resonance in the HSQC spectrum can be assigned to a particular residue in the protein sequence.

Because some amino acid types have similar C_α and C_β chemical shifts, occasionally some resonances can not be assigned unambiguously. In these cases, the **CCC(CO)NH** spectrum provides information of side-chain ^{13}C resonances of residue i , helping to discern one amino acid type from another.

The chemical shifts are strongly correlated with local structure. Thus, the values of the C_α and C_β chemical shifts assigned to a particular residue are informative on the secondary structure element where it is located. This information can be extracted by calculating the secondary chemical shift $\Delta\delta$ (Equation 3.11).

$$\Delta\delta = \delta_{observed} - \delta_{randomcoil} \quad (3.11)$$

Thus, the observed chemical shift of a particular residue in the protein sequence is compared to the chemical shift of this residue type in a random coil. This is a reference value corresponding to the average of chemical shifts that a certain amino acid type can have in unstructured or disordered parts of a protein (random coils). Depending on the value of $\Delta\delta$, the main secondary structure elements can be discerned (Dalgarno et al., 1983):

- For residues located in **α helices** $\Delta\delta_{C_\alpha}$ values tend to be positive, whereas $\Delta\delta_{C_\beta}$ are generally negative. In this case, $\Delta\delta_{C_\alpha} - \Delta\delta_{C_\beta}$ will be positive.
- For residues located in **β sheets** $\Delta\delta_{C_\alpha}$ values are most likely negative, while $\Delta\delta_{C_\beta}$ are tend to be positive. In this situation, $\Delta\delta_{C_\alpha} - \Delta\delta_{C_\beta}$ will be negative.

3.4.1.4. Protein structural determination

The determination of a protein structure by NMR is based on distance restraints that place a particular atom in a defined position of the three-dimensional space relative to all other atoms in the molecule.

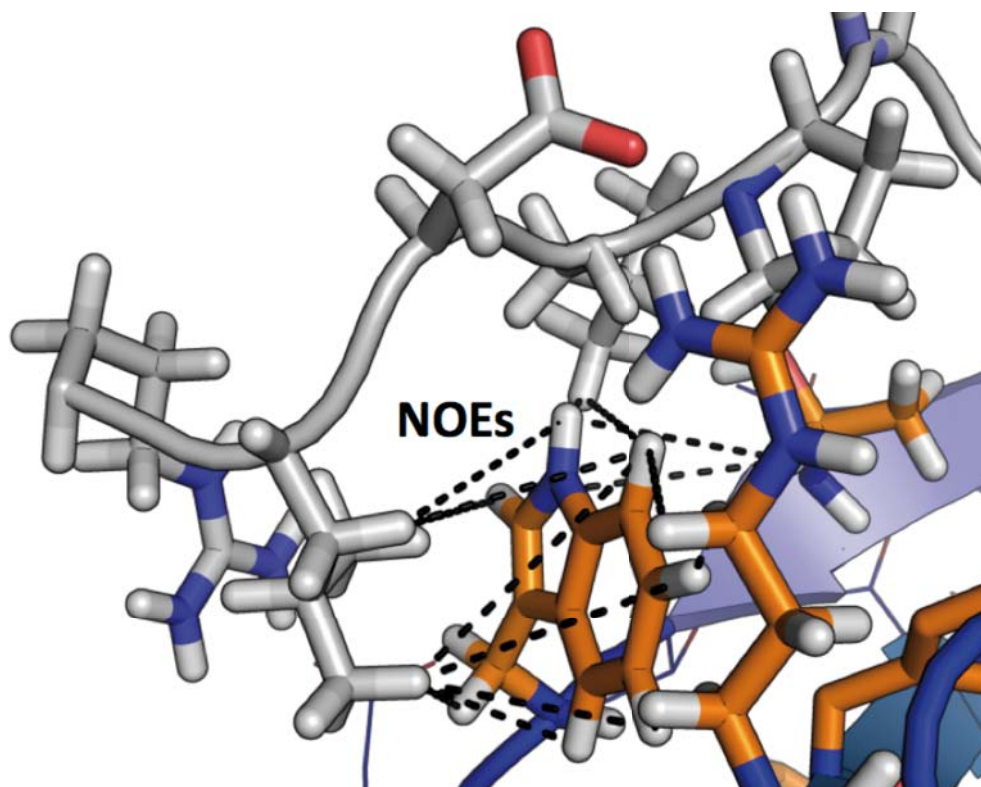


Figure 3.10: NOE restraints in protein structural determination. NOESY spectra display cross-peaks for nuclei located $<5 \text{ \AA}$ apart in the space. In the picture, black dashed lines represent intra and inter-molecular NOE-derived proton distance restraints for a protein (orange side-chains) – peptide (grey) complex.

Magnetization is transferred through the space in a phenomenon known as dipolar coupling. The effective magnetic field on a nucleus is affected by the presence of other nuclei with magnetic activity. The effect of one nucleus on another depends on the angle of the vector between both and the external magnetic field. Because of molecular tumbling caused by thermal energy, the dipolar coupling is averaged to zero in isotropic solutions. However, they are the source of the measurable Nuclear Overhauser Effect (NOE).

The NOE is a measurement of the cross-relaxation effect that two nuclei provoke on each other. It is inversely proportional to the sixth power of the distance between these nuclei, so it can be used as a measurement of this distance. NOE signals are obtained for nuclei located less than 5 \AA apart as the intensity of the effect decays rapidly with distance (**Figure 3.10**).

To determine the 3D structure of a protein by NMR, NOE distance restraints between the protons in the protein are obtained using **NOESY** (NOE Spectroscopy) experiments (**Figure 3.11**). In multidimensional NOESY experiments, the NOE effect gives rise to correlation signals whose integration is a direct measurement of

the inter-atomic distance. Homonuclear 2D ^1H - ^1H and heteronuclear 3D ^{13}C and ^{15}N -edited NOESY spectra are often used to acquire these restraints.

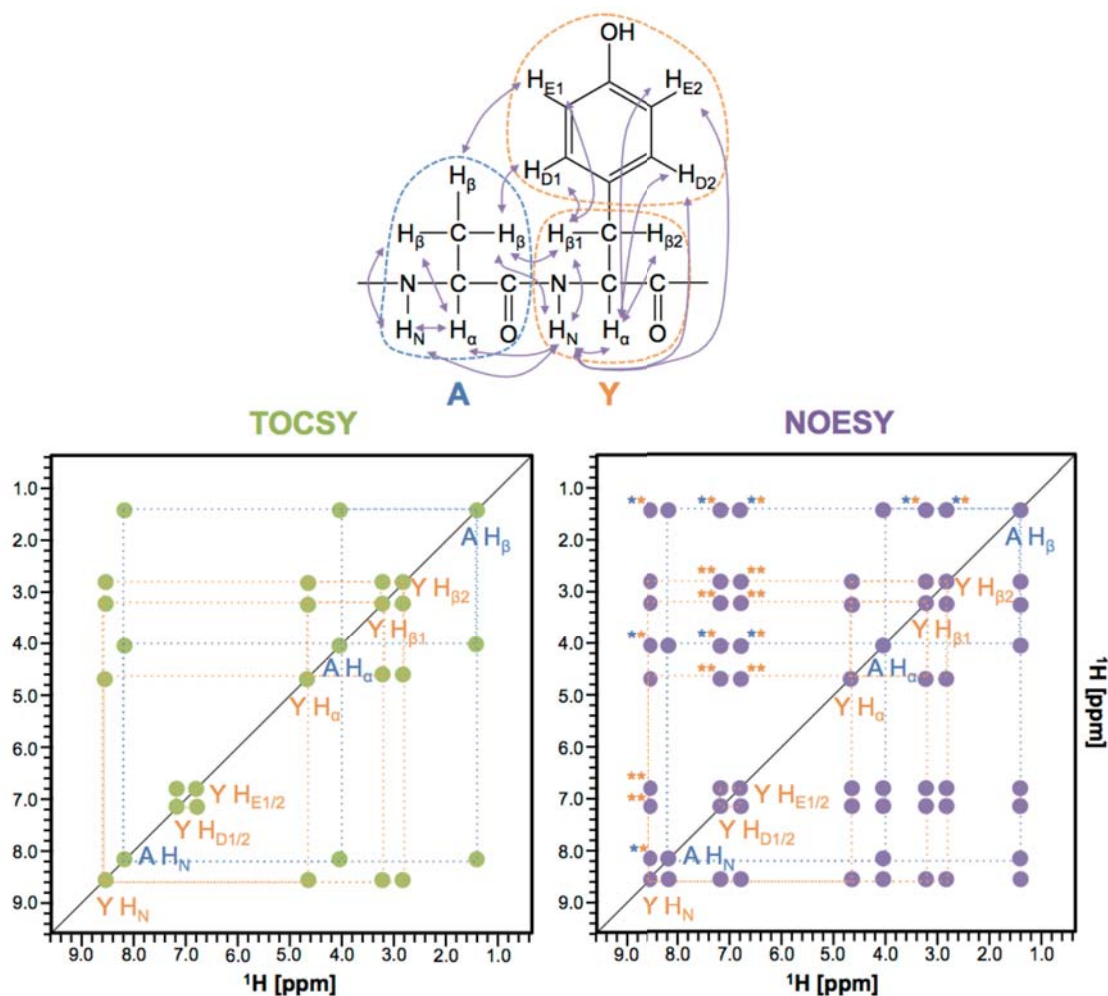


Figure 3.11: Assignment of TOCSY and NOESY spectra to obtain distance restraints in protein structural determination. The dipeptide AY (Alanine – Tyrosine) is used as an example. On top, the chemical formula of the sequence is shown. The side-chain of A constitutes a spin system (blue balloon, β protons are chemically equivalent), whereas Y has two spin systems (aliphatic chain and aromatic ring, orange balloon). In the 2D ^1H - ^1H TOCSY spectrum, there is a cross-peak correlating every nucleus in a single spin system (located at the vertices of the blue squares for A and orange squares for Y). The chemical shift of each nucleus in a particular amino acid type has a value within a known range. The assignment of one residue's side-chain (shown for the peaks in the diagonal) involves matching the TOCSY connectivity pattern expected for that particular residue. In the 2D ^1H - ^1H NOESY spectrum, there is a cross-peak correlating every nucleus that is located $< 5 \text{ \AA}$ apart in the space (indicated by purple arrows on the chemical formula). NOEs are observed for nuclei within the spin system and outside it, as long as they are located in the mentioned spatial range. In the figure, NOEs outside the spin systems are indicated with a double asterisk. One orange asterisk plus one blue asterisk are used to label peaks that correlate nuclei located in the two different amino acids. Two orange asterisks denote NOEs correlating atoms located in two different spin systems of Y. NOEs between nuclei located in the same spin system are at the vertices of orange and blue squares and give redundant information.

Many intra-residue signals are usually observed, as well as signals with residues close in the sequence usually defining secondary structure elements. However, the most valuable NOE signals defining the fold of a protein are those obtained for nuclei that are apart in the sequence but close in the space. When solving the structure of a protein-ligand complex, inter-molecular NOEs help defining the relative orientation of the molecules and the binding mode.

NOESY spectra are usually very crowded with signals and individual amino acids are hard to identify, as intra and inter-residue signals are mixed. To help with the assignment, **TOCSY** (Total Correlation Spectroscopy) experiments are acquired (**Figure 3.11**). Taking advantage of J-couplings this experiment displays correlation peaks between all the atoms in a spin system. Each amino acid's side-chain has a characteristic spin system, and some of them have two (aromatic amino acids). Therefore, all nuclei corresponding to a single residue can be identified and assigned. Again, homonuclear 2D ^1H - ^1H and heteronuclear 3D ^{13}C and ^{15}N -edited TOCSY spectra are often acquired to obtain this data.

3.4.1.5. Relaxation and dynamics

Relaxation is the process by which coherent (precessing in phase) transversal magnetization is converted into de-phased magnetization aligned with the z-axis (longitudinal). There are two relaxation mechanisms:

- **Longitudinal relaxation T_1** is the name given to the re-alignment of transversal magnetization with the z-axis (B_0). Typical values found in proteins are in the range of several hundred milliseconds to a few seconds. Because longitudinal magnetization is required at the start of every experiment scan, the T_1 is a limiting factor of the number of scans that can be acquired in a certain time.
- **Transverse relaxation T_2** consists in the de-phasing of the transverse magnetization in the x-y plane causing the loss of coherence. It is caused by stochastic fluctuations of the magnetic field B_0 that have different effects on the molecules of the sample. In proteins, it has values in the order of several tens of milliseconds.

Both T_1 and T_2 depend on molecular tumbling, described by the correlation time τ_c . Fast tumbling averages the small magnetic field inhomogeneities that cause coherence loss. In folded proteins, the bigger the protein the slower the tumbling and the bigger τ_c . Folded protein domains below 30 kDa have relatively favourable

relaxation properties, and quality data can be obtained with domains of up to 50 kDa. Beyond this value, the T_2 relaxation is so fast that signals can not be obtained.

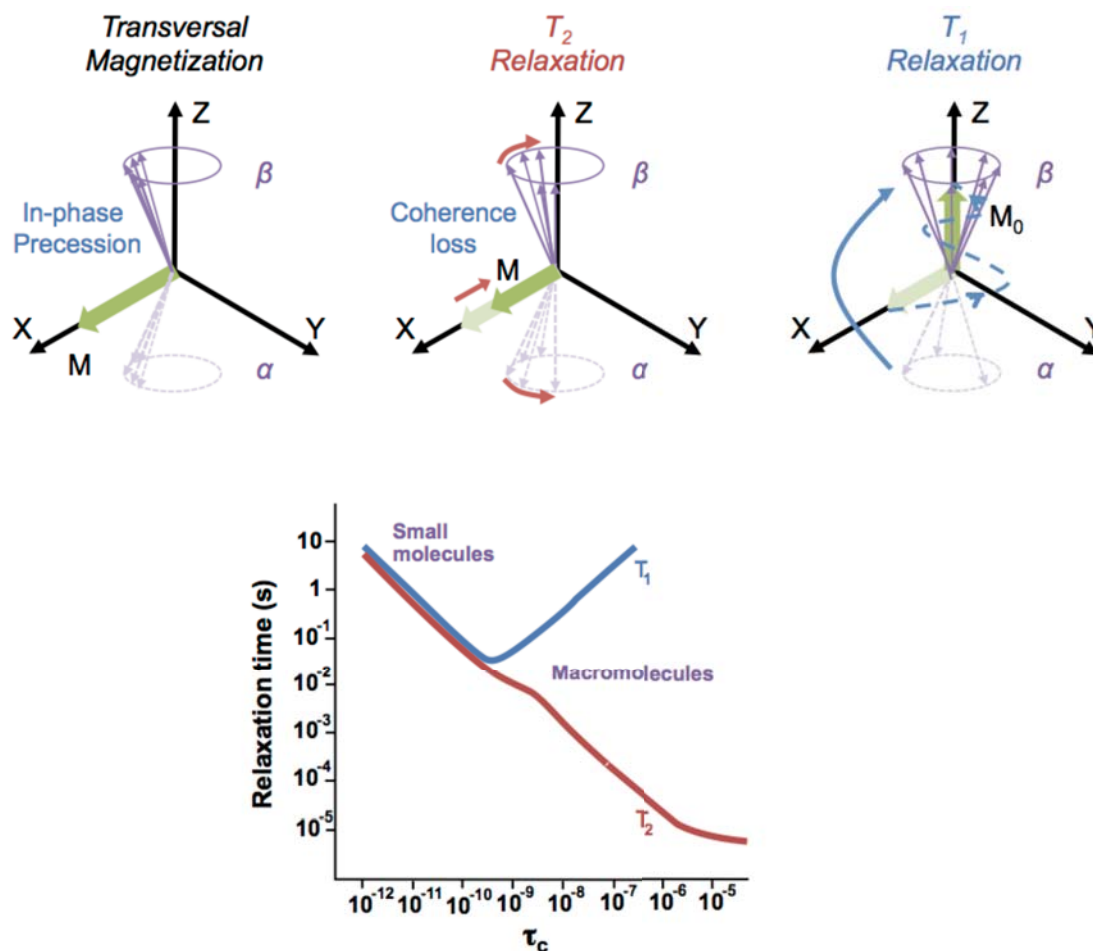


Figure 3.12: T1 and T2 relaxation mechanisms. On top, vectorial representation of the concepts associated with T_2 and T_1 relaxation mechanisms detailed in the main text. Below, plot showing the dependence of both relaxation rates with the correlation time τ_c , which increases with molecular size.

Because deuterons do not take part in ^1H - ^1H dipolar couplings, deuteration of big proteins is frequently used to extend the T_1 and T_2 relaxation rates, sensibly improving the quality of the spectra. Deuteration is achieved by inducing protein expression in a cell culture grown in a medium prepared in D_2O rather than in H_2O . Backbone amide protons are labile and exchange rapidly with protons in the buffer, so dissolving a deuterated sample in a H_2O buffer will result in the exchange of deuterons in these positions with hydrogen atoms in the buffer. The result is a fully deuterated sample except for the backbone amide positions, permitting the acquisition of HSQC and HNCACB/HNCA(CO)CB-like experiments with enhanced favourable relaxation properties.

Disordered regions within a folded protein such as long loops or flexible hinges connecting domains often tumble independently from the rest of the molecule. Both T_1 and T_2 values of residues located in these regions are higher than those observed in residues located in ordered regions. Therefore, the measurement of each residue's T_1 and T_2 values yields information on the internal dynamics of the protein, permitting the identification of folded and flexible areas.

The ^{15}N T_1 and T_2 values of each residue's backbone amide nitrogen nucleus can be measured using series of HSQC experiments. To obtain T_1 measurements, the delay in the spin echo sequence during ^{15}N mixing is progressively increased. The longer the delay, the lower the magnetization remaining in the x-y plane. In the T_2 measurements, the spins are not refocused during ^{15}N mixing. The longer the mixing time, the higher the spin de-phasing, thus increasing the signal linewidth. In both cases, the increase of the ^{15}N mixing time makes the HSQC signals to decay exponentially so relaxation values can be obtained by fitting a curve to this decay.

Internal flexibility measurements are often complemented with the measurement of the ^1H - ^{15}N **heteronuclear NOE** experiment, which provides information on the motion of backbone amide N-H bonds. The relative NOE intensity between these two nuclei is obtained from two HSQC spectra with and without proton pre-saturation, which are acquired in an interleaved manner. Residues in folded regions display $I_{\text{sat}}/I_{\text{insat}}$ ratio values around 0.8, while those located in flexible areas display smaller values.

3.4.1.6. Paramagnetism

Paramagnetic effects are caused by unpaired electrons, whose strong magnetic moment affects the relaxation properties of spatially close nuclei. Depending on the type of relaxation affected, signals may be broadened as in **PRE** (Paramagnetic Relaxation Enhancement) or shifted as in PCS (Pseudo-Contact Shifts). Some paramagnetic centres also cause partial internal molecular alignment that is used to measure RDCs (Residual Dipolar Couplings) (Lian and Roberts, 2011; Su and Otting, 2010). These effects are distance and sometimes orientation-dependent, so restraints can be obtained. Distance restraints up to 60 Å can be measured, which is a twelve-fold longer distance than that sampled with NOEs (up to 5 Å). Therefore, paramagnetism offers the possibility to acquire data for the characterization of protein structures and protein-protein interactions.

PREs are particularly useful in protein NMR because distance restraints can be directly extracted from HSQC spectra. The intensity ratio of a particular peak in the

presence and absence of a paramagnetic centre is measured and converted into a restraint.

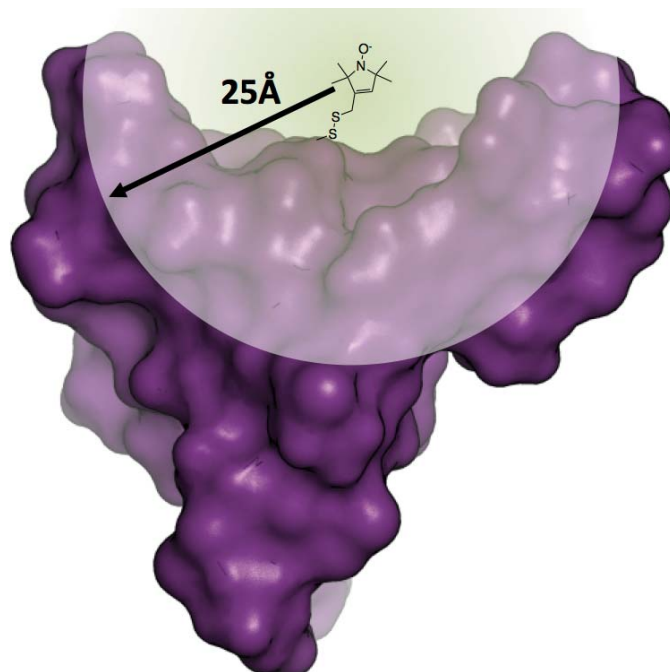


Figure 3.13: Paramagnetic Relaxation Enhancement with a nitroxide tag. Distance-dependent PRE restraints of up to 25 Å can be measured by derivatizing a protein with a nitroxide spin label.

In metal-coordinating domains, a paramagnetic centre can be introduced in a protein by substituting a diamagnetic metal like Ca^{2+} with a paramagnetic one like a lanthanide. Another option is to derivatize the protein with a stable radical used as a paramagnetic tag. The most commonly used tags are the nitroxide labels, which provide distance restraints up to 25 Å.

Usually the thiol group in the side-chain of cysteine residues is used to covalently bind the tag. Using site-directed mutagenesis, the tag is artificially located in the position of interest whereas naturally occurring cysteine residues are mutated to avoid undesired labelling. The obtained distance restraints can be used in structure calculations and in data-driven docking procedures for the characterization of protein-protein interactions.

3.4.2. Experimental procedures

3.4.2.1. Sample preparation and instrumentation

For NMR experiments, purified protein and peptide solutions containing 10% D_2O were adjusted to pH 7 when required and loaded into conventional or Shigemitsu tubes

(Shigemi®, Tokyo, Japan). The experiments were acquired on a Bruker Avance III 600-MHz spectrometer equipped with a z-pulse field gradient. The temperature was set to 298 K unless otherwise stated. ^{13}C -edited experiments were recorded using a triple resonance probe head (^1H , ^{13}C , ^{15}N). ^{31}P -edited experiments were recorded using a quadruple resonance probe head (plus ^{31}P). All other experiments were acquired equally in any of these two probes.

The HSQC spectrum of the HECT domain was acquired at 298 K on a Bruker Avance 900-MHz spectrometer equipped with a z-pulse field gradient and a 5 mm TCI cryoprobe at the *Forschungsinstitut für Molekulare Pharmakologie* (FMP-Berlin). A 90 μM ^2H - ^{15}N labelled sample containing 10% D_2O was used for the experiment.

3.4.2.2. Chemical shift perturbation experiments

Protein-protein HSQC titrations were performed using a constant concentration of the ^{15}N -labelled domain (100 μM) and adding progressively increasing amounts of the unlabelled domain until the interaction was saturated. Each domain was purified in the same buffer, and then mixed and concentrated together to a final volume of 400 μL (including 10% D_2O). Samples used in the WW-HECT C-lobe titrations were equilibrated in 20 mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ and 100 mM NaCl pH 7. Samples used in the C2-HECT C-lobe and C2-HECT titrations were equilibrated in 20 mM Tris-HCl, 250 mM NaCl , 1 mM EGTA (pH 7) and when required 20 mM CaCl_2 .

Titrations of the HECT-PY peptide and each of the four ^{15}N -labelled WW domains were performed using 200 μM samples of the protein domains. The peptide was dissolved in H_2O at a concentration of 10 mM and the pH of the solution was neutralized. Small amounts of the solution (8 μL) were added to the protein sample at each experimental point until saturation of the binding.

Titrations of the IP_3 inositol phosphate and the ^{15}N -labelled C2 domain in the presence and absence of Ca^{2+} were performed in an equivalent manner. At every experimental point, a 1D ^{31}P spectrum was acquired after recording the HSQC experiment.

The analysis of chemical shift perturbations was performed using the CCPNMR Analysis software (Vranken et al., 2005) using a 0.2 weighting of the ^{15}N dimension relative to ^1H (Equation 3.12).

$$\Delta\delta_{AV} = \sqrt{(\Delta\delta_{1H})^2 + \frac{1}{5}(\Delta\delta_{15N})^2} \quad (3.12)$$

3.4.2.3. Resonance assignment

In the case of the WW3-HECT PY complex (protonated sample), backbone ^1H , ^{13}C and ^{15}N assignments were obtained by combining the information from the 3D CBCANH/CBCA(CO)NH experiment pair. Instead, the equivalent 3D HNCACB/HN(CO)CACB experiment pair was acquired and analysed in the case of deuterated samples (HECT C-lobe and C2 domain). Assignment ambiguity was solved when possible by analysing a complementary 3D CCC(CO)NH spectrum. In some cases, TROSY (Transverse Relaxation-Optimized Spectroscopy) versions of these experiments and/or the NUS (Non-Uniform Sampling) acquisition strategy were used to optimize the acquisition time and improve resolution. Ionic strength of the C2 domain buffer was reinforced (400 mM NaCl) to avoid its precipitation during experimental acquisition. All other buffers were as described in Section 3.4.2.2.

For the structural determination of the WW3-HECT PY complex, 3D ^{15}N -TOCSY and ^{15}N -NOESY experiments were used to obtain side-chain assignments. Most intra-molecular NOE proton distance restraints were obtained from the analysis of the 3D ^{15}N -NOESY spectrum. Complementary intra-molecular as well as inter-molecular restraints were assigned in the 2D ^1H - ^1H homonuclear NOESY spectrum of the complex.

Spectra processing was performed with NMRPipe (Delaglio et al., 1995). Spectra analysis and assignment were performed using CCPNMR Analysis (Vranken et al., 2005) and CARA (Bartels et al., 1995). Peak integration of spectra used in structure calculation was performed using the batch integration method of the XEASY package.

3.4.2.4. Structure determination and refinement

Unambiguously assigned NOE inter and intra-molecular restraints were used for the structure calculation. This was performed using the software CNS 1.1 (Brunger et al., 1998). The software TALOS+ (Shen et al., 2009) was used to derive secondary structure-based hydrogen bond patterns as well as backbone torsion angles ψ and ϕ from the $\text{C}\alpha$ and $\text{C}\beta$ resonance assignment. 120 structures were calculated, representing a good fit of the experimental data. All of them were submitted for water refinement using the software Aria (Nilges et al., 1997) with an in-house modified protocol using all experimental restraints during refinement.

The 20 lowest energy structure ensemble was selected and its quality was assessed using the iCing package (Doreleijers et al., 2012), which uses PROCHECK-NMR

(Laskowski et al., 1996) to yield the structure calculation statistics shown in **Table 6.3**. The software PyMOL (DeLano, 2002) was used to generate all structure images displayed in this thesis.

3.4.2.5. Relaxation measurements

Relaxation measurements were acquired using a 500 μM ^{15}N -labelled HECT C-lobe sample. For the measurement of the ^{15}N T_1 values, a series of twelve HSQC spectra were acquired in an interleaved manner with progressively increasing ^{15}N spin echo delays: 20.8, 52, 104, 156, 265, 424, 520, 676, 832, 1040, 1352 and 1664 ms. The ^{15}N T_2 values were measured using a series of ten HSQC spectra with increasing ^{15}N mixing delays: 12, 24, 36, 48, 60, 72, 84, 120, 144 and 168 ms. In both cases, the peaks were integrated using NMRPipe (Delaglio et al., 1995). Using the software Prism, data corresponding to each peak was fitted to an exponential decay curve (Equation 3.13), where I_0 is the normalized intensity at time 0 and $I(t)$ is the relative intensity at data point t .

$$I(t) = I_0 e^{\left(\frac{-t}{T_{1,2}}\right)} \quad (3.13)$$

^1H - ^{15}N heteronuclear NOE data were acquired on the same sample. Two HSQC spectra were acquired in an interleaved fashion, one with proton pre-saturation and the other without it (reference). Remarkably, the inter-scan delay was set to 8 s to allow z -magnetization recovery after pre-saturation. The peak intensities were integrated using NMRPipe and the ratios between the values obtained in the pre-saturated (I_S) and the reference (I_0) spectra yielded the steady-state ^1H - ^{15}N NOEs (Farrow et al., 1994). The NOE standard deviation was calculated using Equation 3.14, which required the measurement of the background noise levels in the pre-saturated (σ_S) and the reference (σ_0) spectra.

$$\frac{\sigma_{NOE}}{NOE} = \left(\left(\frac{\sigma_S}{I_S} \right)^2 + \left(\frac{\sigma_0}{I_0} \right)^2 \right)^{\frac{1}{2}} \quad (3.14)$$

3.4.2.6. PRE experiments

Paramagnetic relaxation enhancement (PRE) measurements were performed using two mutants of the Nedd4L WW3 domain, each with an artificially introduced cysteine residue derivatized with the paramagnetic nitroxide label MTSL and its acetylated diamagnetic counterpart MTS. Mutant DNA was generated using site-directed mutagenesis (Section 3.1.1.2). The proteins were expressed in inclusion bodies, solubilized in denaturing buffer, refolded by dialysis after Ni^{2+} -affinity

purification and digested with the TEV protease (Section 3.1.2.5). The proteins were incubated for 2 h in a 10 mM DTT containing buffer to reduce cysteine residues. After that time, DTT was removed by buffer exchange and 7-10 equivalents of MTSL or MTS were added to the solution. The reaction proceeded overnight at room temperature and protected from the light, and samples were analysed using MALDI-TOF (Section 3.2.1.9) after de-salting a small aliquot using a ZipTip® (Millipore, Billerica, MA, USA) for small-scale reversed-phase chromatography. A final size-exclusion chromatography purification step was performed using an analytical Superdex™ Peptide 10/300 GL column on a Äkta FPLC system (GE Healthcare, Uppsala, Sweden). Pure fractions were mixed and concentrated to a final volume of 400 µL and a final concentration of 200 µM.

HSQC spectra were acquired on both the MTSL and the MTS derivatized samples, and the HECT-PY peptide was titrated up to the saturation point in all of them. Using the spectra of the WW3 derivatized mutants bound to the HECT-PY peptide, the intensity of the peaks was integrated with NMRPipe and normalized with respect to the intensity of the peak corresponding to Asn515, which is located more than 25 Å away from both derivatized positions and is unaffected by the binding of the HECT-PY peptide. For each peak, the ratio between its intensity in the presence of the paramagnetic tag MTSL (I_{MTSL}) and the diamagnetic tag MTS (I_{MTS}) was calculated and plotted on the structure of the Nedd4L WW3 domain.

3.5. Molecular dynamics simulations

3.5.1. System equilibration

Molecular dynamics simulations were performed on the HECT C-lobe fragment (residues 836-949) extracted from the crystal structure of the Nedd4L HECT domain (PDB entry 3JW0). The GROMACS package (Hess et al., 2008) was used to perform and analyse the simulations. The AMBER03 force field and TIP3P model for explicit water were used in the simulations. To guarantee electroneutrality 0.1 M of NaCl ions were added to the system. The solvated 20378-atom system was equilibrated with 1000 steps using the steepest descent algorithm implemented in GROMACS. Prior to the final production simulation the system was equilibrated using the NPT ensemble for 500 ps, followed by 50 ps in the NVT ensemble. Equilibration was assured between steps, by checking the convergence of energetic parameters.

3.5.2. Trajectory at 310 K

After equilibration, the system was simulated for 120 ns using a 2 fs time ramp. Temperature coupling was done with the Nose-Hoover algorithm at 310 K. Pressure coupling was done with the Parrinello-Rahman algorithm at 1 bar. The electrostatic interactions were treated using the Ewald particle mesh methods, also implemented in GROMACS.

3.5.3. Simulated annealing simulations

For simulated annealing simulations the system was gradually heated from 310 K to 600 K over 40 ns, starting from the 310 K equilibrated structure. The equilibration protocol was the same as used for the full 120 ns molecular dynamics trajectory.

Part IV.

Results

CHAPTER 4. SMAD LINKER PHOSPHORYLATION AFFECTS SUBSTRATE BINDING CAPABILITIES OF NEDD4L AND RELATED PROTEINS	91
4.1. Introduction	91
4.2. Solid-phase peptide synthesis	94
4.3. Standard procedures used in the reported syntheses	102
4.4. Synthesis of a set of Smad1 phosphopeptides	104
4.5. Synthesis of a set of Smad3 phosphopeptides	113
4.6. Native chemical ligation of a phosphopeptide	118
4.7. Summary	123
CHAPTER 5. ACTIVITY REGULATION OF THE E3 UBIQUITIN LIGASE NEDD4L: CA²⁺/IP₃ SIGNALLING TRIGGERS THE TRANSITION FROM THE LATENT TO THE ACTIVE CONFORMATION	127
5.1. Introduction	127
5.2. Deciphering the contacts that maintain Nedd4L in a latent state and its switch to the active form	130
5.3. Summary	148
CHAPTER 6. FULLY FUNCTIONAL NEDD4L MOLECULES TARGET PARTIALLY UNSTRUCTURED COUNTERPARTS AS A MEANS TO REGULATE THE TURNOVER RATE OF THE LIGASE	151
6.1. Introduction	151
6.2. Ubiquitination of the ubiquitin ligase: a bi-molecular reaction	154
6.3. Summary	169
CHAPTER 7. PARAMAGNETIC LABELLING OF THE NEDD4L WW3 – HECT-PY COMPLEX	173
7.1. Introduction	173
7.2. MTSL derivatization of Nedd4L WW3 cysteine mutants	174
7.3. Summary	179

CHAPTER 4. Smad linker phosphorylation affects substrate binding capabilities of Nedd4L and related proteins[‡]

4.1. Introduction

In the TGF- β signalling pathway, the receptor-activated R-Smads associate with the Co-Smad and are shuttled into the nucleus, where they modulate gene expression in cooperation with other effector proteins (**Figure 1.10**). Pin1 is a peptidyl-prolyl *cis/trans* isomerase that acts together with Smad2/3 to transduce signals in the TGF- β signalling pathway. In BMP signalling, the transcription factor YAP (Yes-Associated Protein) contributes with Smad1 signalling.

The signal transducing activity of the R-Smads is regulated at different levels. Their turnover rates are tuned by E3 ubiquitin ligases, which label them for degradation. Nedd4L and Smurf1 target Smad3 and Smad1 respectively. All the mentioned Smad binders (Pin1, YAP, Smurf1 and Nedd4L) are modular multi-domain proteins with different functions. They all have WW domains that they use to bind Smads (**Figure 4.1**).

Smad proteins have all a similar architecture consisting of two MH (Mad Homology) domains. The MH1 domain binds DNA, whereas the MH2 domain interacts with partner proteins. These domains are connected through a flexible linker that contains a PY motif. Additionally, R-Smads have an adjacent cluster of S/TP sites. The WW domains of Pin1, YAP, Smurf1 and Nedd4L bind the Smads in this linker region depending on its phosphorylation state.

Once the R-Smads are in the nucleus, the S/TP cluster in their linker region is phosphorylated by two kinases, namely CDK8/9 (cyclin-dependent kinase 8/9) and GSK3- β (glycogen synthase kinase-3 β). First, CDK8/9 phosphorylates its defined positions on the Smad sequence. This primes GSK3- β to phosphorylate its sites (positions $n-4$, where n are the CDK8/9 sites).

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

The action of these two kinases conditions the binding of other proteins to the R-Smads. The main question arising from such scenario refers to how the R-Smads manage to mediate signal transduction in cooperation with partners such as Pin1 and YAP before being labelled for disposal by E3 ubiquitin ligases like Smurf1 and Nedd4L.

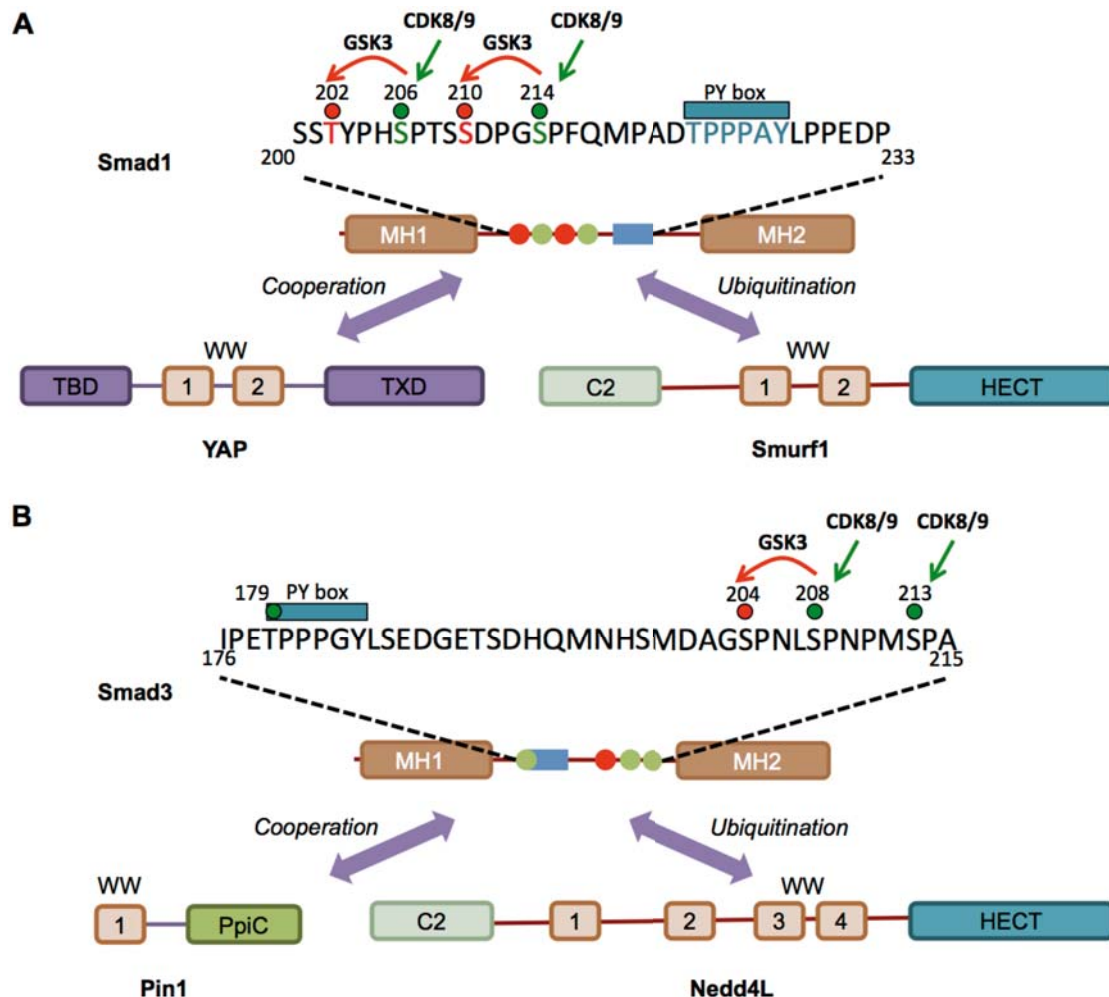


Figure 4.1: Phosphorylation of the R-Smad linker region creates docking sites for WW domains. (A) The modular structure of Smad1 is displayed together with a magnification of the linker region containing the phospho-S/TP sites and the PY box. Both YAP and Smurf1 interact with this region in Smad1 using their respective WW domains. (B) Similar to (A) but for the case of Smad3. Pin1 and Nedd4L use their WW domains to bind the phosphorylated Smad3 linker.

We hypothesized that this may respond to a chronological order of events based on hierarchical phosphorylation. In collaboration with the group of Dr Joan Massagué (Memorial Sloan Kettering Cancer Centre, New York) we found evidence in this direction. Molecular biology experiments performed at Dr Massagué's laboratory using HEK293 (human embryonic kidney) cell cultures showed that incubation with both TGF- β /BMP cytokines and a CDK8/9 pharmacological inhibitor (flavopiridol) prevented the establishment of all R-Smad complexes with their

binding partners cited above (Aragón et al., 2011). In contrast, when LiCl was used instead of flavopiridol (LiCl is a pharmacological inhibitor of GSK3- β) only the interactions with the ubiquitin ligases were hampered. Oppositely, the interactions with Pin1 and YAP were slightly favoured (**Figure 4.2**).

These experiments suggested a switch mechanism for the kinase-mediated regulation of the R-Smad proteins. CDK8/9 enables them to carry out their function by favouring the interactions with Pin1 and YAP. Later, GSK3- β phosphorylation results in R-Smad disposal by switching the Smads preferred binding partners to the E3 ubiquitin ligases Smurf1 and Nedd4L.

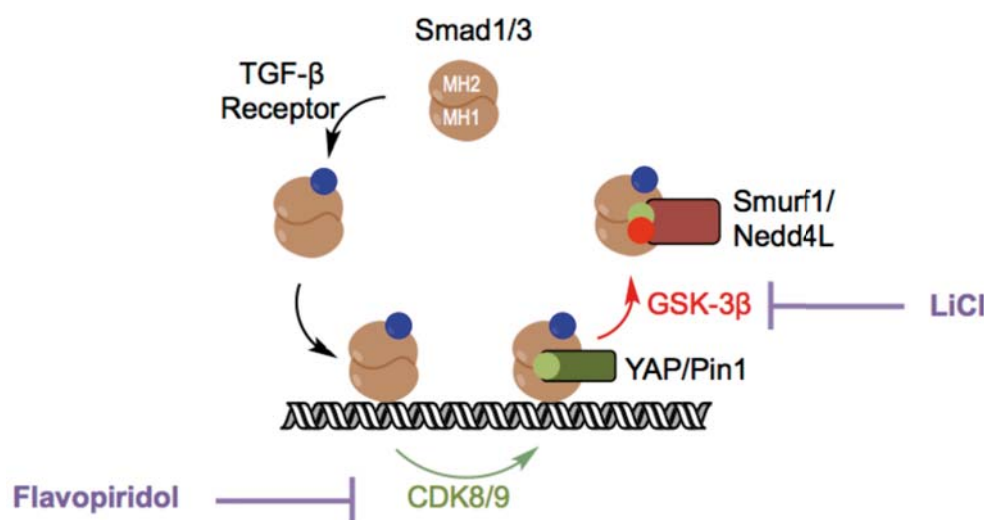


Figure 4.2: Pharmacological inhibitors of Smad kinases affect association with partner proteins. Flavopiridol blocks CDK8/9 action, which prevents the interactions with all partner proteins. LiCl blocks GSK3- β activity, which favours YAP/Pin1 binding to Smad1/3 and avoids the interaction with the ubiquitin ligases Smurf1 and Nedd4L.

We set to investigate the biophysical and structural bases governing this switch mechanism. We hypothesized that proteins with multiple WW domains (Nedd4L, Smurf1 and YAP) may bind both the PY motif and the phospho-S/TP cluster using tandems of such domains, while Pin1 would just bind the PY motif in Smad3 with its single WW domain. The preferred R-Smad binding partners would depend on the phosphorylation state of the S/TP cluster, resulting in an efficient way to regulate the activity of the associated proteins by modulating their substrate binding capabilities. We planned to measure the affinity and to perform structural studies of the interactions involving tandem or single WW domains constructions with peptides representing different phosphorylation states of the R-Smads.

The work presented in this thesis chapter consisted in the synthesis of two sets of peptides representing the linker regions of Smad1 and Smad3. All of them contain

the PY motif, and peptides in each set differ in the phosphorylation state of the S/TP cluster. These were synthesized using solid-phase peptide synthesis (SPPS). The incorporation of phosphorylated residues to relatively long sequences required the optimization of several steps in the synthesis related to amino acid coupling, peptide cleavage and purification.

4.2. Solid-phase peptide synthesis

Since its introduction by Merrifield in 1963 (Merrifield, 1963), solid-phase peptide synthesis (SPPS) has become an universal technique for the synthetic generation of polypeptides. Today it is broadly used to specifically engineer and introduce [various](#) modifications in [peptide](#) sequences, such as unnatural or [chemically](#) modified residues. Given the biological activity of peptides and their variants, this has been extensively exploited in drug design, despite the difficulties that the field permanently needs to deal with (low stability and bioavailability, difficult transmembrane transport, solubility issues, etc.).

For biochemical studies in the context of basic research, SPPS constitutes an advantageous tool to produce large amounts of peptides containing one or more biologically relevant modifications. Post-translational modifications (PTM) that polypeptides receive in living cells such as phosphorylation, glycosylation, acetylation, methylation, lipidation, etc. can be introduced in the synthetic sequence. This is valuable in *in vitro* biophysical and structural studies based on techniques such as NMR, ITC or crystallography. These usually require large amounts of highly pure samples.

Organisms commonly used as biological factories to produce samples for such studies usually apply PTM patterns that thoroughly differ from human ones, as it is the case of yeasts. Closer systems based on insect or mammal cell cultures (including humans) present the inconveniences of low yields and sample inhomogeneity. Widely used prokaryotes such as *E. coli* are unable to perform many of such modifications.

Expression and purification of short amino acid sequences (about ≤ 20 residues) in such biological factories is usually hampered too. Typical problems include low expression yields (if not null) combined with low efficiency of chromatography-based purification techniques (steps based on affinity tags give low yields and crude lysates are often hard to resolve by size exclusion and reversed-phase chromatography). SPPS is the technique of choice in all these cases given its high yields and relatively low complexity of the crudes for such short sequences.

The introduction in the mid 1990s of chemical ligation techniques, including native chemical ligation (NCL) (Dawson et al., 1994) and disulphide-bridge based (Baca et al., 1995), boosted the capabilities of SPPS and broadened the possibilities for its application in *in vitro* studies. The technique permits the fusion of synthetic and recombinant peptide segments. Entire globular proteins have been produced in this manner (Kent et al., 2012).

4.2.1. Fundamentals of the technique

In SPPS, the amino acid chain is grown on a solid polymeric functionalized support (the resin) (**Figure 4.3**). An organic linker is connected to the solid support with a free reactive group (hydroxyl, amino...) that is cleavable under certain conditions, usually acidic. The sequence is grown by reacting the free reactive group of the linker with the residue's backbone carboxyl group. Except for this one, amino acids used in SPPS have all other functional groups protected. The chemical groups used to protect the side chains and the backbone amino group are labile in different chemical conditions (orthogonal protecting strategy). This permits the selective deprotection of the backbone amino group no effects on the side-chain protection.

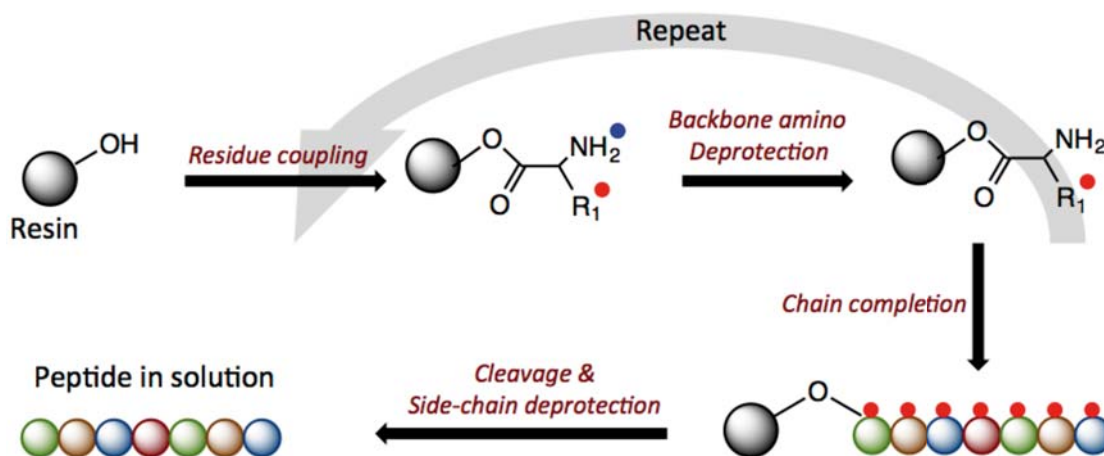


Figure 4.3: Fundamentals of SPPS. The amino acid chain is grown on a solid support (resin). Protected amino acid derivatives are added one by one. Using an orthogonal protecting strategy, backbone amino groups are deprotected without affecting side-chain protecting groups. Once the chain is completed, the peptide is cleaved from the resin and the side-chains are deprotected to yield the final peptide in solution.

The residues are added stepwise as building blocks, sequentially from the C- to the N-terminus. After the addition of the first amino acid to the resin linker, its backbone amino group is selectively deprotected and prepared to react with the backbone carboxyl group of the next amino acid in the sequence. These steps are cyclically repeated up to the completion of the synthesis. The peptide is then fully deprotected and cleaved from the resin using strong acidic conditions.

Two different strategies are commonly used in SPPS (**Table 4.1**). They mainly differ in the chemical nature of the protecting groups and the deprotection/cleavage reagents.

	BOC/Bzl	Fmoc/^tBu
N_α protecting group	Boc	Fmoc
N_α deprotection	TFA	Piperidine
Side-chains protecting groups	Benzyl	<i>t</i> -butyl
Side-chains deprotection	Catalytic hydrogenation	TFA
Cleavage	HF	TFA

Table 4.1: Main strategies used in SPPS. Reagents used in Fmoc chemistry are safer and less aggressive than those used in BOC chemistry.

In Boc chemistry, repeated exposure to TFA to deprotect the backbone amino group is a potential cause of peptide bond and side chain alterations. In addition, the use of HF for the final resin cleavage step is chemically aggressive and dangerous, and it requires special facilities. These are the main reasons why the Fmoc chemistry was developed and is nowadays the most widely used strategy. In Fmoc chemistry, an orthogonal protecting strategy is applied relying on the use of a base for backbone amino deprotection (piperidine in DMF). This enables the use of side-chain protecting groups labile to milder acidic conditions (*t*-butyl, benzyl...) that are removed by TFA during the final resin cleavage step.

In the particular case of phosphopeptide synthesis, recurrent acidolytic conditions used in Boc chemistry specially compromise the stability of phosphate groups. This combined to a higher accessibility for biomedicine-focused laboratories are the reasons for the widespread use of Fmoc chemistry to synthesize phosphopeptides, even though they can be synthesized using Boc chemistry as well.

4.2.2. Reagents in SPPS using Fmoc chemistry

The **resin** is the solid support where the amino acid chain is grown. It is made of a cross-linked polymer. Its physical properties have a great impact in the global synthesis yield. These include:

- *Chemical stability*, resistance to all chemicals used along the synthesis.
- *Chemical inertia*, non-reactivity under the synthesis conditions.
- *Swelling*, capacity to be inflated by the action of organic solvents.
- *Functionalization*, reactive groups (mmol) per gram of resin.

The two latest properties are of special relevance given that the polypeptide chain grows not only on the surface of the resin beads but also inside them. High swelling combined with low functionalization means higher reagent accessibility and solvent drainage.

Polystyrene-based resins have long been the primary choice. The *Rink Amide* resin, designed to introduce a C-terminal amide in the cleaved peptide, is an example of such resin that has been used in some syntheses presented in this work. When synthesizing long sequences, the hydrophobicity of the material may have a negative effect by favouring aggregation. Polyethylene glycol (PEG)-based resins are an alternative in such cases as they present increased polarity and swell extensively in solvents like DMF and acetonitrile. The *ChemMatrix* is a PEG-based resin (García-Martín et al., 2006) that [gives](#) good yields with complicated sequences and has been used to synthesize most of the phosphopeptides presented in this work.

Common organic **solvents** are used in SPPS. DMF is used to dissolve the coupling and amino deprotecting reagents given its low volatility, whereas DCM serves to swell the resin. DMF traces cause false positive results in the Kaiser test so extensive DCM rinsing is also useful in this sense.

Coupling activating agents are used to favour the formation of the peptide bond. They increase the electrophilic character of the free carboxyl group in the amino acid to be coupled. In this manner the nucleophilic attack by the free amino group in the resin-bound amino acid is enhanced. The use of triazoles such as PyBOP, HOAt, HOBt and their derivatives HATU and [HBTU](#) is of advantage over carbodiimides (DCC, DIC) as it avoids amino acid racemization. A typical coupling mixture contains a combination of two different activating agents. To provide the slightly basic conditions required for their activity they are used in combination with DIPEA, a tertiary amine that acts as a base while being a bad nucleophile.

The most common **Fmoc deprotecting agent** is piperidine, a nucleophilic base that displaces and covalently binds the Fmoc moiety. The resulting compound can then be rinsed away. Undesired side-reactions (phosphate β -elimination, aspartimide formation...) may require the use of alternative bases such as DBU or piperazine.

Once the amino acid chain is complete, TFA is used for **resin cleavage**. Benzyl and *t*-butyl based side-chain protecting groups are labile in these conditions and therefore are removed in this step.

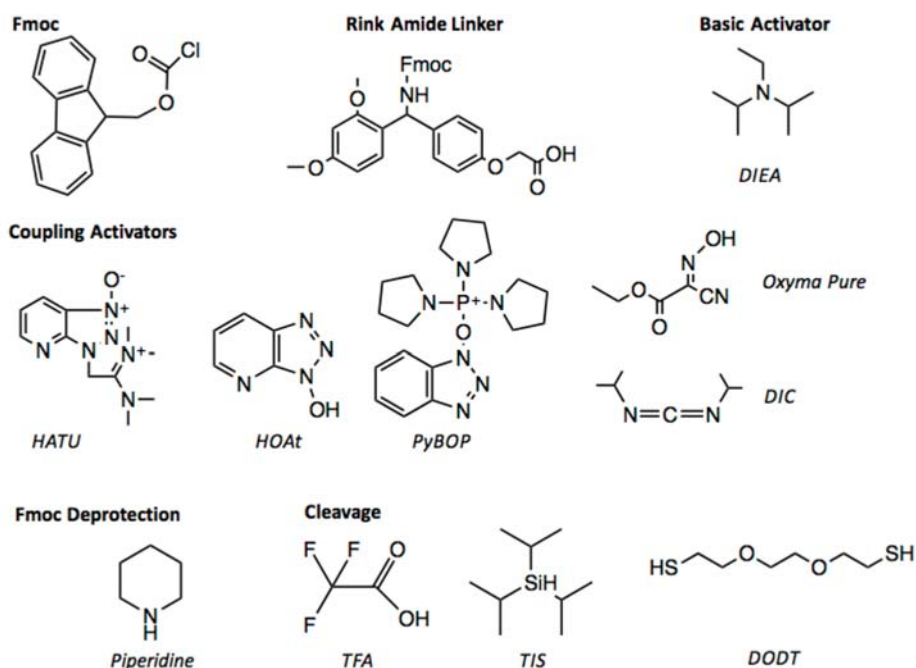


Figure 4.4: Reagents frequently used in Fmoc SPPS. The chemical formula of some of the reagents most frequently used in Fmoc chemistry is shown.

Highly reactive cationic species are formed during the cleavage that can react with side-chain nucleophile functional groups such as those in Cys, Met, Tyr and Trp. Combinations of **scavengers** such as water, TIS or other thiols (EDT, DODT, 2-PySH) are added to the cleavage mixture at 1-2.5% each to bind these species and avoid their negative effects.

When the cleavage is complete, most of the TFA is evaporated and the peptide is **precipitated** using cold ether, typically diethyl or tertmethylbutyl ether. The crude can then be dissolved in a water/acetonitrile mixture and purified by reversed-phase chromatography.

4.2.3. Coupling/deprotection checking tests

The success of coupling and deprotection steps can be easily tested using the **Kaiser test** (Kaiser et al., 1970). This is based on the colorimetric reaction of ninhydrin with primary amines. A few resin beads are transferred to a test tube where ninhydrin is added. The tube is heated for a short time (100°C, 3 min). The supernatant turns blue in case there are free amines attached to the resin (positive result).

The Kaiser test is far less sensible to secondary amines as those present in the Pro backbone. If unprotected, reaction with ninhydrin gives a brownish coloration of the beads. In these cases the use of the **chloranil test** (Vojkovsky, 1995) is more reliable,

which is based on the reaction of chloranil with such amines (5 min, RT). Resin beads turn green in case they contain free secondary amines.

Some automated synthesis protocols use a different approach. Instead of detecting free amines, deprotection is checked taking advantage of the **UV absorbance** of the Fmoc group released to the solution.

All methods described above are subject to possible failure thus yielding false positive (or negative) results. This may happen due to the presence of interfering reagents (such as DMF in the Kaiser test) or hindered accessibility of the amino groups to be detected. Ambiguous results in the colorimetric tests are common after the coupling of a phosphorylated amino acid. In such cases performing a **mini-cleavage** is a more reliable manner to check the coupling. It consists in cleaving the peptide from a small fraction of the resin used in the synthesis followed by its characterization by MS. Its main drawback is the much longer time that it requires when compared to the other methods, usually about 3 hours.

4.2.4. Manual and automated synthesis strategies

Given its repetitive and cyclic nature, peptide synthesis can be highly automated. **Manual synthesis** has the advantage of providing a high control along the synthesis, with the possibility to check and decide on every step. It is also the most economic method in terms of reagents. However, it is a demanding and time-consuming procedure. Coupling reactions at room temperature require from 60 to 90 minutes, piperidine deprotection lasts for around 20 minutes, and extensive washing with both DMF and DCM is required in between each step. This makes an approximate time of 2 hours per cycle. The synthesis of a complete sequence as the ones presented below (35 residues) takes several days.

Automated room temperature synthesis makes the procedure much quicker and less tedious. However, coupling and deprotection reactions are not speeded up. In addition, decision-making after every step is no longer possible.

Automated microwave synthesis sensibly reduces the overall synthesis time by heating the mixture using microwaves during the coupling and the deprotection steps. In these conditions, the synthesis of a 35 amino acid sequence takes less than a day. Again, control over each individual step is lost, but good yields and amenable crudes are often obtained. The strategy is suitable for phosphopeptide synthesis. Nonetheless, certain undesired side-reactions such as phosphate β -elimination may be favoured by heating.

4.2.5. Coupling efficiency and intrinsic by-products

A major complication in SPPS arises from the fact that every coupling step has a limited yield, typically around 99%. After every coupling step, stochastically a small but still sensible fraction of the growing peptide chains will remain uncoupled, with an unprotected amine. Similarly, the efficiency of backbone amide deprotection steps is limited as well.

Considering only a limited efficiency in the coupling steps, for a relatively short peptide (15 amino acids), the yield of desired product will be around 86%. In the case of the longest peptides presented below (35 amino acids) this will be reduced to around 70%.

Consequently, by-products are intrinsically generated in every synthesis that are very similar to the desired product in terms of physicochemical properties. In order to obtain it with the purity required for further experiments, a case-by-case optimized reversed-phase chromatography step is thus needed after almost every synthesis.

This situation may be aggravated when complicated sequences (in terms of length, impeded couplings, use of modified residues...) are synthesised, as it is the case for some of the phosphopeptides presented below. Factors such as steric impedance may lower the yield of a specific coupling step down to 50-60%. When this happens once in a synthesis, it gives raise to two major products in the crude. Purification of the desired product is in these cases a real challenge given the high similarity and abundance of the contaminant.

4.2.6. Chemical ligation

The complications presented in the previous section limit the potential of SPPS to the synthesis of relatively short sequences of up to 50 residues, enough to cover only a very small fraction of the functional protein domains existing in nature. This has been extensively overcome with the development of chemical ligation [techniques](#). Using them entire proteins have been synthesized so far of great value in protein and medicinal chemistry (Kent et al., 2012): fully active human insulin, active enzymes, D-proteins to be used as antagonists of the naturally occurring L-proteins (as in the case of VEGF), easily crystallizing racemic mixtures of otherwise recalcitrant proteins, etc.

Chemical ligation is a convergent strategy where two or more peptide fragments are independently synthesized and covalently condensed in a single molecule. Fully deprotected peptides are used, as the ligation reaction is chemo and regioselective.

A variety of strategies have been developed. Disulphide-based chemical ligation has been successfully used in our lab to achieve segmental labelling of recombinant proteins (Aragón et al., 2011). The growing length and complexity of the synthetic phosphopeptide ligands required for our protein-protein interaction studies has led to the exploration of native chemical ligation (NCL) in this sort of molecules.

In NCL a native peptide bond is created after the ligation of the peptide moieties. The N-terminal segment is synthesized as a peptide thioester in its C-terminus. A N-terminal cysteine residue (or otherwise a modified amino acid with a thiol group) is required in the C-terminal segment. A transesterification chemoselective reaction occurs between the side-chain free thiol and the C-terminal thioester. An intermediate is established that spontaneously rearranges through a S->N acyl shift to yield a native peptide bond. [Furthermore](#), cysteine can be converted into alanine by desulfurization, which can be applied to thiol variants of natural amino acids as well.

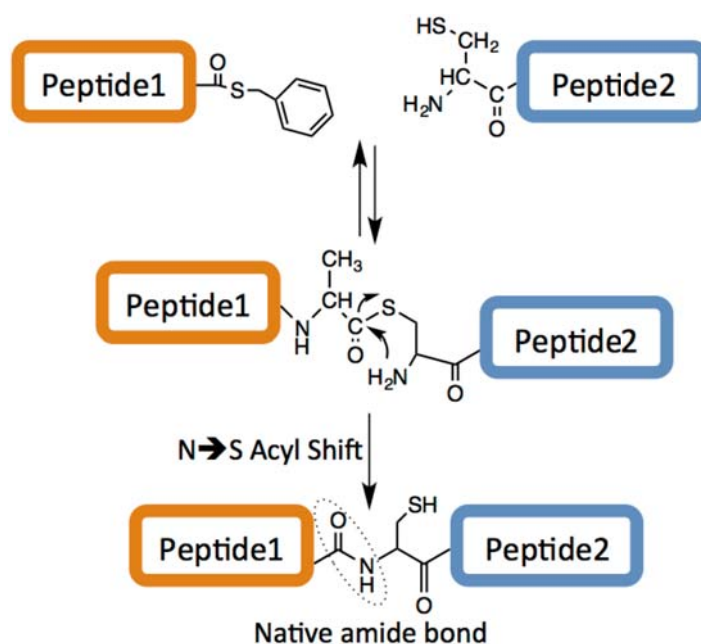


Figure 4.5: Basics of native chemical ligation. A C-terminal peptide thioester (Peptide1) is fused to a peptide with an N-terminal cysteine residue (Peptide2). A transesterification reaction occurs between the cysteine side-chain thiol group and the C-terminal thioester, and the intermediate rearranges spontaneously to yield a native amide bond.

In the last section of this chapter the synthesis of a peptide thioester and its ligation to a Smad3 phosphopeptide are presented, as well as a cysteine

desulfurization assay performed on a phosphopeptide. This settled the bases for the application of this technique for our particular research interests.

Recently, optimal conditions have been found for cysteine-free peptide ligation (Payne et al., 2008). Our research group is currently working in the optimization of the protocols to apply this technique to the synthesis of the phosphopeptides required for our studies.

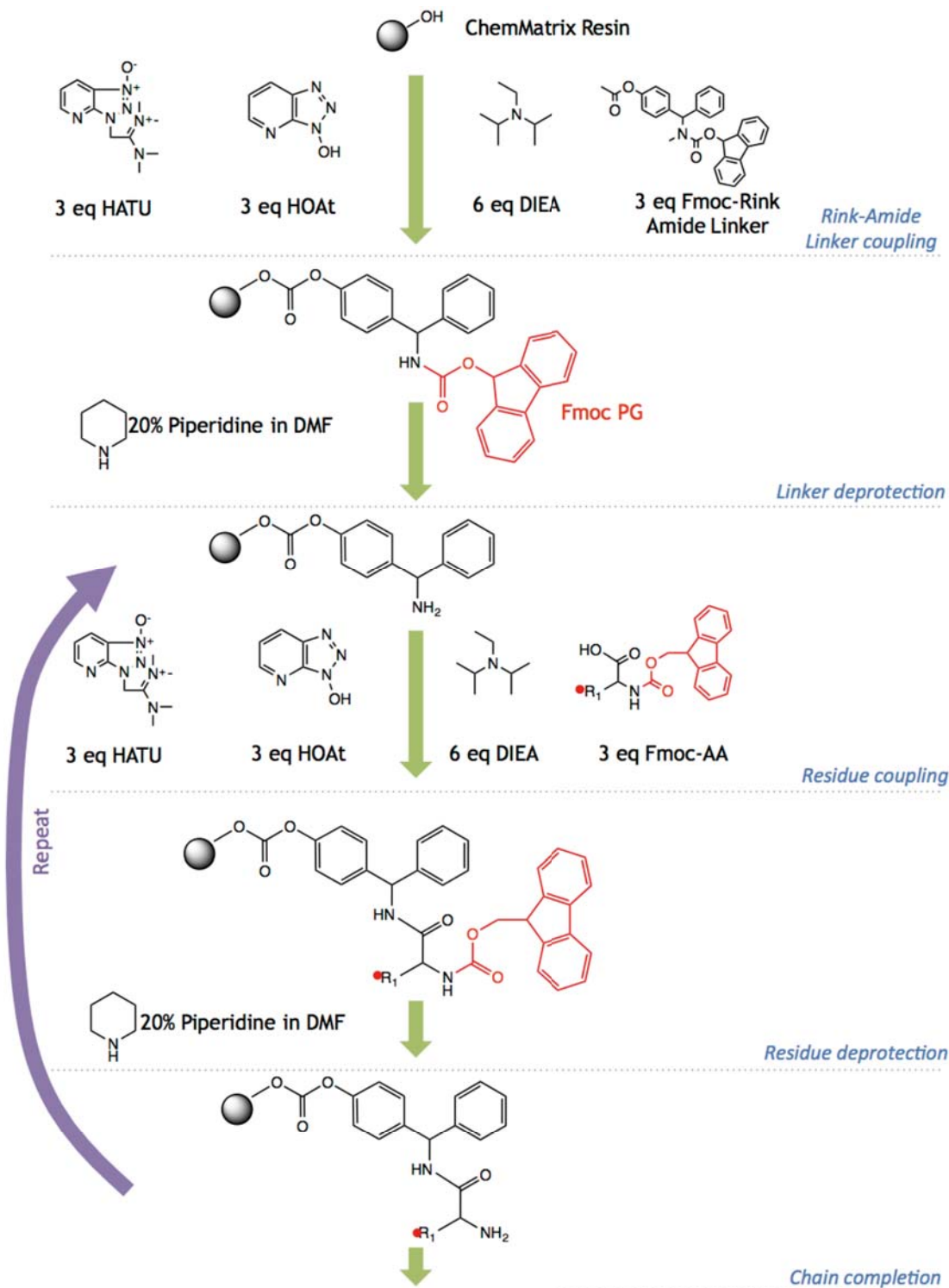
4.3. Standard procedures used in the reported syntheses

In general, similar conditions were used to synthesize all the peptides reported below (summarized in **Figure 4.6** and **Table 4.2**). Variants were applied when necessary to overcome particular complications in some of the syntheses.

Strategy	Fmoc/ ^t Bu
Resin	Aminomethyl ChemMatrix®
Resin loading	0.74 mmol/g
Scale	0.2 – 0.25 mmol
C-terminal amidation	Rink-Amide linker
Amino acid equivalents	3 (manual and RT automatic) or 4 (MW automatic)
Coupling activator mixture	3 eq HATU + 3 eq HOAt + 6 eq DIEA (manual) 4 eq HATU + 4 eq HOAt + 8 eq DIEA (MW automatic)
N-terminal acetylation	Acetic anhydride:DIEA:DMF at 5:8.5:86.5
Cleavage mixture	TFA:H ₂ O:TIS at 95:2.5:2.5
Preparative RP-HPLC	<p><i>Stationary phase:</i></p> <ul style="list-style-type: none"> • Waters SunFire™ PREP C18 OBD™ <ul style="list-style-type: none"> ○ Stationary phase: C₁₈ ○ Column size: 19x100 mm ○ Particle size: 5 μm ○ Pore size: 100 Å <p><i>Mobile phase:</i></p> <ul style="list-style-type: none"> • ACN gradient in H₂O with 0.1% TFA <ul style="list-style-type: none"> ○ 0-20% ACN in 5 min + 20-30% ACN in 20 min ○ 10 mL/min flow rate

Table 4.2: Conditions used in the syntheses detailed in this chapter. Standard conditions unless stated otherwise.

In all the cases, the free carboxyl group of the C-terminal amino acid was amidated. This was achieved by coupling a Rink-Amide linker to the functionalized resin before starting the synthesis. Once the amino acid sequence was completed, the free amino group of the N-terminal residue was acetylated using acetic anhydride.



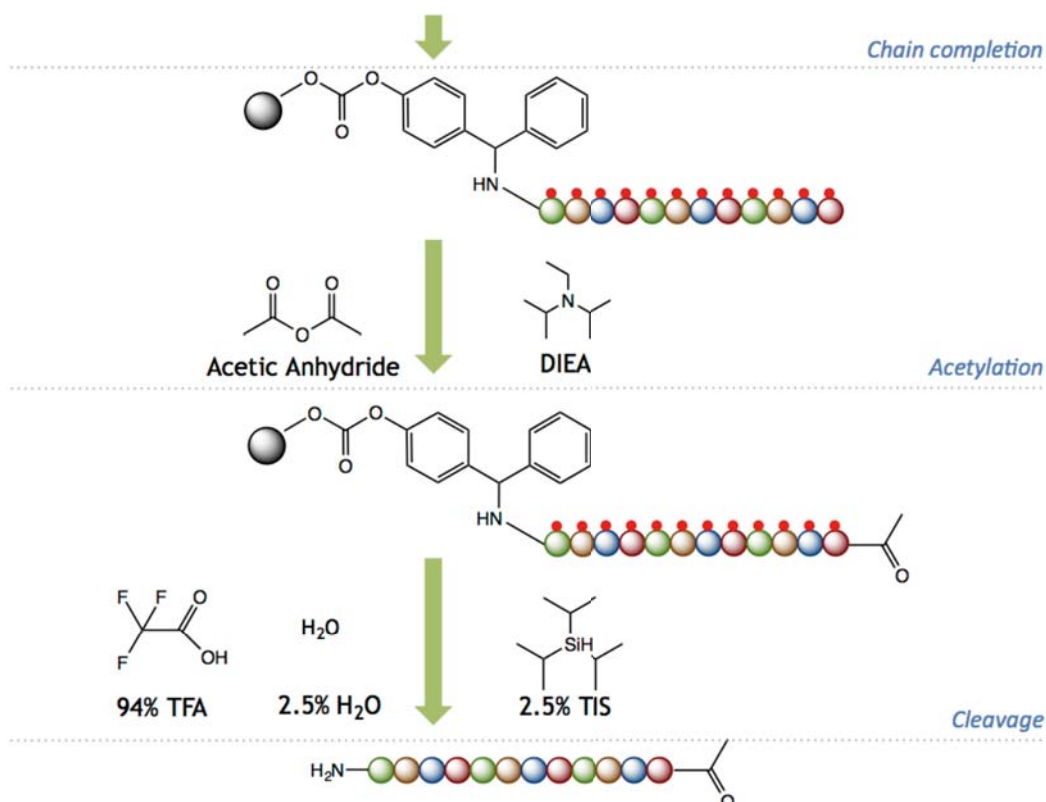


Figure 4.6: Standard pipeline used in the reported syntheses. The reagents and the equivalents shown in this figure were used at each synthesis step unless otherwise stated.

The purified fraction of all the synthesized peptides was characterized using LC-MS and, when necessary, MALDI-TOF. The chromatograms and the MS reports are detailed in Section 9.1.

4.4. Synthesis of a set of Smad1 phosphopeptides

In the nucleus, CDK8/9 phosphorylates the linker region of Smad1 at residues S206 and S214, located downstream with regards to the PY motif (positions 223 – 227). The modification of these positions primes for GSK3- β phosphorylation of the T202 and S210 sites respectively (n-4).

In order to perform comparative ITC and NMR binding studies with YAP and Smurf1, we synthesized six variants of the Smad1 sequence each representing a different phosphorylation pattern.

Smad1 NP corresponds to the unphosphorylated sequence. Due to its amenable size (34 residues), recombinant expression was used instead of SPPS in this case. Together with SPPS synthesis of Smad1 pS206, this was performed by Dr Eric Aragón and Dr Nina Görner in our laboratory, and is therefore not detailed below.

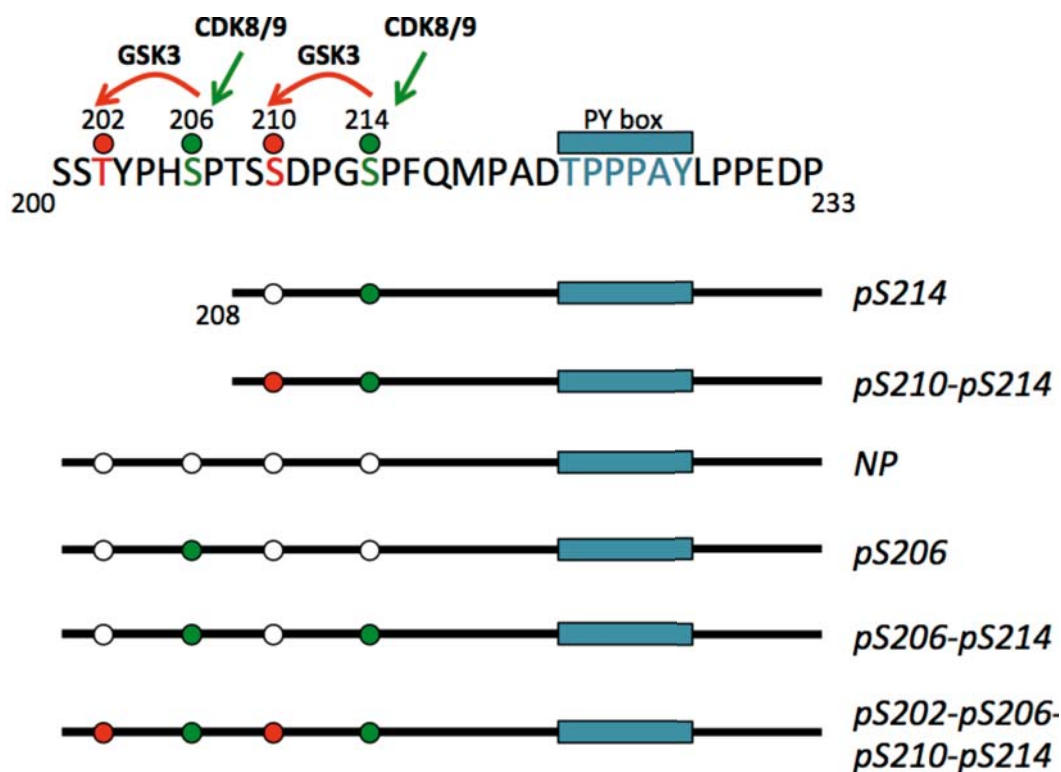


Figure 4.7: Set of Smad1 phosphopeptides synthesized for this study.

Smad1 pS214 and Smad1 pS210-pS214 cover a shorter segment of the Smad1 linker sequence (26 residues) compared to all other variants (34 residues). This was enough to represent the phosphorylation variants of interest in these cases, and facilitated both their synthesis and the interpretation of the NMR experiments they were used in.

4.4.1. Synthesis of Smad1 pS214

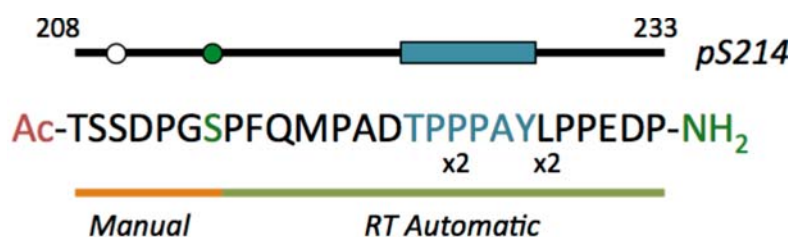


Figure 4.8: Synthesis of Smad1 pS214.

Smad1 pS214 was synthesized using a combined strategy (Figure 4.8). The C-terminal segment from P213 to P233 (19 residues) was synthesized by Fmoc/^tBu-chemistry on a room-temperature synthesizer using 10 equivalents of the amino acid Fmoc derivatives activated with a HATU + HOAt mixture in DMF (0.45 M each) and DIEA (2 M) in DMF at the 0.25 mmol scale (Table 4.2). P224 and L228 were

coupled twice (double coupling steps). The phosphorylated residue (pS214) was incorporated manually by activating the corresponding amino acid Fmoc derivative (5 equivalents) again with HATU, HOAt and DIEA in DMF (same equivalents).

The rest of the synthesis was continued manually. After the addition of the phosphoserine to the amino acid chain results obtained from the Kaiser and the chloranil tests turned ambiguous. Instead, most of the couplings performed manually (5 out of 7) were checked using the mini-cleavage protocol.

The peptide was acetylated, cleaved and purified by preparative RP-HPLC using the conditions specified in **Table 4.2**. The peptide was separated from its oxidized and Met-truncated versions detected in the crude. In the pure fraction, it was detected by LC-MS as the doubly and triply ionized species (m/z calculated: 1416.97 and 944.98; m/z observed: 1471.78 and 945.23).

4.4.2. Synthesis of Smad1 pS210-pS214

The Smad1 pS210-pS214 peptide was completely synthesized using the automated microwave-assisted strategy (**Figure 4.9**). The synthesis was carried out by Fmoc/^tBu-chemistry on a CEM Liberty1 microwave synthesizer using 5 equivalents of the amino acid Fmoc derivatives activated with 2 mL of a HATU + HOAT solution (0.5 M each) and 1 mL of 2 M DIEA in DMF at the 0.25 mmol scale (**Table 4.2**).

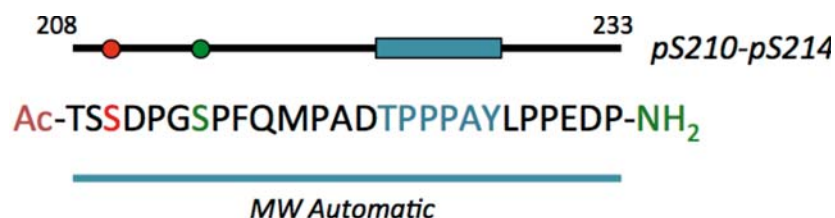


Figure 4.9: Synthesis of Smad1 pS210-pS214.

Optimal microwave reactor conditions were used for the coupling of the phosphorylated and the immediately consecutive residues (Harris et al., 2008). These slightly differ from the conditions used for all other amino acids:

Standard conditions	Phosphorylated and consecutive residues
5 minutes per coupling	15 minutes per coupling
25W at T _{Max} 75°C	25W at T _{Max} 72°C

Table 4.3: MW coupling conditions for phosphopeptides.

After the incorporation of a phosphorylated residue, microwave heating during the Fmoc deprotection step may result in an undesired side-reaction: the β -elimination of a phosphate group (Attard et al., 2009).

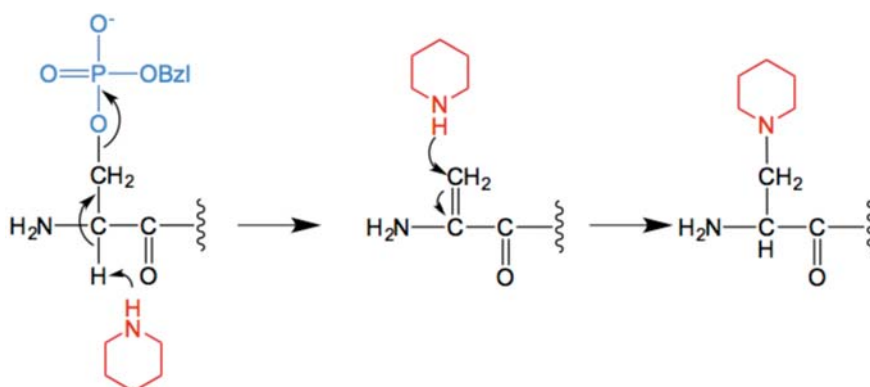


Figure 4.10: β -elimination of a phosphate group by piperidine. The phosphate group is highlighted in blue, whereas piperidine is shown in red.

The nitrogen atom in the piperidine ring is able to perform a nucleophilic attack on the α carbon of the phosphorylated residue. This results in the displacement of the phosphate moiety and the incorporation of piperidine to the side-chain, yielding a by-product with $m/z = M-13$. In order to prevent this, the synthesizer was paused after the addition of each phosphorylated residue. These were deprotected manually at room temperature using DBU instead of piperidine (**Table 4.4**).

When a consecutive amino acid is coupled to the phosphorylated residue, the efficiency of the nucleophilic attack by piperidine becomes negligible so it can be used again for subsequent deprotection steps.

Standard conditions	Phosphorylated residues
20% piperidine in DMF	5% DBU in DMF
30 sec + 3 min	5 min + 20 min + 20 min
60W at T_{Max} 75°C	RT, manual

Table 4.4: Deprotection conditions in MW-assisted phosphopeptide synthesis.

LC-MS analysis of the resulting crude revealed the presence of several truncation species. Notwithstanding that, the target peptide was resolved from most of the impurities in the major peak of the UV profile when using a preparative column (RP-HPLC conditions specified in **Table 4.2**). The peptide was obtained with 99.3% purity and detected as the singly ionized specie by MALDI-TOF (m/z calculated: 2898.89; m/z observed: 2894.80) and the doubly and triply ionized species by ESI-Q-TOF (m/z calculated: 1449.95 and 966.96; m/z observed: 1450.63 and 967.11).

4.4.3. Synthesis of Smad1 pS206-pS214

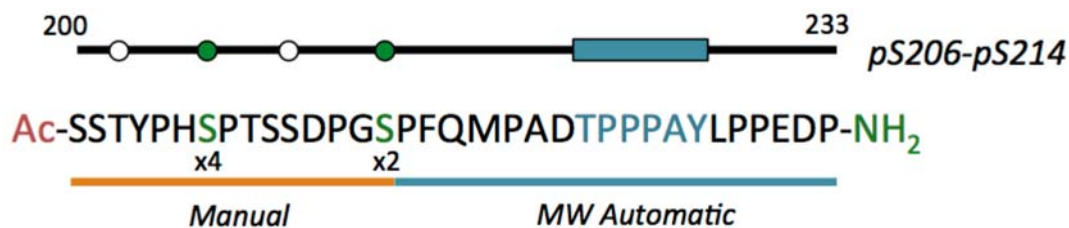


Figure 4.11: Synthesis of Smad1 pS206-pS214.

Similar to the case of Smad1 pS214, the synthesis of Smad1 pS206-pS214 was carried out using a combined strategy (**Figure 4.11**). In this case, microwave-assisted automated synthesis was used to couple the amino acids corresponding to the C-terminal segment from P213 to P233 (19 residues). The phosphorylated segment was added manually.

Synthesis of the C-terminal segment in the microwave automatic synthesizer was performed as reported in Section 4.4.2 and **Table 4.2**. Its performance was checked using the mini-cleavage strategy. The crude was analysed by LC-MS and the expected m/z (calculated 2080.38) was detected as the [singly and](#) [doubly charged](#) ion (m/z [experimental](#) = 2080.35 and 1040.64) in the major peak of the UV chromatogram ($\approx 60\%$ of the crude material).

During the synthesis of the N-terminal segment, the coupling of both phosphoserine residues required mini-cleavage checking due to ambiguous results in the colorimetric tests. Both residues were re-coupled (twice for pS214 and four times for pS206). PyBOP (3 eq) or Oxima Pure + DIC (3eq each) were used in the activator mixture as a functional alternative to the HATU + HOAt (3 eq each) combination used in all other coupling steps, which improved the coupling efficiency. Before deprotecting the backbone amino group, uncoupled growing chains were acetylated to avoid the incorporation of new amino acids.

When the peptide chain was complete, a final mini-cleavage step revealed the presence of three major contaminants in the crude mixture that co-eluted with the target product in the LC-MS chromatographic analysis:

- *M-167*, corresponding to the target peptide lacking a phosphoserine residue.
- *M+16*, typically observed when a Met residue is oxidized during the cleavage step.

- *M+90*, arising due to the alkylation of a Met residue with the phosphate protecting group (Bzl) (Harris et al., 2008).

The *M+16* and *M+90* by-products were generated during the TFA cleavage step. Apart from decreasing the synthesis yield, the purity of the target peptide was compromised as their similarity made them very hard to resolve by RP-HPLC. Therefore, the cleavage conditions were optimized to avoid the formation of these species.

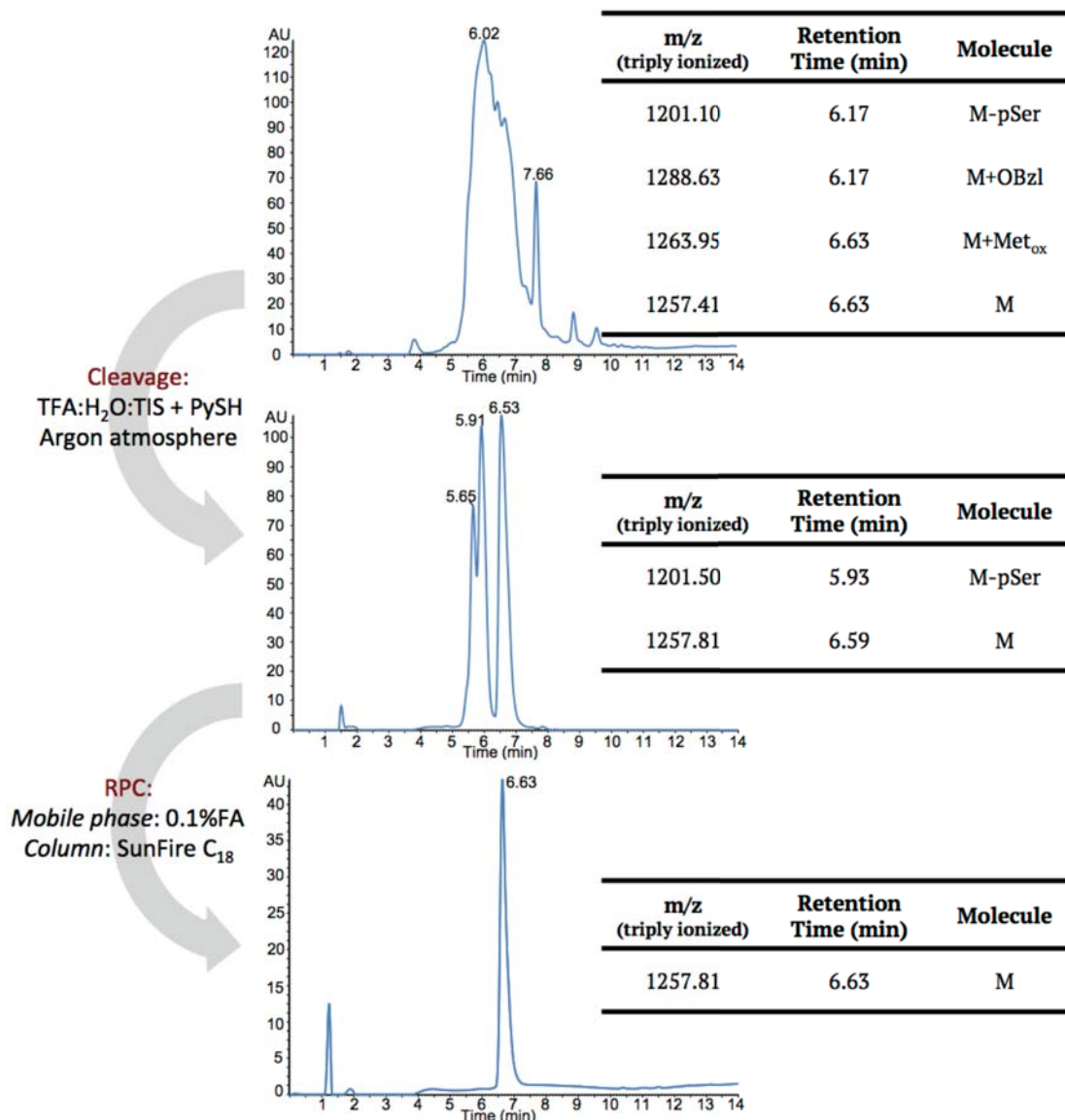


Figure 4.12: Cleavage and purification steps optimization for Smad1 pS206-pS214. On the top panel, the LC-MS analysis of the crude obtained using cleavage conditions detailed in **Table 4.2** is shown (chromatogram and detected ions). The middle panel shows the LC-MS analysis of the crude after applying the cleavage-optimized conditions. The lower panel displays the LC-MS analysis of the pure fraction after optimizing the purification step. Analyses performed using a SunFire™ C₁₈ 4.6x100 3.5µm column, gradient 0-100% ACN with 0.1% FA over 14min.

2-PySH had been described to be a good scavenger for Bzl reactive ions in BOC chemistry (Taichi et al., 2009). An efficiency of >90% was reported for HF based cleavages. No application in Fmoc chemistry had been published so far. Optimal conditions were found using a cleavage mixture with TFA, H₂O and TIS (95:2.5:2.5) together with 5 equivalents of 2-PySH. In addition, the reaction was performed in argon atmosphere to avoid Met oxidation. After the cleavage, the crude was treated with a 0.1% TFA + 5 eq 2-PySH solution (RT, 2 hours) to remove residual alkylation of Met. The crude was then lyophilized in presence of 2-PySH.

With these cleavage conditions, the M+16 and M+90 by-products were not detected any more (**Figure 4.12**). It was therefore applied to the whole synthesis after the acetylation of the N-terminus.

The truncation M-167 by-product lacking a phosphoserine co-eluted with the target peptide when using a Waters SunFire™ PREP C18 OBD™ column combined with 0.1% TFA in the mobile phase (Section 4.3). It was finally removed using formic acid (FA, 0.1%) instead of TFA (**Figure 4.12**). The Smad1 pS206-pS214 peptide was obtained in isolation with a purity of 98.9% and detected by MALDI-TOF (singly ionized, *m/z* calculated: 3769.80; *m/z* detected: 3769.70) and ESI-Q-TOF (doubly and triply ionized, *m/z* calculated: 1885.40 and 1257.27; *m/z* observed: 1886.20 and 1257.81).

4.4.4. Synthesis of Smad1 pT202-pS206-pS210-pS214

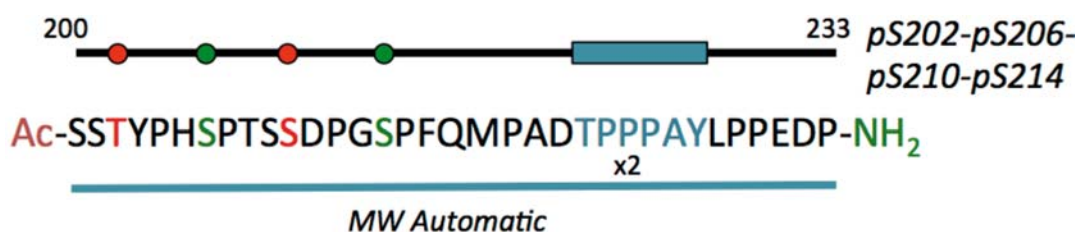


Figure 4.13: Synthesis of Smad1 pT202-pS206-pS210-pS214.

Differently from all previous syntheses, the Rink-Amide resin was used for the tetra-phosphorylated peptide Smad1 pT202-pS206-pS210-pS214. In this case all the amino acids were added using a microwave-assisted automatic synthesizer (**Figure 4.13**, conditions specified in **Table 4.2**). A double coupling step was set for P224.

The coupling and deprotection of the four phosphorylated residues were performed with the optimal conditions reported for the case of Smad1 pS210-pS214 (Section 4.4.2).

Once acetylated, the peptide was cleaved from the resin using a mixture of TFA:DODT:H₂O:TIS (94:2.5:2.5:1). DODT is a thiol with proven scavenging efficiency in Fmoc chemistry that is widely used in substitution of EDT due to its neutral odour. The target peptide was detected in the crude by ESI-based LC-MS as the triple ionized specie (m/z calculated: 1339.99; observed: 1339.55).

First, a purification step using the conditions mentioned in **Table 4.2** resulted in the isolation of the target peptide together with several truncated versions, many of them corresponding to the peptide lacking one or more phosphorylated residues. A second step was required to completely isolate the target peptide. Again, FA (0.1%) was used in the mobile phase instead of TFA. None of the by-products were detected in the purified fraction, and the peptide was obtained with a final purity of 97.5%.

4.4.5. Biophysical bases for the Smad1 action turnover switch

The peptides represented in **Figure 4.7** were used to elucidate the biophysical bases explaining the switch in Smad1's preferred binding partners. The results reported in this sub-section were presented as part of Dr Eric Aragón's thesis.

The binding affinities of both YAP and Smurf1 to the phosphopeptides were measured using ITC (**Figure 4.14A**). The recombinant YAP WW1-WW2 segment binds Smad1 with the highest affinity when both CDK8/9 positions (S206 and S214) are phosphorylated ($8.4 \pm 1 \mu\text{M}$). In contrast, the affinity of the recombinant Smurf1 WW1-WW2 segment for this interaction is lower ($18.8 \pm 6 \mu\text{M}$). Phosphorylation of the GSK3- β positions drastically decreases the affinity of the interaction with YAP ($60.6 \pm 7 \mu\text{M}$), whereas Smurf1 binding is favoured ($10.4 \pm 1 \mu\text{M}$). The highest affinity for the Smurf1-Smad1 interaction was measured when titrating the Smad1 pS210-pS214 peptide ($1.2 \pm 0.3 \mu\text{M}$), which is a shorter version of the tetra-phosphorylated sequence. This revealed a preference of Smurf1 for the pS214 position over pS206 that was confirmed *in vivo* by our collaborators using HEK293 cells.

ITC measurements suggest a scenario where CDK8/9 phosphorylation creates a docking site in Smad1 for the competitive binding of YAP and Smurf1. After GSK3- β phosphorylation, the binding is unbalanced in favour of Smurf1. To investigate this switch with atomic resolution, multidimensional NMR was used to solve the structures of YAP and Smurf1 in complex with different Smad1 phosphopeptides. For clarity, peptide residues are named with the one character code whereas protein amino acids are named with the three characters code from now on.

Figure 4.14C shows the structure of the YAP WW1-WW2 segment binding Smad1 pS206-pS214. WW1 coordinates pS206 in the proximities of the first loop, whereas the WW2 domain binds the PY box in a canonical manner.

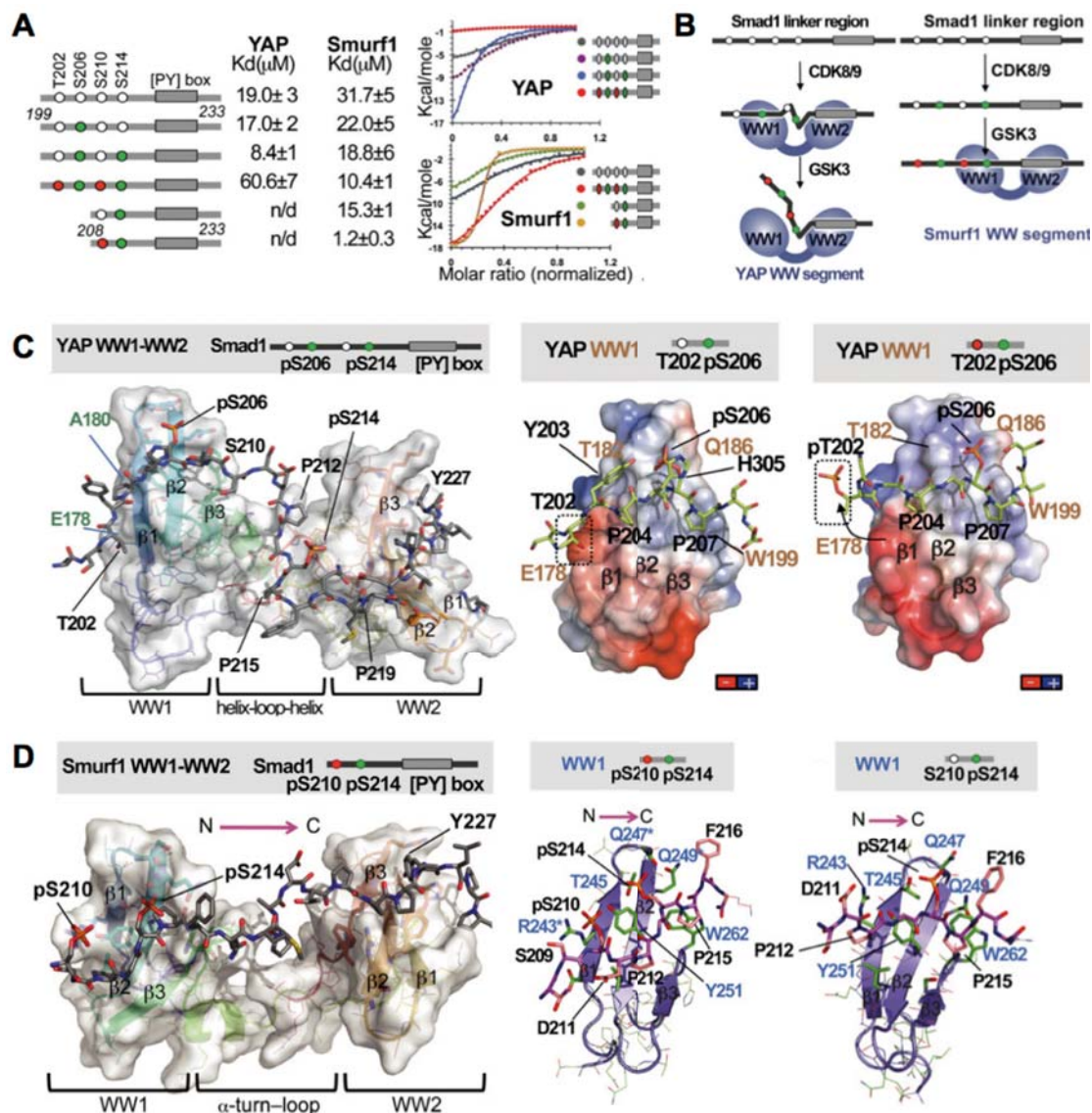


Figure 4.14: Biophysical and structural basis for the Smad1 action/turnover switch. (A) ITC measurements for the interaction of YAP and Smurf1 tandems of WW domains with peptides representing different phosphorylation patterns of Smad1. (B) Models for YAP and Smurf binding to Smad1 depending on its phosphorylation state. (B) Structural studies on YAP binding the phospho-S/TP cluster on Smad1. (C) Structural studies on Smurf1 binding the phospho-S/TP cluster on Smad1.

The surface charge distribution of the YAP WW1 domain bound to this peptide is shown in **Figure 4.14C** (middle panel). T202 is accommodated close to Glu178 in a negatively charged patch of the domain. Upon phosphorylation of T202 by GSK3-β this interaction is hampered by electrostatic repulsion, explaining why this event decreases the YAP binding affinity.

For the YAP WW1 domain to bind the distal pS206 position, the Smad1 pS206-pS214 peptide adopts a conformation where both D211-P212 and pS214-P215 bonds are in the *trans* configuration (**Figure 4.14C**, left). This results in the formation of two β -turns that favour the interaction of pS206 with YAP WW1. This could explain the lower affinity for Smad1 pS206, as P215 is not locked in the *trans* configuration in this case. Phosphorylation of S210 by GSK3- β results in electrostatic repulsion with D211, thus affecting the β -turn centred in this position and locking P212 in the *cis* configuration. This contributes to the low affinity of YAP for Smad1 pT202-pS206-pS210-pS214 (**Figure 4.14B**, left).

Differently, the Smurf1 WW1-WW2 pair prefers the extended configuration that Smad1 adopts after GSK3- β phosphorylation as it binds the proximal pS210 and pS214 positions (**Figure 4.14B** and **D**). In the structure, WW1 binds the phospho-S/TP cluster and WW2 coordinates the PY motif. A close-up view of the phospho-sites binding by Smurf1 WW1 (**Figure 4.14D**, middle) shows that pS214 is coordinated by the residues around the first loop Thr245, Gln247 and Gln249. pS210 contacts Arg243 and Tyr251 in β 1 and β 2 respectively.

For comparison, the structure of Smurf1 WW1 in complex with Smad1 pS214 was solved too (**Figure 4.14D**, right). In this case, the protein-peptide contacts are much less abundant than in the presence of a phosphate group in position S210, explaining the lower affinity of Smurf1 for Smad1 pS214 and Smad1 pS2016-pS214.

4.5. Synthesis of a set of Smad3 phosphopeptides

In Smad3 the PY motif (PPPGY) occupies the positions 180 to 184. It is preceded by T179, which is phosphorylated by CDK8/9. Together, the sequence is known as the pT[PY] motif. Two more positions are phosphorylated by this kinase, namely Ser208 and Ser213. Different from the case of Smad1, these are located upstream with regards to the PY box. Phosphorylation of position Ser208 primes GSK3- β to act on Ser204.

Alanine mutation of the CDK site Ser208 diminishes Nedd4L binding to Smad3 in human HEK293 cells, as well as mutation of the GSK3- β position Ser204 (Aragón et al., 2011). In contrast, mutation of Ser213 to alanine has no effect on this binding.

A set of four peptides representing different phosphorylation variants of Smad3 was synthesized to characterize its interaction with Pin1 and Nedd4L (**Figure 4.15**). Dr Nina Görner in our laboratory prepared the Smad3 pT179-pS208 peptide and its synthesis is not detailed below.

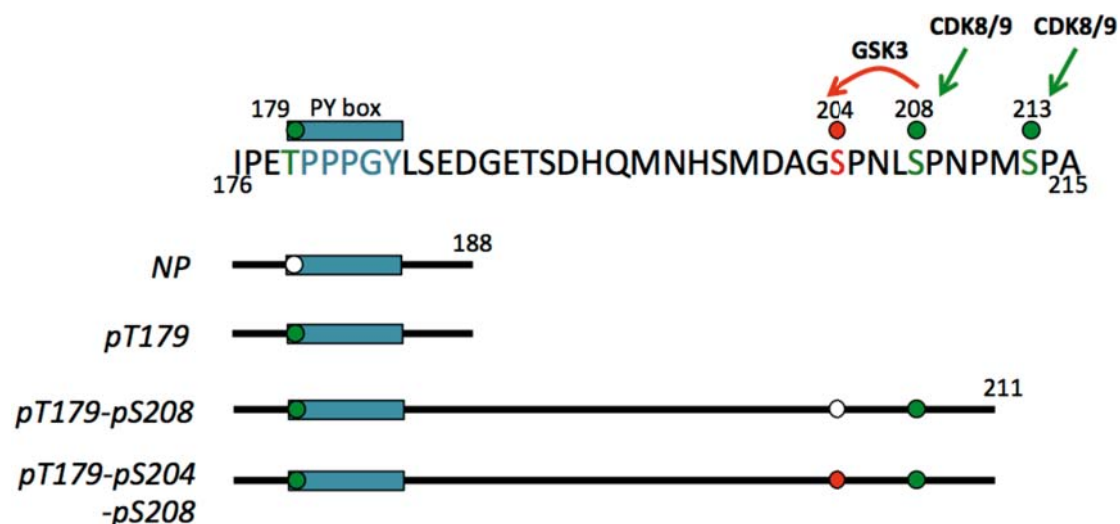


Figure 4.15: Set of Smad3 phosphopeptides synthesized for this study.

4.5.1. Syntheses of Smad3 NP and Smad3 pT179

Smad3 NP and pT179 syntheses were performed using the microwave-assisted automatic strategy (Figure 4.16) with the conditions detailed in Table 4.2.

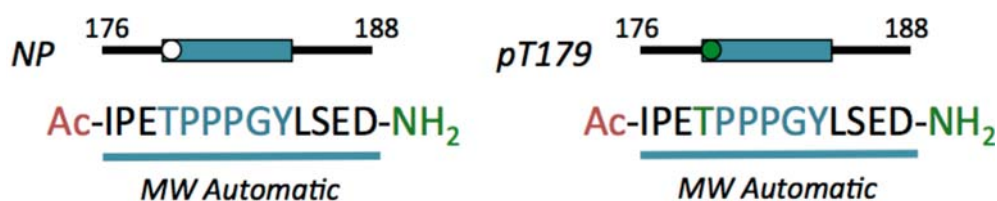


Figure 4.16: Syntheses of Smad3 NP and Smad3 pT179.

Optimal microwave-assisted conditions for phospho-amino acid coupling (Table 4.3) and deprotection (Table 4.4) were used for the incorporation of pT179. After completion of the syntheses, the peptides were acetylated and cleaved using the TFA:DODT:H₂O:TIS (94:2.5:2.5:1) solution. Singly ionized species corresponding to both target peptides were detected in the major peak of their respective crudes (Smad3 NP m/z calculated: 1455.57; m/z observed: 1455.88; Smad3 pT179 m/z calculated: 1536.58; m/z observed: 1536.70). After RP-HPLC, Smad3 NP was obtained with a final purity of 97.7% whereas Smad3 pT179 was 90.3% pure.

During the characterization of the pure fraction of Smad3 NP by MS (using both ESI and MALDI ionization strategies) several ions were detected apart from the singly ionized specie. Their m/z values were smaller than the target and by using MS/MS analysis it was found that they corresponded to γ -fragmentation products of the target molecule and were therefore artefacts of the analytical technique.

4.5.2. Synthesis of Smad3 pT179-pS204-pS208

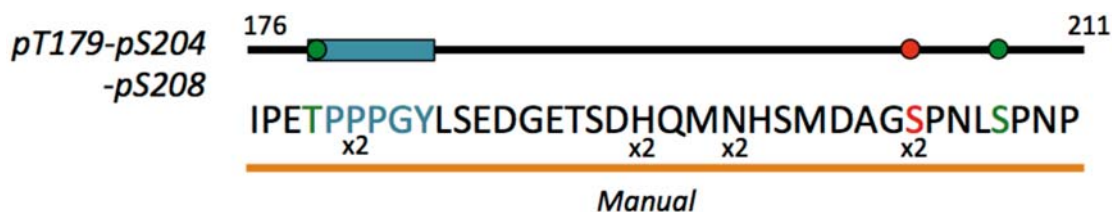


Figure 4.17: Synthesis of Smad3 pT179-pS204-pS208.

In contrast to the syntheses presented in the previous section, obtaining Smad3 pT179-pS204-pS208 was not trivial. Its length (36 residues) together with the introduction of three phosphorylated amino acids posed remarkable challenges to this synthesis.

In a first trial, a combined strategy was designed. [The C-terminal segment from P211 to D193 was synthesized using an automated microwave synthesizer.](#) The [rest of the sequence \(residues S192 to I176\)](#) was added manually. After the synthesis was complete and the peptide was cleaved from the resin, the MS analysis of the crude did not reveal the presence of the corresponding target ion. [Therefore, the synthesis was repeated completely in manual mode.](#) The growing chain was checked after every few coupling-deprotection cycles using the mini-cleavage strategy described in Section 4.2.3, with up to 12 tests performed along the synthesis.

As revealed by mass spectrometry, pS204, N197, H194 and P181 needed to be re-coupled. Oxima Pure + DIC (3eq each) were used as activator agents instead of HATU + HOAt (3 eq each) in these steps, yielding successful couplings in all cases. After this, acetylation of the remaining free amines prevented the formation of the truncated by-products.

After completion of the amino acid chain, the peptide was acetylated and cleaved using the same mixture as described in Section 4.4.3, which contains 2-PySH as a complementary scavenger to TIS and water in order to avoid methionine alkylation by the phosphate protecting group (Bzl).

[RP-HPLC](#) purification of the peptide using 0.1% FA in the mobile phase combined with a SunFire C₁₈ column yielded a well-resolved profile. The target peptide was obtained with 97.5% purity and detected as the [singly charged ion](#) by MALDI-MS (m/z calculated: 4102.11; m/z observed: 4099.6) and as the [triply charged ion](#) by ESI-MS (m/z calculated: 1368.04; m/z observed: 1368.18).

4.5.3. Biophysical bases for the Smad3 action turnover switch

The kinase-mediated switch on the preferred Smad3 binding partners was investigated using the set of peptides shown in **Figure 4.15**. These studies were presented as part of Dr Nina Görner's thesis.

Again, ITC was used to measure the binding affinities of Pin1 and Nedd4L for the different Smad3 phosphorylation patterns (**Figure 4.18A**). CDK8/9 phosphorylation of T179 enables the single WW domain of Pin1 to bind the pT[PY] box ($12.6 \pm 1 \mu\text{M}$). The WW2 domain of Nedd4L recognizes this motif as well ($8.5 \pm 0.3 \mu\text{M}$).

Because both flavopiridol (CDK8/9 inhibitor) and LiCl (GSK3- β inhibitor) diminished the binding of Nedd4L to Smad3 *in vivo*, we postulated that another WW domain in Nedd4L binding to the phospho-S/TP cluster must help in the interaction. The WW2-WW3 was the only pair showing high affinity for the CDK8/9 phosphorylated sequence (pT179-pS208, $2.9 \pm 0.1 \mu\text{M}$). The binding was even tighter when the GSK3- β position S204 was phosphorylated as well ($0.7 \pm 0.2 \mu\text{M}$).

Nedd4L WW2 and WW3 domains are connected through an 80 residues linker. In NMR spectra residues in this region show poor signal dispersion. When solving the structure of the WW2-WW3 segment bound to Smad3 pT179-pS204-pS208, a segmental labelling strategy was used (**Figure 4.18B**). Fully deuterated WW2 and protonated WW3 were purified separately and ligated using a disulphide bond strategy (Baca et al., 1995). This allowed the assignment of the residues in the domains and their contacts with the peptide. 80% of the residues in the linker were assigned as well, and the analysis of the ^{13}C chemical shifts revealed that they do not adopt any secondary or tertiary structure. A model of the interaction is shown in **Figure 4.18C**.

A detailed view of the pT[PY] box coordination by Nedd4L WW2 is shown in **Figure 4.18D**. Residues P181 and P182 are accommodated in the XP groove formed by Tyr382 ($\beta 2$) and Trp393 ($\beta 3$) while Arg380 in $\beta 2$ coordinates pT179 through an electrostatic interaction. His386 and Arg389 in the second loop bind Y184. The coordination of the phospho-S/TP cluster by WW3 is displayed in **Figure 4.18E**. Arg486 in $\beta 1$ and Arg492 in $\beta 2$ establish electrostatic contacts with pS204 and pS208 respectively. Importantly, Nedd4L binds Smad3 with all proline residues in the phospho-S/TP cluster in *trans* configuration.

The WW domain of Pin1 binds the pT[PY] box in a different orientation (**Figure 4.18F**). pT179 is coordinated by Arg18 in $\beta 1$ and Arg21, while P180 occupies the XP groove. Pin1 also binds Smad3 with proline residues in trans configuration.

Collectively these results point towards a scenario where Pin1 and Nedd4L compete for binding Smad3 after CDK phosphorylation. Pin1 may control the ability of Nedd4L WW3 domain binding to the Smad3 phospho-S/TP cluster through *cis* to *trans* isomerization of the proline residues located there. GSK3- β phosphorylation further enhances Nedd4L binding.

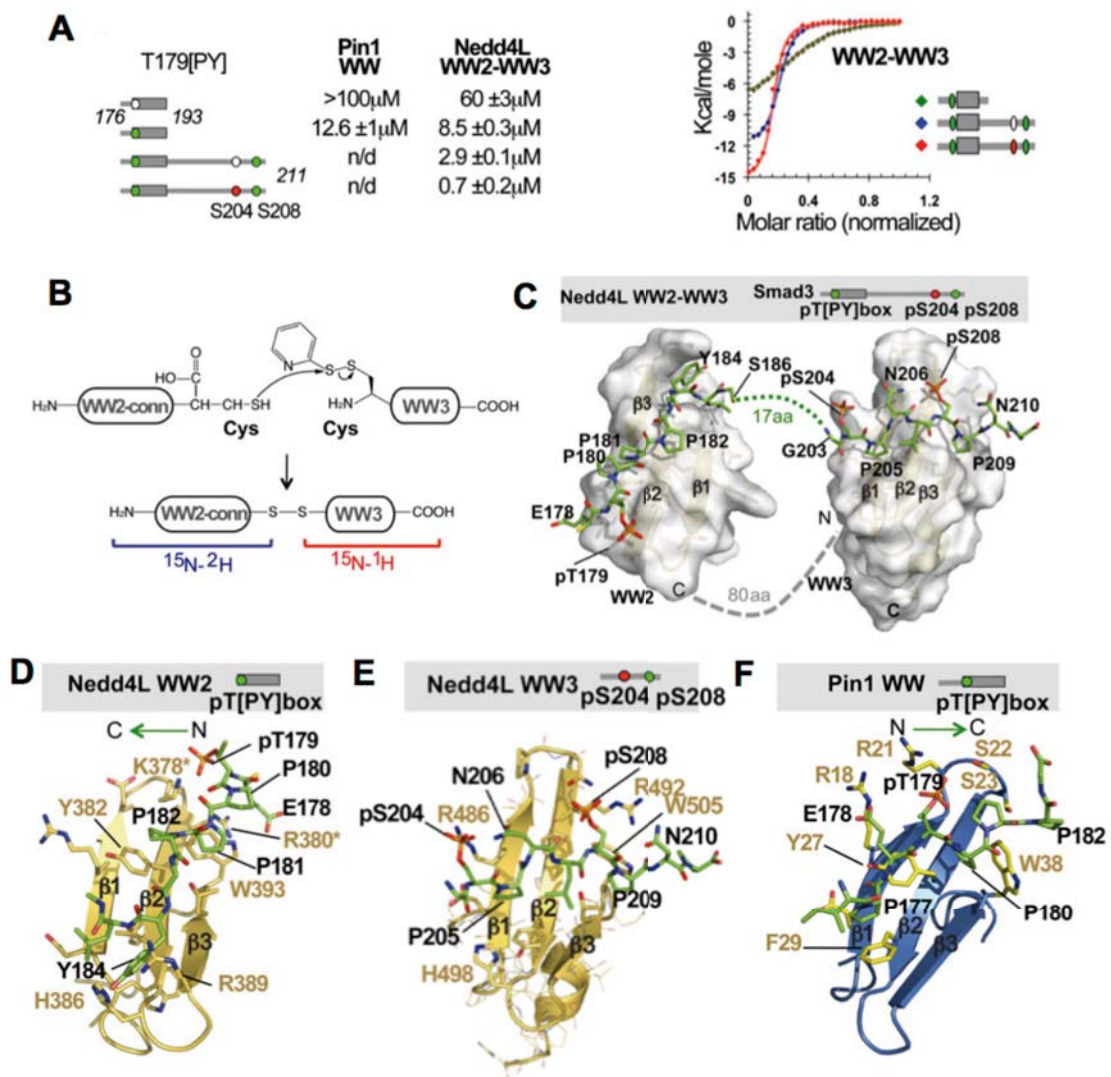


Figure 4.18: Biophysical and structural bases of the Smad3 action/turnover switch. (A) ITC binding isotherms and K_d values derived from these experiments. (B) Chemical ligation protocol for sequential labelling of the Nedd4L WW2-WW3 tandem. (C) Model for the interaction of Nedd4L WW2-3 with Smad3 pT179-pS204-pS208 derived from the structures of each individual domain with the corresponding peptide segment. (D) Structure of Nedd4L WW2 in complex with the Smad3 pT[PY] box. (E) Structure of Nedd4L WW3 in complex with the Smad3 phospho-S/TP cluster. (F) Structure of Pin1 WW domain in complex with the Smad3 pT[PY] box.

4.6. Native chemical ligation of a phosphopeptide

As described in the previous sections, our studies often require the synthesis of long sequences by SPPS combined with the introduction of phosphorylated residues. Obtaining the desired products is often demanding and challenging. These are the reasons why we have explored the use of native chemical ligation (NCL) as a means to simplify our syntheses and overcome these challenges in an easier manner.

With the aim to test this technique on a phosphopeptide, we used a segment of Smad3 comprising residues from D193 to P211 with two phosphorylated residues (pS204 and pS208, **Figure 4.19**). This peptide was called Smad3 pS204-pS208 and it was obtained by laying away a fraction of the resin used to synthesize Smad3 pT179-pS204-pS208 (Section 4.5.2). A cysteine residue was added to the N-terminus of the sequence (position 192) and the peptide was cleaved and purified by RP-HPLC.

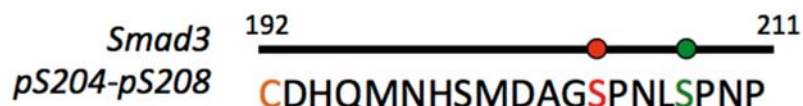


Figure 4.19: Sequence of Smad3 pS204-pS208.

A short peptide thioester was synthesized for this trial. This synthesis and the ligation experiment are detailed below (Section 4.6.1).

As long as there are no cysteine residues in the Smad1 and Smad3 sequences synthesized in previous sections, the possibility to convert the cysteine residues required for the ligation into alanine by desulfurization is attractive to our particular case. The application of this reaction on a phosphopeptide is presented as well (Section 4.6.3).

4.6.1. Synthesis of the peptide thioester GYPGA-COS-Bzl

The synthesis of peptide thioesters by SPPS presents substantial differences when compared to conventional peptides (**Figure 4.20**). For our first synthesis of a peptide thioester we selected a short and simple sequence that was not presenting major difficulties *a priori*: GYPGA.

The 4-Sulfamylbutyryl AM resin with a 0.79 mmol/g load was used for this synthesis, which was performed at a 0.15 mmol scale. The coupling of the first amino acid to this resin requires special conditions.

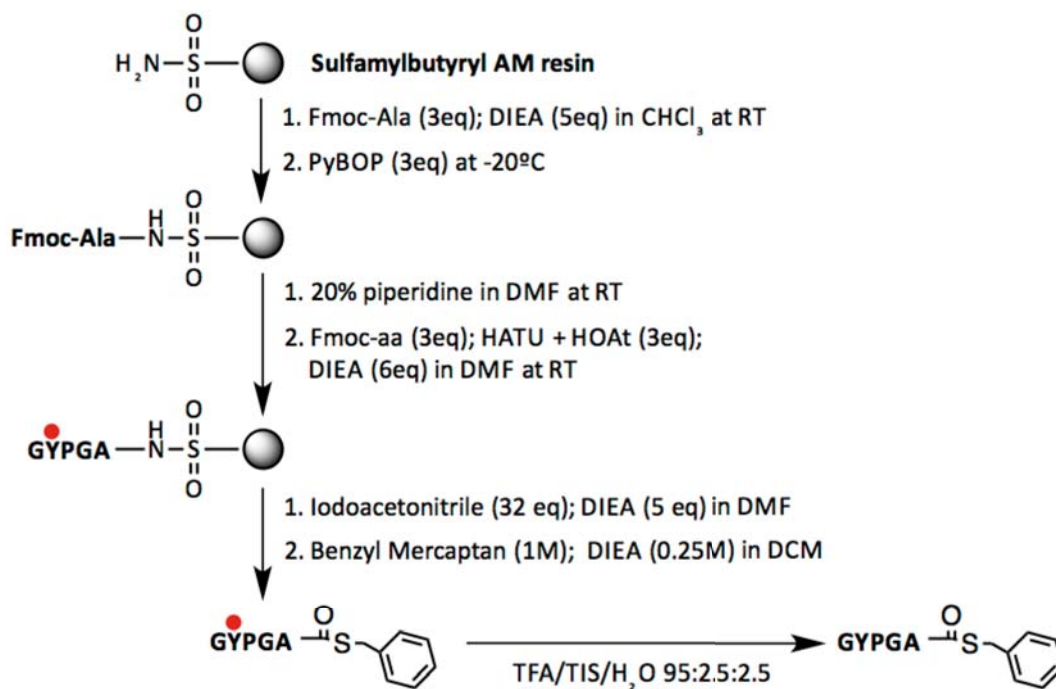


Figure 4.20: Synthesis of the peptide thioester GYPGA-COS-Bzl.

The protocol for this first coupling was designed according to reference methods (Backes and Ellman, 1999; Huse et al., 2000). 3 equivalents of the Fmoc-Ala amino acid were dissolved in chloroform (previously filtered in basic alumina) together with 5 equivalents of DIEA. The mixture was added to the resin in a round-bottom flask and agitated for 10 minutes.

The flask was then immersed in an isopropanol bath and cooled down to -20°C using a cryocooler. The mixture was equilibrated for 20 minutes before adding 3 equivalents of PyBOP. The coupling reaction proceeded for 8h at -20°C with continuous agitation. After that, the resin was washed with pre-filtered chloroform and DMF before swelling with DCM. The next four residues were coupled normally and the peptide was acetylated.

The resin cleavage protocol also differs extensively from that used in conventional peptides. The thioester bond between the peptide and the resin is not labile to acidic conditions. Instead, the peptide is cleaved using a thiol reagent that displaces the peptide from the resin.

To activate the resin for the cleavage, it was agitated overnight with a 2 mL mixture of iodoacetone nitrile (32 eq) with DIEA (5 eq) in DMF. After that, the resin was thoroughly washed with DMF and DCM. The peptide was cleaved by treating the resin for 6 hours with a 43 mL solution of 1 M benzyl mercaptan and 0.25M DIEA in DCM.

DCM was then evaporated using a nitrogen flow and the side-chain of the Tyr residue in the sequence was deprotected using a TFA:H₂O:TIS (95:2.5:2.5) mixture. The TFA was evaporated and the peptide was precipitated using cold tert-butyl methyl ether. It was then solubilized in 5% acetonitrile in H₂O and lyophilized.

The LC-MS analysis of the crude revealed that the peptide was notably pure (89.6%). The singly ionized peptide was detected (m/z calculated: 611.71; m/z observed: 612.21). Therefore, no RP-HPLC purification step was required and the peptide thioester was used straightaway in the ligation reaction.

4.6.2. Ligation of GYPGA-COS-Bzl with Smad3 pS204-pS208

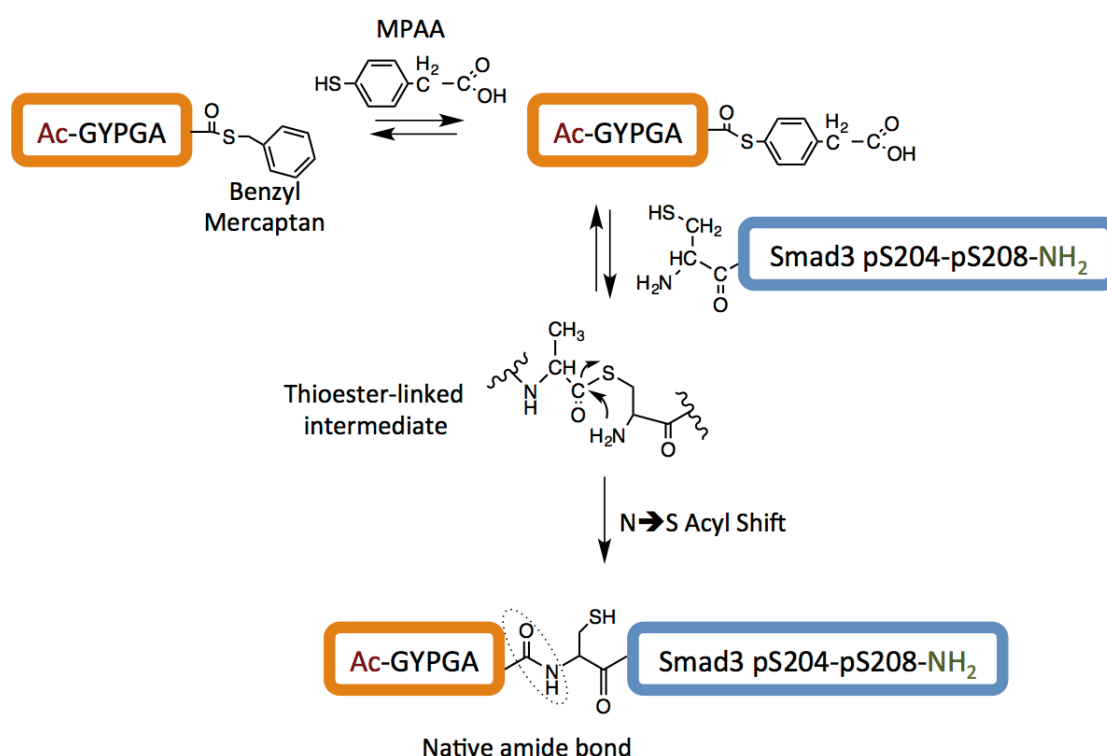
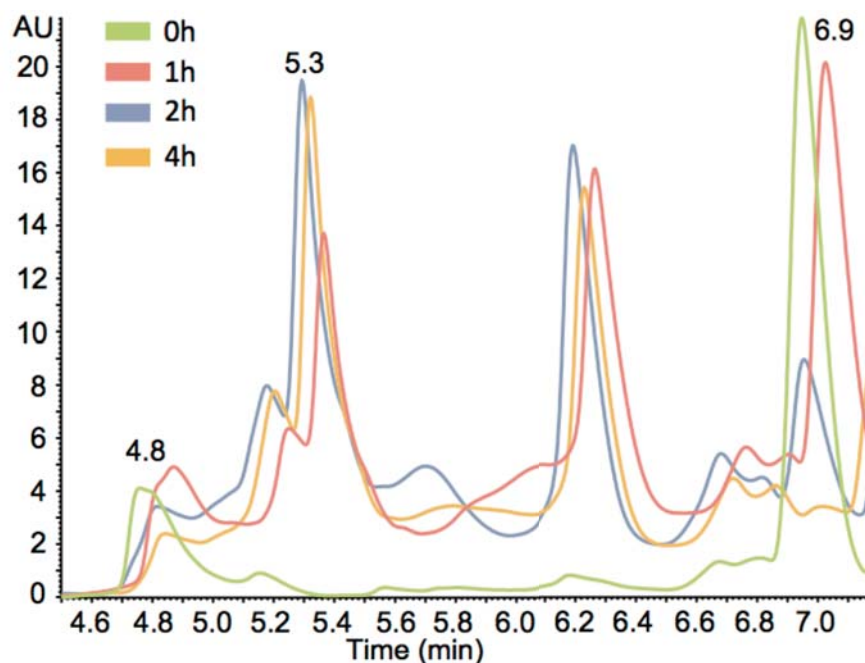


Figure 4.21: Scheme of the ligation of GYPGA-COS-Bzl with Smad3 pS204-pS208.

For the ligation of these two peptides, we used an optimized protocol where a substituted thiophenol with $pK_a > 6$ (MPAA) is added as a catalyst (Johnson and Kent, 2006). MPAA is able to exchange rapidly with the benzyl mercaptan moiety in the GYPGA-COS-Bzl peptide thioester to act as a better leaving group in the transesterification reaction with Smad3 pS204-pS208 (**Figure 4.22**). It has been shown to fasten the reaction time in an order of magnitude, from 24-48h to a few hours.

Both peptides were dissolved to a final concentration of 1mM in 5mL of degassed NCL buffer, containing 6M guanidine hydrochloride, 50mM MPAA, 20mM TCEP and 100mM phosphate at pH 7. The reaction proceeded for 4 hours in a round-bottom flask with continuous orbital agitation.



Retention Time (min)	m/z	Molecule	Area (E+05)			
			0h	1h	2h	4h
4.8	1157.53 (doubly charged) 757.85 (triply charged)	Smad3 pS204-pS208	6.11	3.35	1.10	1.15
5.3	1400.58 (doubly charged) 934.02 (triply charged)	Ligation product	0.00	3.17	5.63	8.42
6.9	612.49 (singly charged)	Ac-GYPGA-COS-Bzl	23.4	8.00	10.3	0.12

Figure 4.22: Progress of the ligation reaction monitored by LC-MS. The UV detected chromatogram for samples taken at time points 0, 1h, 2h and 4h is shown in green, pink, blue and yellow respectively. The table below displays the detected ions and their corresponding molecules together with the integrated peak area at each data point.

During the reaction, 10 μ L aliquots were taken at time points 0, 1h, 2h and 4h. These were diluted in 190 μ L of 50% ACN in H₂O with 0.1% TFA and analysed with LC-MS straightaway. **Figure 4.22** shows the progress of the reaction as monitored using the HPLC profiles of these aliquots. The intensity of the peaks corresponding to each of the separate reacting peptides rapidly decreased during the first two hours of the reaction (retention times 4.8 and 6.9). The peak corresponding to the ligated product (retention time 5.3) appeared already after one hour of reaction and kept

growing one more hour. The intensity of all peaks was stable at time point 4h, when the reaction was stopped.

The ligated peptide was detected as the [doubly and triply charged ions](#) (m/z calculated: 1400.4 and 933.93; m/z observed: 1400.58 and 934.16). Purification with RP-HPLC yielded a final product with 92.1% purity.

4.6.3. Desulfurization of Smad3 pS208

This reaction was tested on Smad3 pS208 (**Figure 4.23**). This short peptide comprises Smad3 residues from P205 to P211 (including pS208). In the N-terminus, glycine was added at position 204 and cysteine at position 203. Using LC-MS, it was detected as the [singly charged](#) and the [doubly charged ions](#) (m/z calculated: 977.0 and 489.5; m/z observed: 977.61 and 489.42).

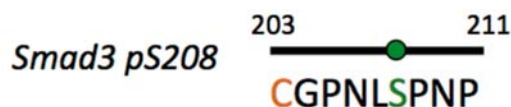


Figure 4.23: Sequence of Smad3 pS208

The peptide was dissolved in a round-bottom flask to a concentration of 1.6 mg/mL in 20% acetic acid (Yan and Dawson, 2001). 10% palladium on activated carbon was added to the solution and the atmosphere was purged with H₂. The reaction proceeded overnight with orbital agitation.

After that, an aliquot was taken, filtered and analysed by LC-MS. A decrease of the m/z values for both the singly ($\Delta m/z = -32.19$) and doubly charged ($\Delta m/z = -16.19$) ions was detected (**Figure 4.24**), corresponding to the loss of the thiol group in the N-terminal cysteine (m/z calculated: 944.93 and 473.47; m/z observed: 945.42 and 473.23). The original peptide was no longer detected. The reaction was then filtered and the desulfurized peptide was lyophilized.

With this test we could check that a cysteine residue in a phosphopeptide can be converted into alanine using desulfurization without having negative consequences on the phosphorylated residues. This is useful in our particular case given that all Smad1 and 3 sequences synthesized do not contain cysteine residues.

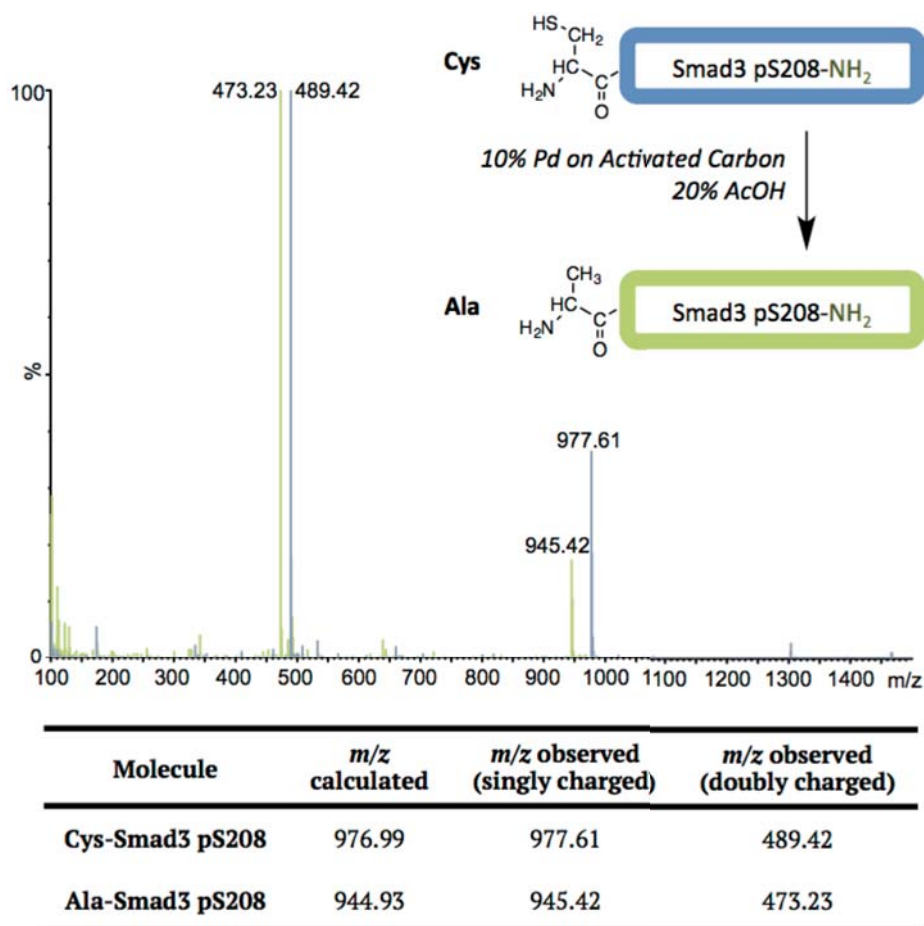


Figure 4.24: Overlaid ESI-MS spectra of Cys-Smad3 pS208 and its desulfurization product.

4.7. Summary

In this chapter the synthesis of two sets of phosphopeptides corresponding to different CDK/GSK3- β phosphorylation patterns of Smad1 and Smad3 is presented. Due to the incorporation of modified amino acids in relatively long sequences (up to 36 residues), some syntheses required optimization at different steps including amino acid coupling, peptide cleavage from the resin and purification.

Optimal conditions for phospho-amino acid coupling in a microwave synthesizer were used. In these cases, β -elimination of the phosphate group by piperidine was avoided by using DBU for the deprotection step (Section 4.4.2). In the manual mode, some hindered residues required re-coupling, and the efficiency of these steps was improved with the combined use of triazoles and carbodiimides as the coupling activating agents.

Oxygen-free conditions were applied to avoid Met oxidation during cleavage. 2-PySH, a scavenger used in Boc chemistry, was added to the cleavage mixture to avoid

Met alkylation by the phosphate protecting group of phosphorylated amino acids (Bzl) (Section 4.4.3).

Truncated by-products lacking one or more phospho-amino acids are sometimes intrinsic in the synthesis of phosphopeptides given the eventual low coupling yield of these residues. Purification conditions were found to separate the target peptides from these undesired by-products (Section 4.4.3).

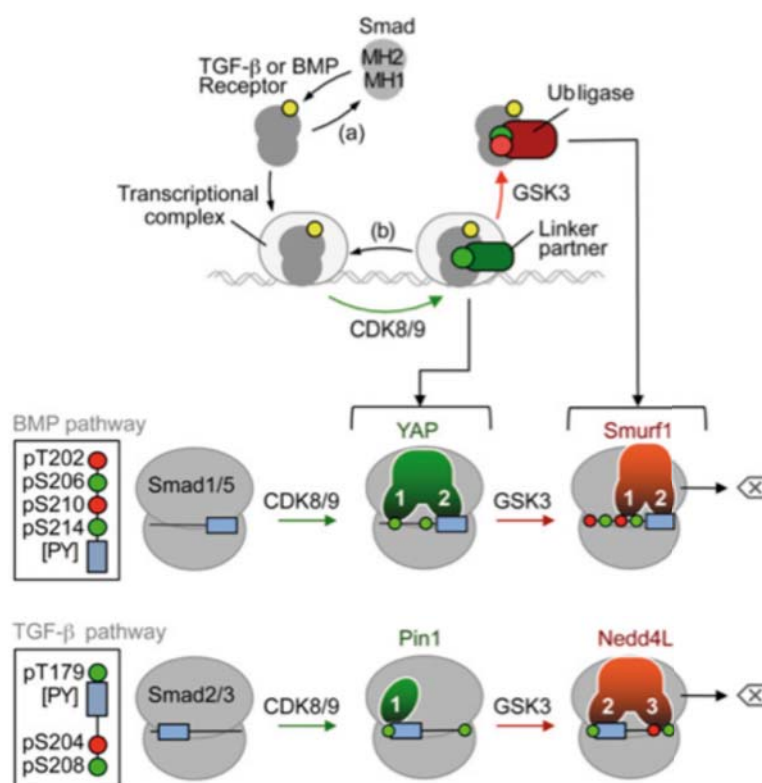


Figure 4.25: Model for the Smad activation turnover switch. (Figure prepared by Dr Joan Massagué). In both the BMP and the TGF- β signalling pathways, phosphorylation of the Smad1 and Smad2/3 linker regions by CDK8/9 creates a docking site for the preferential binding of YAP (a transcription factor) and Pin1 (a peptidyl-prolyl isomerase). This event also triggers the phosphorylation of the region by GSK3- β , which switches the preferred binding partners to Smurf1 and Nedd4L, both E3 ubiquitin ligases of the Nedd4 family that label the Smads for proteasomal degradation.

Collectively this enabled us to obtain peptide samples with sufficient amounts and purity for our interaction studies with YAP and Smurf1 in the case of Smad1, and Pin1 and Nedd4L in the case of Smad3. From these experiments we obtained the biophysical and structural basis explaining the regulation of the R-Smad fate by the linker kinases CDK8/9 and GSK3- β , which modulate the substrate binding capabilities of the pathway associated proteins (Sections 4.4.5 and 4.5.3). Thus, substrate phosphorylation is a regulatory mechanism of the ubiquitinating activity of Nedd4L and the closely related ligase Smurf1. The results from these studies were reported in Dr Eric Aragón and Dr Nina Görner theses respectively and are

summarized in this chapter to illustrate the application of the reported syntheses. **Figure 4.25** displays the model we draw in agreement with the results obtained.

Finally, the application of native chemical ligation to a phosphorylated sequence is presented together with the synthesis of a peptide thioester and the desulfurization of a phosphopeptide to convert Cys residues to Ala. The incorporation of these techniques to our repertoire enables us to expand the size of the phosphorylated sequences we can use in our studies.

CHAPTER 5. Activity regulation of the E3 ubiquitin ligase Nedd4L: Ca²⁺/IP₃ signalling triggers the transition from the latent to the active conformation[‡]

5.1. Introduction

Even though the effect that E3 ubiquitin ligases have on their substrates is characterized to a good extent (**Figure 1.1**), the regulation of the activity of many of them is still to be explored in depth. While the abundance of these proteins is regulated both at the transcriptional level (Chen and Matesic, 2007) and through auto-ubiquitination (CHAPTER 6), phosphorylation of their ligands (CHAPTER 4) and the ligases themselves modulates their substrate recognition capabilities and thus their activity.

Recently another major means for the activity regulation of these ligases has been unveiled. They have been found to exist in latent conformations that are able to respond to specific triggers that render them active. This involves conformational rearrangements consisting in the transition from a closed, inactive conformation to an open and active one.

This activity regulation mechanism has so far been found in several ligases of the Nedd4 family (**Figure 5.1**). Biochemical data showed that in the case of Itch a closed conformation is established through contacts between the catalytic HECT domain and either the WW domains or a Proline Rich Region (PRR) in the linker region between the C2 and the WW1 domains (Gallagher et al., 2006). JNK-1 phosphorylation of the PRR activates the ligase by triggering the transition from the latent to the active form.

Differently, the closed and inactive conformation of Smurf2 is achieved through C2-HECT contacts as shown by extensive biochemical and structural characterization (Wiesner et al., 2007). These contacts are displaced by WW recognition of the ligase's substrate Smad7. So far, biochemical studies have proposed similar closed conformations mediated by C2-HECT contacts in the cases

[‡] Escobedo, A., Gomes, T., Aragón, E., Martín-Malpartida, P., Ruiz, L., Macias, M.J. (2014). *Structural basis of the activation and degradation mechanisms of the E3 ubiquitin ligase Nedd4L*. **Structure** 22, 1446-1457.

of Nedd4L and Smurf1. These contacts have been shown to be inter-molecular in the case of Smurf1 (Aragón et al., 2012; Wan et al., 2011). Regarding Nedd4L, it remains unclear whether these contacts are inter or intra-molecular.

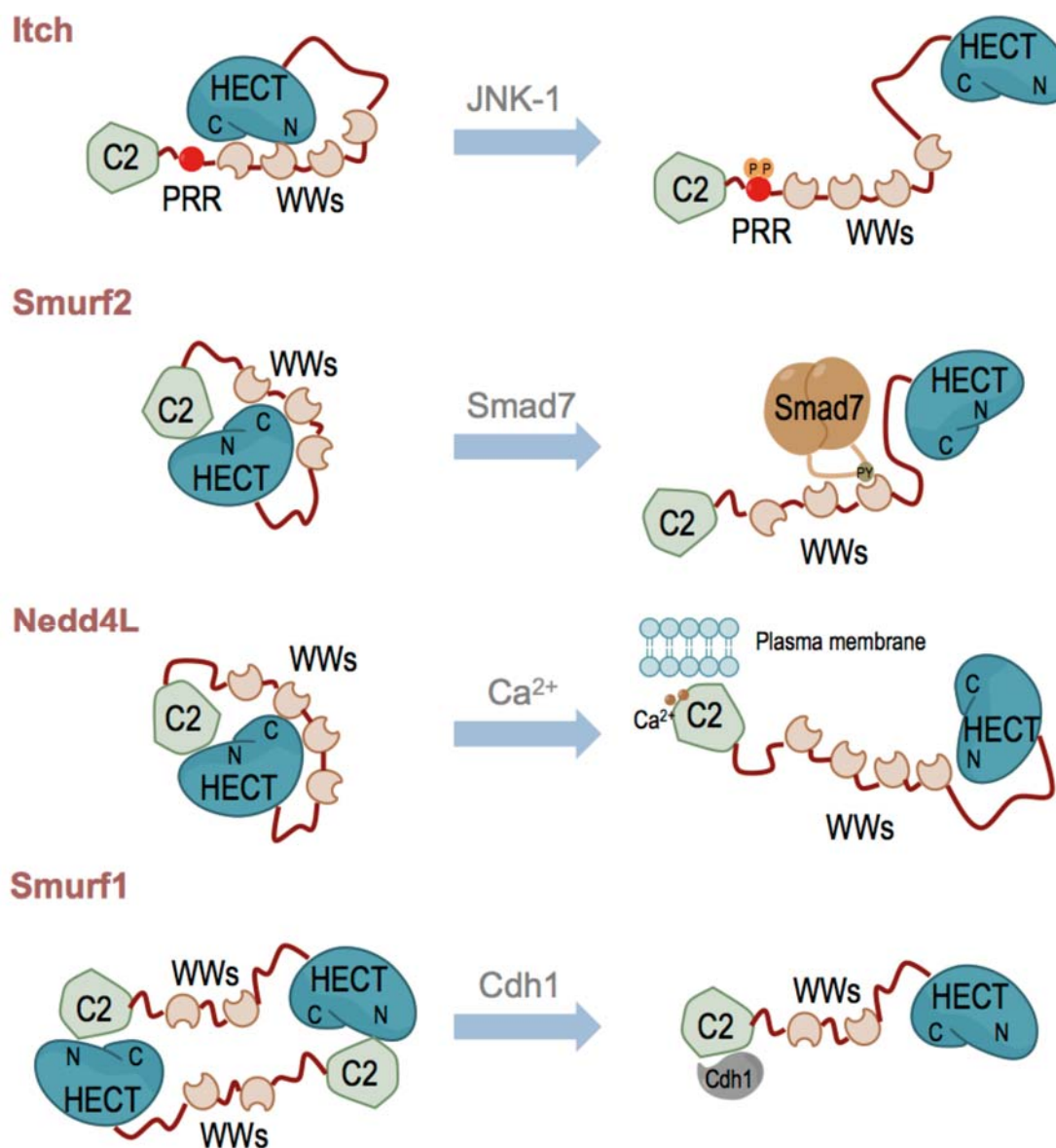


Figure 5.1: Closed/inactive to open/active conformational transitions of the ligases in the Nedd4 family. Schematic representation of the different models proposed for this mechanism of activity regulation. The closed conformation of Itch is established through contacts between the HECT domain and the central region of the ligase. Phosphorylation of the proline rich region (PRR) by JNK-1 renders the ligase active. Differently, Smurf2 presents intramolecular contacts between the C2 and HECT domains in its closed conformation, which are displaced by WW recognition of the substrate Smad7. A similar situation has been described for Nedd4L, although the activating event is a Ca^{2+} level increase in this case, which is also related to the relocation of the ligase to the plasma membrane. It is not clear whether the contacts are intra or inter-molecular. The C2-HECT contacts in the closed conformation of Smurf1 are inter-molecular. Binding of Cdh1 to the C2 domain displaces these contacts and activates the ligase.

Available data report the existence of Nedd4L in a closed and inactive conformation mediated by C2-HECT contacts, whose transition to the open and active form is triggered by an increase of intracellular calcium levels (Wang et al., 2010). Nedd4L is mainly localized in the cytoplasm, but Ca^{2+} -mediated activation contributes to re-localizing the ligase to the plasma membrane, which is mediated by the C2 domain (Garrone et al., 2009).

Cytoplasmic Ca^{2+} delivery is a major signalling event in a wide variety of fundamental cellular processes (Section 1.2). The membrane phospholipid phosphatidylinositol 4,5-bisphosphate (PIP_2) plays a central role in Ca^{2+} signalling. In response to external stimuli, plasma membrane receptors recruit phospholipase C (PLC), which hydrolyses PIP_2 yielding diacylglycerol (DAG) and the cytoplasmic second messenger inositol 1,4,5-trisphosphate (IP_3). IP_3 diffuses in the cytoplasm and binds to its receptor in the endoplasmic reticulum (ER), a ligand-gated Ca^{2+} channel that responds releasing Ca^{2+} to the cytoplasm. The action of IP_3 together with other second messengers such as cADP is fundamental at initiating the intracellular Ca^{2+} signalling cascade.

Several C2 domains have been described to bind membrane phospholipids including PIP_2 in a Ca^{2+} -dependent manner (Section 1.2.2). Similar to PH domains, C2 domains bind both PIP_2 and IP_3 , as their inositol moiety is identical. Thus, we set to investigate the structural bases for the roles that the Ca^{2+} and IP_3 levels play in Nedd4L regulation.

We have used NMR as the central technique to perform a structural study of the inter-domain interactions that are relevant for the activity regulation of the ligase. The C2-HECT contacts have been characterized at the residue level, and we have found the structural basis for the regulation of this interaction by Ca^{2+} and IP_3 . We have observed extensive overlapping of the binding sites on the surface of the C2 domain used by the HECT domain, Ca^{2+} and IP_3 . The correlated intracellular increase of these two intracellular messengers constitutes an efficient switch for the transition from the closed and inactive to the open and active conformation of Nedd4L.

5.2. Deciphering the contacts that maintain Nedd4L in a latent state and its switch to the active form

5.2.1. Optimization of HECT and C2 protein sample preparation and labelling

The architecture of HECT domains comprises two subdomains linked together by a flexible hinge, namely the C- and the N-terminal lobes. The C-terminal lobe contains the catalytic cysteine residue that covalently binds ubiquitin before transferring it to the final substrate. It also contains a PY motif, which is a target for WW domain binding.

E2 ligases bind the N-terminal lobe to donate the ubiquitin moieties to the E3 ligases. Recent crystallographic studies on Nedd4L have revealed major conformational rearrangements concerning the relative positions of HECT subdomains that allow ubiquitin loading and transfer (Kamadurai et al., 2009).

To characterize the interaction between the HECT and C2 domains that impairs Nedd4L function (Wang et al., 2010), we prepared recombinant constructs of both domains for expression in *E. coli*. The catalytic cysteine residue of the HECT domain (C922) was mutated to aspartic acid to avoid potential stability-related complications. The expression and purification of enough amounts of the HECT domain labelled in NMR active nuclei was achieved as described (Marley et al., 2001) after extensive optimization. Adequate conditions were found when using a His₆ tagged construct and very mild expression induction with 50 µM IPTG at 15 °C, as well as mild cell disrupting sonication in a buffer containing protein stabilizers such as glycerol and ZnCl₂ (Section 3.1.3). The protein was expressed in solution and kept in native conditions along the purification process. The construct included a TEV (Tobacco Etch Virus) protease site between the His₆ tag and the HECT domain that was used to remove the tag during protein production by controlled TEV protease digestion. Due to the big size of the HECT domain (380 residues, 45 kDa), its relaxation properties were unfavourable and it required deuteration to yield a well-resolved HSQC spectrum (**Figure 5.2**). The experiment was acquired in a high field spectrometer (900 MHz) equipped with a cryoprobe, which extensively improved the data acquired in lower fields (600 MHz) using room temperature probes. Unfortunately, the protein was not stable enough to acquire experiments that required longer acquisition times, such as those used for the sequential assignment of the backbone resonances.

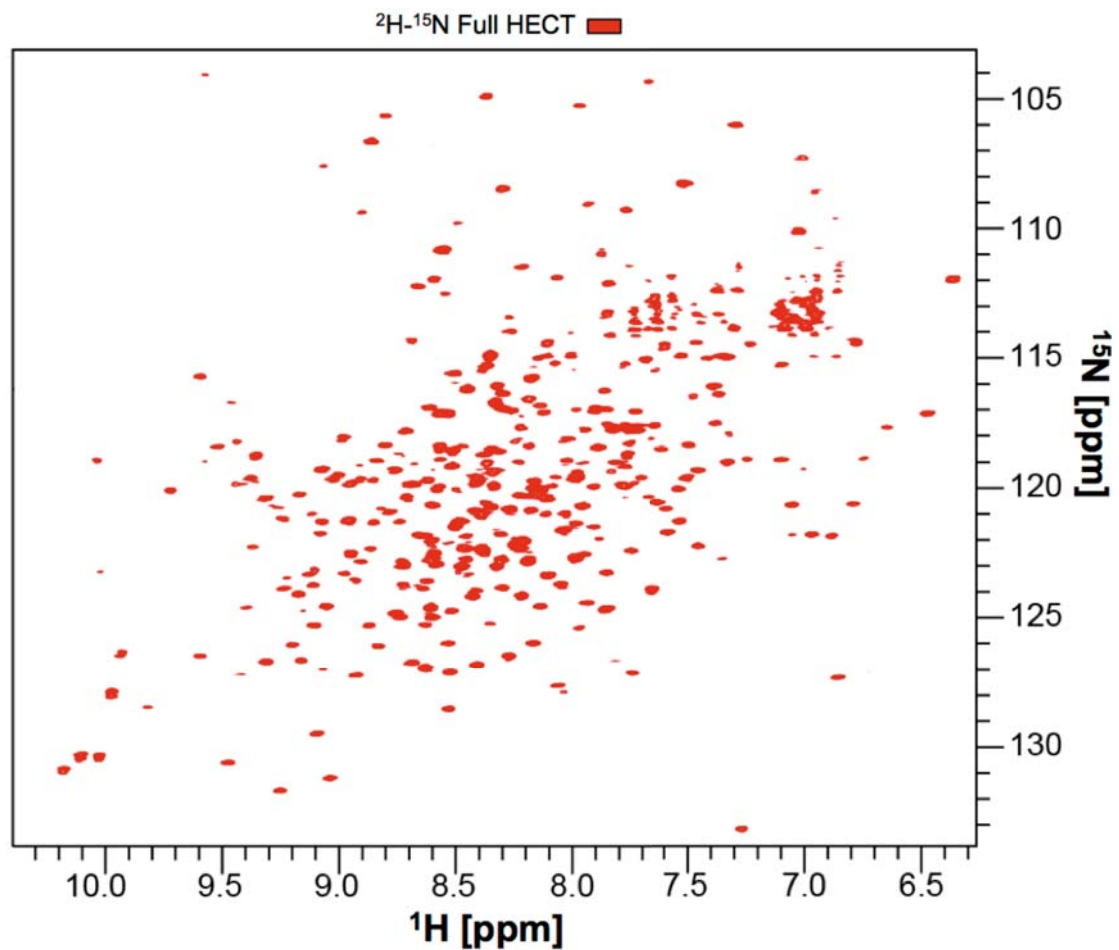


Figure 5.2: HSQC spectrum of the HECT domain. The good dispersion of the signals and the relatively low overlapping are indicative of a folded protein. 93.4% of the expected backbone amide peaks can be observed (339 out of 363 signals). After the experiment acquisition, the solution displayed turbidity as a result of the protein's tendency to aggregate and precipitate. The experiment was acquired on a deuterated 90 μM sample in a Bruker Avance 900 MHz spectrometer equipped with a z pulse gradient 5 mm TCI cryoprobe (FMP, Berlin).

To circumvent this difficulty, a HECT C-terminal lobe recombinant construct was prepared. Because of its smaller size (117 residues, 13.5 kDa) and its higher stability, its properties were more favourable for NMR experiments. No major complications aroused regarding the production and labelling of this protein fused to a His₆ tag. In order to increase the production yields, especially when labelling the protein, it was expressed in inclusion bodies, extracted in a denaturing buffer with 6M guanidine hydrochloride, bound to a Ni²⁺ affinity resin and refolded in a single buffer exchange step while bound (**Figure 3.5**). The protein was then eluted, digested with TEV protease to remove the His₆ tag and purified by size-exclusion chromatography in native conditions.

The expression of the C2 domain required the optimization of the construct boundaries and the final buffer conditions. Although the N-terminal segment comprising the first 15 residues of the domain appears mostly disordered in the

available X-ray structure (PDB code 2NSQ), it was needed for the protein to be soluble (**Figure 5.4B**). A construct lacking this segment was never solubilized despite cloning it in fusion with several tags including His₆, GST, MBP and NusA. In contrast, a construct including this segment was soluble even when expressed at 37 °C, 0.5 mM IPTG. This protein was also expressed in inclusion bodies to improve its production yields. However, it required a slower re-folding procedure by dialysis after Ni²⁺ affinity purification in the denaturing buffer (**Figure 3.5**). High ionic strength in the final NMR buffer was needed as well to prevent precipitation (250 to 400 mM NaCl depending on the length of the experiment to be acquired).

Table 5.1 summarizes the constructs and samples that were prepared to perform the experiments presented in this chapter.

Molecule	MW ^a (kDa)	Cloning ^b	Samples	Production
HECT	44.9	C922D mutant <i>In fusion with:</i> • His ₆ tag • NusA • MBP	• Unlabelled → CSP assays • ¹⁵ N → poor spectrum • ² H ¹⁵ N → good spectrum	• Expression at 15°C, 50 μM IPTG • Native conditions
HECT C-lobe	13.3	C922D mutant • His ₆ tag	• ¹⁵ N → CSP assays • ² H ¹³ C ¹⁵ N → bb. resonance assign.	• Expression in inclusion bodies • Fast refolding bound to Ni ²⁺
C2	17.9	<i>Without 15 aa N-ter segment:</i> • His ₆ tag • NusA • MBP <i>With it:</i> • His ₆ tag	• Unlabelled → CSP assays • ¹⁵ N → CSP assays • ² H ¹⁵ N → CSP assays • ² H ¹³ C ¹⁵ N → bb. resonance assign.	• Expression in inclusion bodies • Slow refolding by dialysis

Table 5.1: Summary of HECT and C2 domains constructs and samples prepared. (a) Molecular weight values have been calculated for unlabelled samples. (b) Colour dots next to each solubility tag in the “cloning” column refer to the actual solubility of the resulting construct. Green: soluble construct in good amounts. Orange: insufficient amounts of soluble construct. Red: insoluble construct.

5.2.2. Backbone resonance assignment of the C2 domain and the HECT C-lobe

Samples labelled in ²H-¹³C-¹⁵N of both the C2 domain and the HECT C-lobe were produced to acquire the HNcoCACB/HNCACB pair of 3D NMR experiments. These were used to assign their backbone resonances.

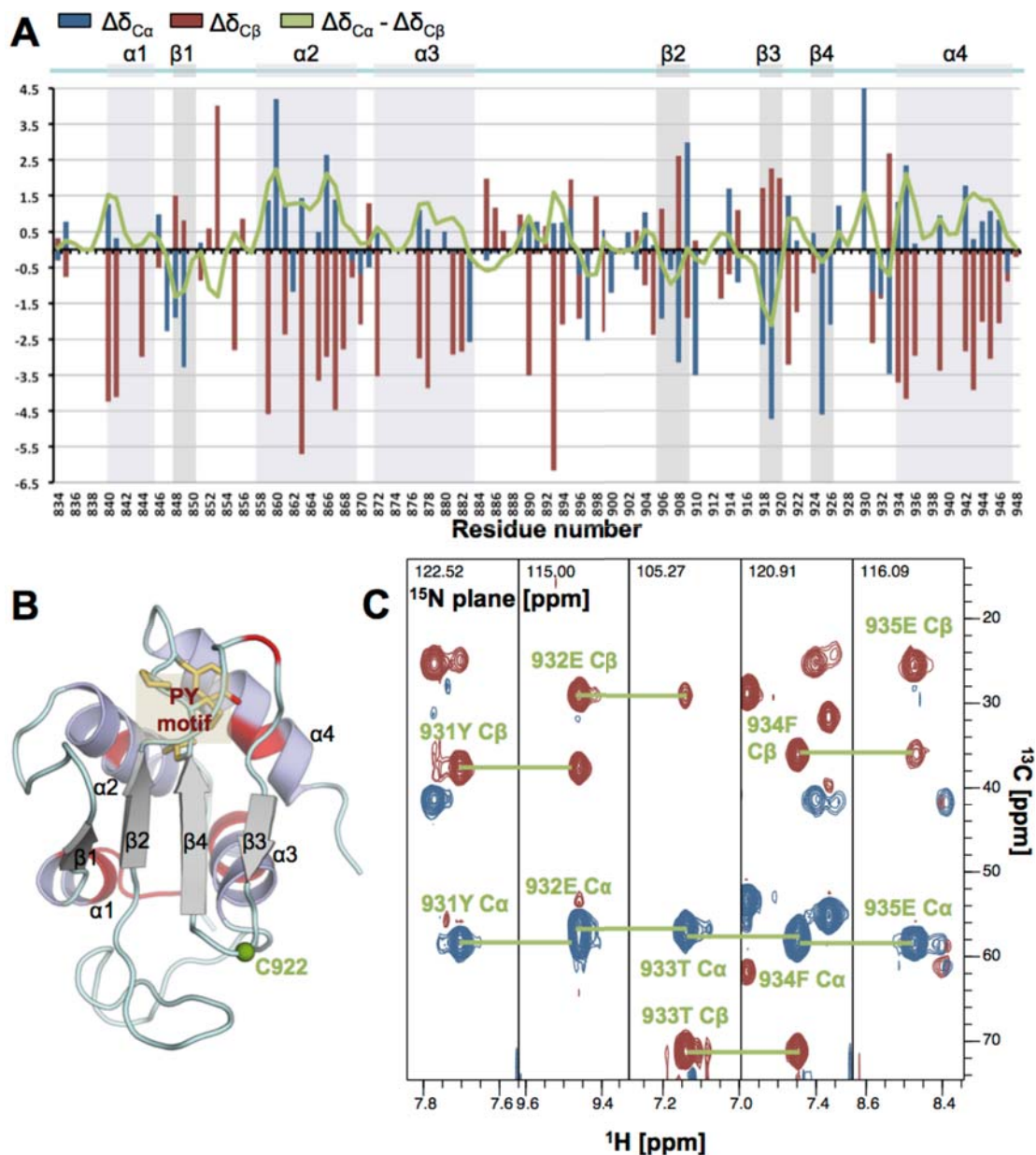


Figure 5.3: HECT C-lobe backbone resonance assignment. (A) Analysis of the $\Delta\delta_{C\alpha}$ and $\Delta\delta_{C\beta}$ secondary chemical shifts and their correlation with the residue location in a secondary structure element, expressed by the value of $\Delta\delta_{C\alpha} - \Delta\delta_{C\beta}$. (B) Crystal structure of the HECT C-lobe displaying the residues with backbone resonances unassigned (red, 15.6%) and assigned (rest, 84.4%). The catalytic residue is shown as a green sphere. The conserved PY motif is displayed in sticks representation (C) Snapshot of the HNCACB spectrum showing the sequential assignment of the Y931-E935 segment (corresponding to the HECT-PY motif).

The CCONH experiment, which shows all side-chain carbon resonances, was acquired as well to help solving ambiguity in certain cases where the C_{α} and C_{β} chemical shifts of two or more residues were indistinguishable.

In the case of the HECT C-lobe, the final construct had 117 residues including 8 prolines. The backbone resonances of 92 residues (84.4%) out of the 109 non-proline residues in the sequence were assigned using the sequential amino acid assignment

strategy (**Figure 5.3C**). The 17 missing resonances (15.6%) were either not present in the spectra or overlapped with other signals, which hampered unambiguous assignment. Approximately half of these resonances (9) corresponded to residues located among the segment comprising the first 12 N-terminal amino acids, which is an area that often presents assignment difficulties (**Figure 5.3B**). All residues that are located in secondary structure elements in the crystal structure of the protein (PDB entry 3JW0) display $\Delta\delta_{C\alpha}$ and $\Delta\delta_{C\beta}$ secondary chemical shifts values that correlate with their expected arrangement (Section 3.4.1.3): $\Delta\delta_{C\alpha} - \Delta\delta_{C\beta}$ values are predominantly positive in α -helices whereas they tend to be negative in β -sheets (**Figure 5.3A**).

The same strategy was used to assign the backbone resonances of the C2 domain (**Figure 5.4C**), which was expressed as a 155 residues-long construct, including 10 prolines. As reported in Section 5.2.4 the set of experiments for backbone resonance assignment was acquired twice for this protein, in Ca^{2+} -free and Ca^{2+} -saturating conditions respectively.

In Ca^{2+} -free conditions, several resonances corresponding to residues located in the $\beta 1$ - $\beta 2$, $\beta 5$ - $\beta 6$ and $\beta 7$ - $\beta 8$ loops were not present in the spectra, which is typical of mobile loop regions sampling different environments. In this case, 80% of the non-proline backbone resonances were assigned (corresponding to 116 residues).

Upon Ca^{2+} binding to the domain, some of the missing resonances appeared in the spectra, particularly in the $\beta 1$ - $\beta 2$ and $\beta 5$ - $\beta 6$ loop regions, which take part of the coordination of the Ca^{2+} ions (Section 5.2.4). Therefore, Ca^{2+} binding to these loop regions results in a reduction of amide broadening, which may be related to a decrease of the mobility of these regions. Finally, the backbone resonances of 122 non-proline residues were assigned (84.1%). Still, the residues that could not be assigned were mostly located in the $\beta 5$ - $\beta 6$ and $\beta 7$ - $\beta 8$ loops (**Figure 5.4B**). The analysis of the $\Delta\delta_{C\alpha}$ and $\Delta\delta_{C\beta}$ secondary chemical shift values shown in **Figure 5.4A** is in good agreement with their expected location in secondary structure elements based on the crystal structure of the domain (PDB entry 2NSQ).

Residues Asp36 – Gly39, in the $\beta 1$ - $\beta 2$ loop, are missing in the available crystal structure of the C2 domain of Nedd4L. As reported below, this region is affected upon Ca^{2+} binding. For clarity reasons in the representations, the loop was reconstructed using Rosetta. Further details are provided in Section 9.2.

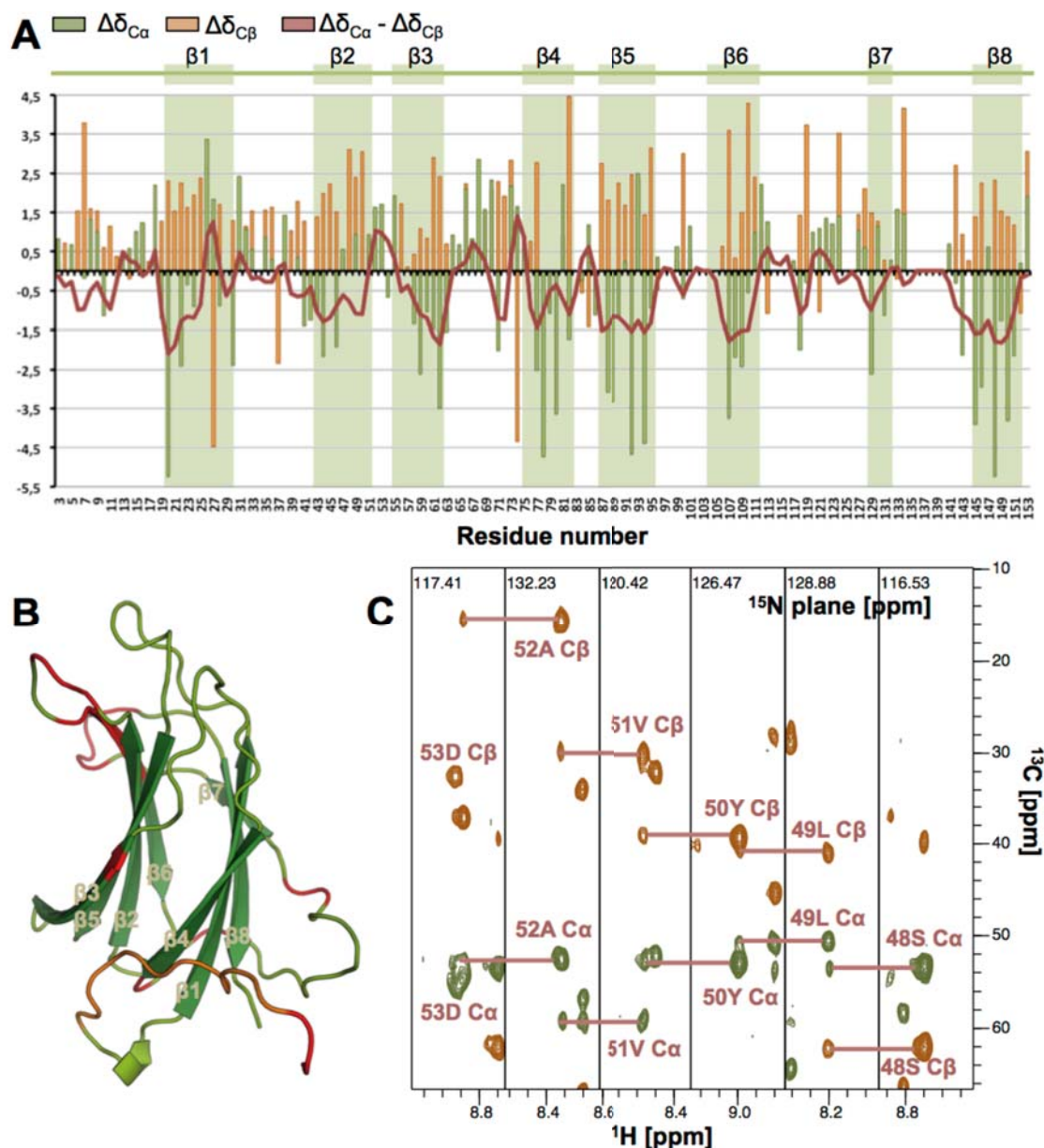


Figure 5.4: C2 domain backbone resonance assignment. (A) The secondary chemical shift $\Delta\delta_{C\alpha}$ and $\Delta\delta_{C\beta}$ values are in agreement with their expected location in secondary structure elements, as $\Delta\delta_{C\alpha} - \Delta\delta_{C\beta}$ values for residues located in β -sheets tend to be negative. (B) Cartoon representation of the crystal structure of the C2 domain (PDB entry 2NSQ) with the unassigned residues displayed in red. Highlighted in orange are the assigned residues corresponding to the N-terminal segment that was required for the solubility of the protein (Section 5.2.1). (C) Sequential assignment of the residues in the 48S – 53D segment in the HNCACB spectrum.

5.2.3. Characterization of the interaction between the HECT and the C2 domains in the closed conformation

With this information in hand, we performed chemical shift perturbation assays to test the interaction between the HECT C-lobe and the C2 domain. Remarkably, the experiment was done using a Ca^{2+} -free buffer, given the reported data suggesting that Ca^{2+} impairs the C2-HECT interaction. The buffer contained 1 mM EGTA, a Ca^{2+}

chelating compound, to neutralize residual Ca^{2+} in the reagents used. Increasing amounts of an unlabelled C2 sample were progressively added to a ^1H - ^{15}N HECT C-lobe sample, and an HSQC spectrum was acquired at every experimental point. After the addition of up to 4 equivalents of the C2 domain, no effect was detected in the HECT C-lobe resonances, indicating that these two proteins do not interact (**Figure 5.5**).

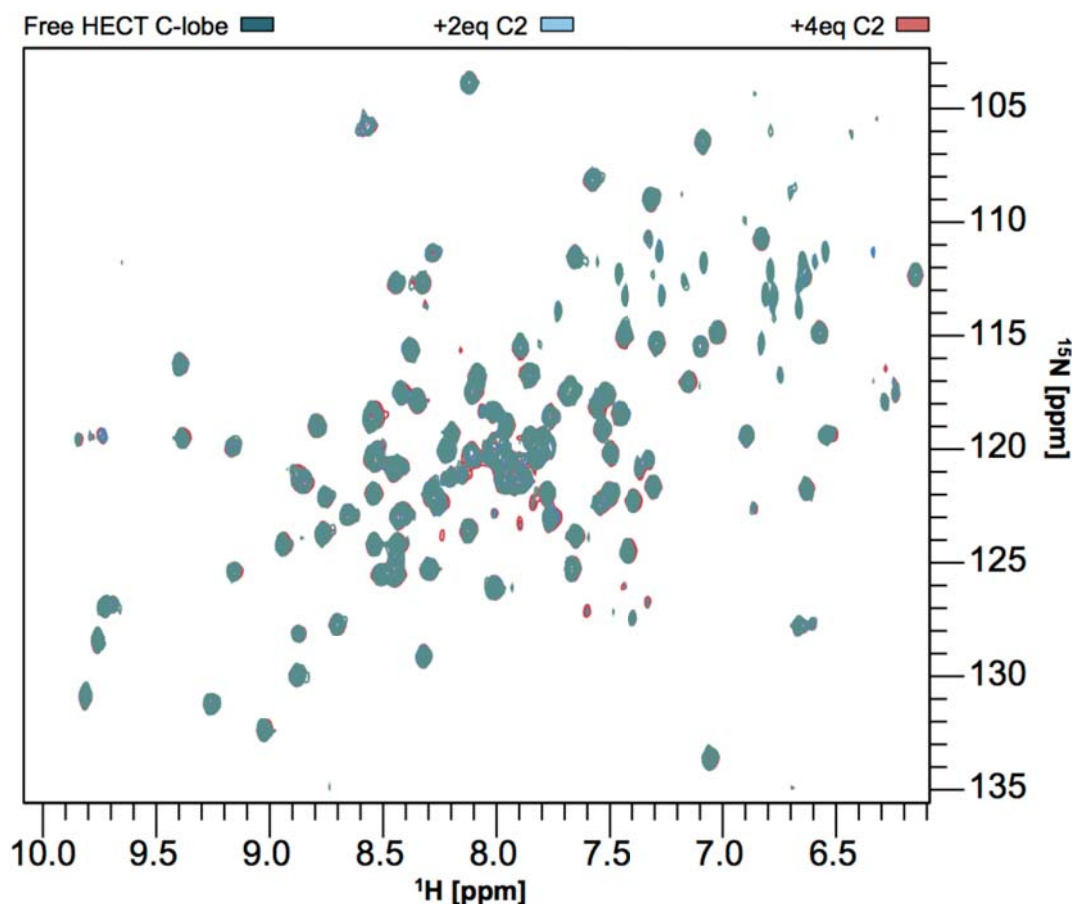


Figure 5.5: HSQC titration reporting no contacts between the C2 domain and the HECT C-lobe. Overlaid view of the spectra acquired on a ^1H - ^{15}N labelled HECT C-lobe sample with added unlabelled C2 domain. Shown in green, spectrum of the free HECT C-lobe (reference). Spectra obtained after the addition of 2 and 4 equivalents of the C2 domain are shown in blue and red respectively. None of the backbone amide peaks of the HECT C-lobe was affected after the addition of up to 4 equivalents of the C2 domain, indicating that these domains do not interact.

Prompted by the observations reported for Smurf2, another case of an E3 ligase in the Nedd4 family where the C2 contacts the HECT domain in its N-terminal lobe (Wiesner et al., 2007), we repeated the chemical shift perturbation assay using the full HECT domain instead of the isolated C-terminal lobe (**Figure 5.6**). Again, the experiment was performed in a Ca^{2+} -free buffer. In this case, we added increasing amounts of unlabelled HECT domain to a ^1H - ^{15}N labelled C2 domain sample. After the addition of 1 equivalent of the HECT domain, about 85% of the C2 resonances were no longer visible in its HSQC spectrum. Only 19 resonances were detected at

that point all of them corresponding to the N-terminal segment of the construct, that was key for the solubility of the protein. Most of the remaining signals are sharp and intense, which is a general characteristic in flexible regions.

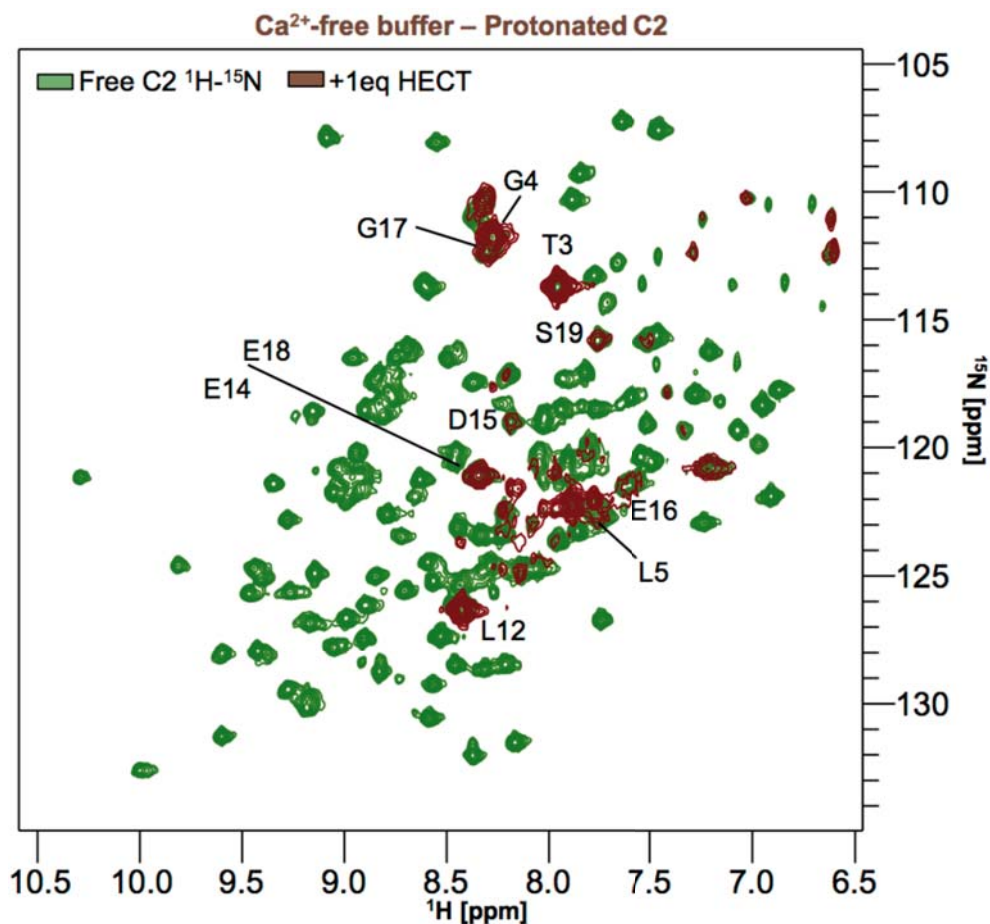


Figure 5.6: HSQC titration showing the establishment of a complex between the C2 and the HECT domains. Shown in green, HSQC spectrum of the free C2 domain labelled in ^1H - ^{15}N in a Ca^{2+} -free buffer. The spectrum in brown corresponds to the same sample after the addition of 1 equivalent of the unlabelled HECT domain. Approximately 85% of the signals are missing in this spectrum, indicative of the establishment of a C2-HECT complex with high molecular weight. The remaining 19 signals correspond to the mobile N-terminal segment.

Given the unfavourable relaxation properties of the HECT domain, we attribute the disappearance of the C2 resonances to the binding of the two proteins. When the C2 domain binds to the HECT domain, its correlation time increases sensibly making anisotropy averaging by tumbling less effective. Chemical shift anisotropy (CSA) is a source of coherence loss, so its inefficient averaging leads to a reduction of the longitudinal relaxation (T_2) times of residues in ordered areas and the subsequent signal broadening. Highly flexible segments are not as much affected as they present independent high tumbling rates. Therefore, the broadening and loss of the C2 resonances in ordered areas upon the addition of the HECT domain is indicative of an interaction between the two domains. The N-terminal lobe of the HECT domain

is thus required for the interaction with the C2 domain and is indeed likely to contain the contacting surface.

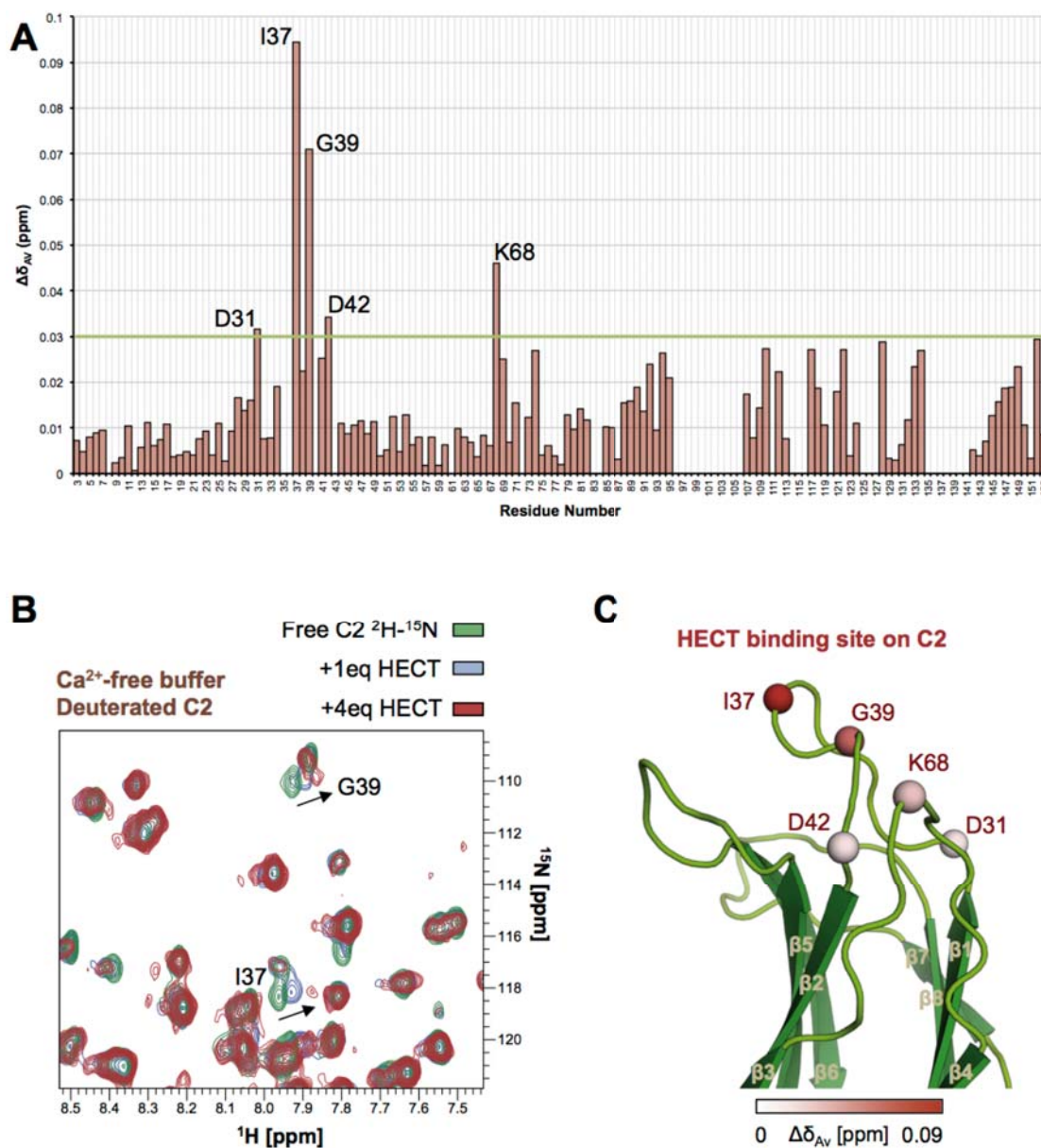


Figure 5.7: HECT binding site mapping on the surface of the C2 domain. Data resulting from the HSQC titration of progressively increasing amounts (up to 4 equivalents) of the unlabelled HECT domain on a ^2H - ^{15}N labelled C2 domain sample. (A) Histogram showing the $\Delta\delta_{AV}$ value (in ppm) per each unambiguously assigned and well-resolved residue of the C2 sequence. Residues displaying values over 0.03 ppm (green line) were considered for the HECT binding site mapping. (B) Close-up view of the overlaid spectra showing the two residues presenting the strongest chemical shift perturbations (CSPs). Shown in green, HSQC spectrum of the free C2 sample. Displayed in blue and red, HSQC spectra acquired on the same sample after the addition of 1 and 4 equivalents of the HECT domain, respectively. (C) Representation of the HECT domain binding site on the surface of the C2 domain. Residues displaying $\Delta\delta_{AV}$ values over 0.03 ppm in (A) are shown as spheres, which have been coloured using a red gradient depending on their individual values.

In order to counteract the signal loss effect that the HECT binding has on the C2 domain signals, we produced a ^2H - ^{15}N labelled sample of the C2 domain. Deuterons have more favourable relaxation properties than protons because they do not participate in ^1H - ^1H dipolar couplings. The sample was prepared by inducing the expression of the protein in an *E. coli* culture in minimum media prepared in D_2O . As usual, the Ca^{2+} -free buffer used in NMR experiments was a 9:1 mixture of H_2O and D_2O . Because protein amide hydrogen atoms are especially labile, they have fast exchange rates with hydrogen atoms in the buffer. Therefore, backbone deuterons in the ^2H - ^{15}N labelled sample exchange with protons in the buffer, making them detectable in NMR experiments including HSQC. Generally, deuterons in all other positions of the protein can not exchange with buffer protons, and they have favourable effects by slowing down the relaxation rates of the protein.

After adding up to 4 equivalents of the unlabelled HECT domain to the ^2H - ^{15}N C2 domain sample, all signals were still visible in the C2 HSQC spectrum. The chemical shift of some resonances was affected, thus reporting their implication in the binding to the HECT domain. These resonances corresponded to residues located in the $\beta 1$ - $\beta 2$ loop (Asp31, Ile37, Gly39 and Asp42) and in the $\beta 3$ - $\beta 4$ loop (Lys68) (**Figure 5.7**). The measured chemical shift perturbations (CSPs) were up to 0.1 ppm ($\Delta\delta_{\text{AV}}$, with a 0.2 ^{15}N weighting with respect to ^1H , Equation 3.12). Because the HECT domain tends to aggregate and precipitate at high concentrations the addition of more than 4 equivalents to the C2 domain was not feasible, which precluded the determination of the apparent affinity of this interaction by either ITC or NMR.

The mapped HECT binding site on the C2 domain of Nedd4L presents notable similarities with that of Smurf2 (Wiesner et al., 2007), both in terms of residue location (mainly $\beta 1$ - $\beta 2$ and $\beta 3$ - $\beta 4$ loops) and of the observed CSP values (0.22 ppm in the case of Smurf2). Therefore, this may be a conserved feature in other ligases of the Nedd4 family that use an analogous mechanism to exist in a latent state within the cell.

5.2.4. Ca^{2+} binding to the C2 domain of Nedd4L

The C2-HECT contacts reported in the previous section had been previously observed *in vivo* in the context of the full-length protein using HEK293 cell cultures and molecular biology methods including immunoprecipitation (Wang et al., 2010). In addition, ubiquitination assays had shown that when these contacts were present the catalytic action of the ligase was impaired. When intracellular Ca^{2+} levels were artificially increased, the contacts were no longer detected and the ligase was rendered active, as shown using β -ENaC and auto-ubiquitination assays.

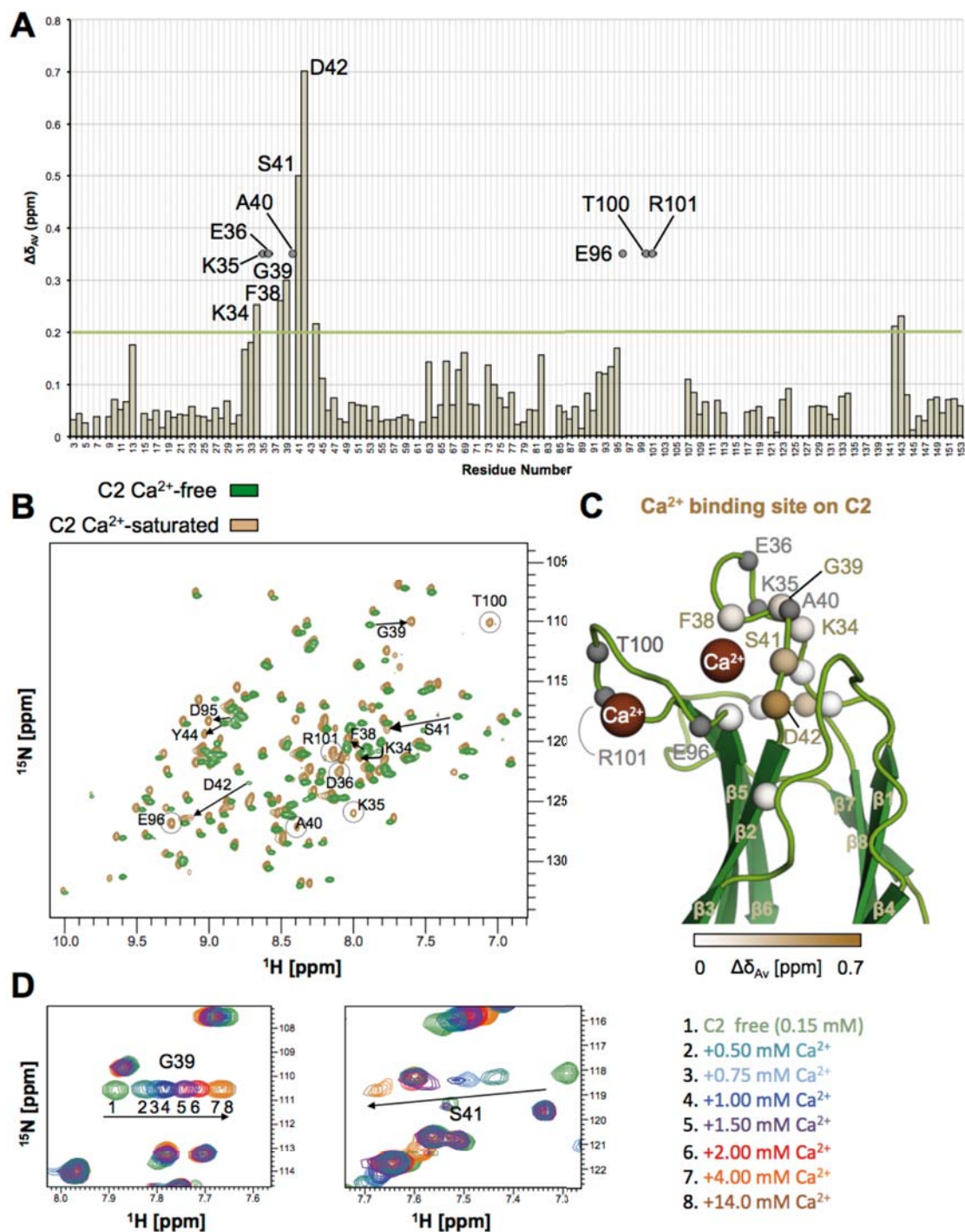


Figure 5.8: Ca^{2+} binding to the C2 domain as monitored by NMR. Data resulting from the comparison of the HSQC spectra of two ^1H - ^{15}N labelled samples of the C2 domain, one in a Ca^{2+} -free buffer and the other one in the same solution containing 20 mM CaCl_2 . (A) Histogram representation of the chemical shift perturbations ($\Delta\delta_{AV}$, ppm) displayed per each residue. Amino acids labelled with a grey circle were only present in the spectrum of the Ca^{2+} -saturated sample. (B) Complete view of the two HSQC spectra. The trajectory of residues with the biggest $\Delta\delta_{AV}$ values is indicated with an arrow. Residues present only in Ca^{2+} -saturated conditions are shown in dashed circles. (C) Mapping of the Ca^{2+} binding site on the surface of the C2 domain. The two Ca^{2+} ions are displayed in agreement with the structural alignment of the Nedd4 and the Nedd4L C2 domains discussed in the main text. Residues displaying $\Delta\delta_{AV}$ values over 0.2 ppm in (A) are shown as spheres and coloured using a brown gradient in accordance with their values. Residues present only in the spectrum of the Ca^{2+} -saturated sample are shown as grey spheres. (D) HSQC titration consisting in the progressive addition of Ca^{2+} to the C2 domain.

To gain structural information on these observations, we prepared a ^1H - ^{15}N labelled C2 domain sample in the same buffer used for experiments reported in Section 5.2.3 with added Ca^{2+} at a saturating concentration (20 mM CaCl_2). Again, the protein was expressed in inclusion bodies, recovered with a buffer containing 6M guanidine hydrochloride, re-folded by dialysis and gel-filtered using a Ca^{2+} -saturated buffer. We acquired an HSQC spectrum of this sample and compared it with the spectra acquired on the Ca^{2+} -free samples.

Because new resonances appeared (**Figure 5.8B**) and the amide chemical shifts of various residues were highly affected, the protein backbone chemical shifts were assigned again in the presence of Ca^{2+} . Using a ^2H - ^{13}C - ^{15}N labelled C2 sample prepared in a Ca^{2+} -saturated buffer we acquired the HNcoCACB/HNCACB pair of 3D NMR experiments. To support the assignment, an HSQC titration consisting in the addition of progressively increasing amounts of Ca^{2+} to a Ca^{2+} -free C2 sample was performed in order to follow the peak trajectory (**Figure 5.8D**).

In the presence of Ca^{2+} , the magnitude of the observed chemical shift perturbations (CSPs) was sensibly bigger than that reported in the previous section, up to 0.702 ppm ($\Delta\delta_{\text{AV}}$) (**Figure 5.8**). The residues that show the biggest effects are located in the $\beta 1$ - $\beta 2$ loop (Lys34, Phe38, Gly39, Ser41 and Asp42).

Apart from the CSPs, the spectra acquired on the Ca^{2+} -saturated samples displayed some extra peaks when compared to the ones acquired in Ca^{2+} -free conditions. These resonances corresponded to residues located in the $\beta 1$ - $\beta 2$ loop (Lys35, Glu36 and Ala40) as well as in the $\beta 5$ - $\beta 6$ loop (Glu96, Thr100 and Arg101). Loop segments connecting structured regions are often mobile and yield poor NMR signals, as residues in these areas sample different environments leading to signal broadening. For this reason, we attributed the appearance of these signals to a reduction of this broadening related to a decrease in the mobility of these areas.

The C2 domains of Nedd4L and the closely related ligase Nedd4 share 73% sequence identity and 83% similarity. The crystal structure of the C2 domain of Nedd4 in complex with Ca^{2+} is publicly available at the PDB database (entry 3B7Y). Its structural alignment with the crystal structure of the Ca^{2+} -free C2 domain of Nedd4L (PDB entry 2NSQ) reveals a very similar fold.

The position of the two Ca^{2+} ions in the Nedd4 C2 domain structure is in agreement with the Ca^{2+} binding site we propose for the Nedd4L C2 domain. Therefore, combining our data with this structural alignment we inferred that the C2 domain of Nedd4L has two Ca^{2+} binding sites located in the $\beta 1$ - $\beta 2$ and the $\beta 5$ - $\beta 6$

loops (**Figure 5.8C**). An electrostatic surface representation of the C2 domain reveals the presence of two negatively charged pockets at these positions where the Ca^{2+} cations are accommodated (**Figure 5.9**). Remarkably, the defined sites partially overlap with the HECT domain contacting surface on the C2 domain, specially regarding the $\beta 1$ - $\beta 2$ loop.

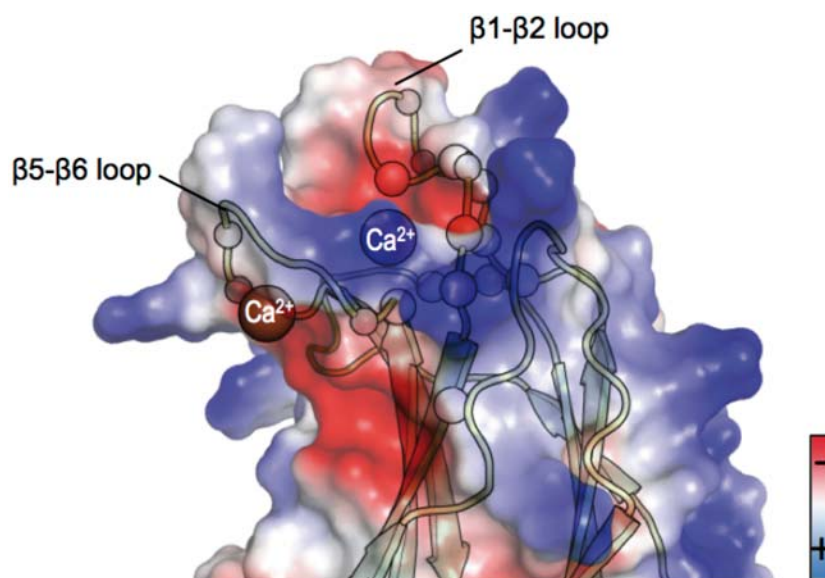


Figure 5.9: Electrostatic surface representation of the C2 domain displaying two Ca^{2+} coordination pockets. **Figure 5.8C** is reproduced including the electrostatic surface representation of the C2 domain. The $\beta 1$ - $\beta 2$ and $\beta 5$ - $\beta 6$ loops form two negatively charged pockets where the Ca^{2+} cations are accommodated.

The binding of Ca^{2+} to the C2 domain was further investigated using Isothermal Titration Calorimetry (ITC) (**Figure 5.10**). The experiment consisted in the progressive addition of a 1.17 mM Ca^{2+} solution to a 50 μM C2 domain sample. Of note, no EGTA was present in the buffer where the Ca^{2+} and the C2 domain samples were prepared (which consisted of 20 mM Tris-HCl, 250 mM NaCl at pH7) to avoid its interference with the assay. The data is consistent with the presence of two Ca^{2+} binding sites on the domain, and the multiple sites model was required to fit an adequate curve. One of the binding sites binds Ca^{2+} with a nanomolar affinity ($K_{D,1} = 26.5 \pm 10.9$ nM). This binding is endothermic, as indicated by positive ΔH and ΔS values. The affinity of the second binding site is in the low micromolar range ($K_{D,2} = 1.7 \pm 0.2$ μM). In this case, the binding is exothermic as ΔH and ΔS have negative values. Due to the difference in two orders of magnitude in the binding affinities, the curve displays a fast saturation of the first binding site. After that, a stationary phase is reached where the second site is slowly occupied until the amount of added Ca^{2+} is enough to eventually start to saturate it in a progressive manner. The experimental setup required notable optimization in order to saturate both sites in the same assay while still acquiring sufficient data points permitting an appropriate curve fitting.

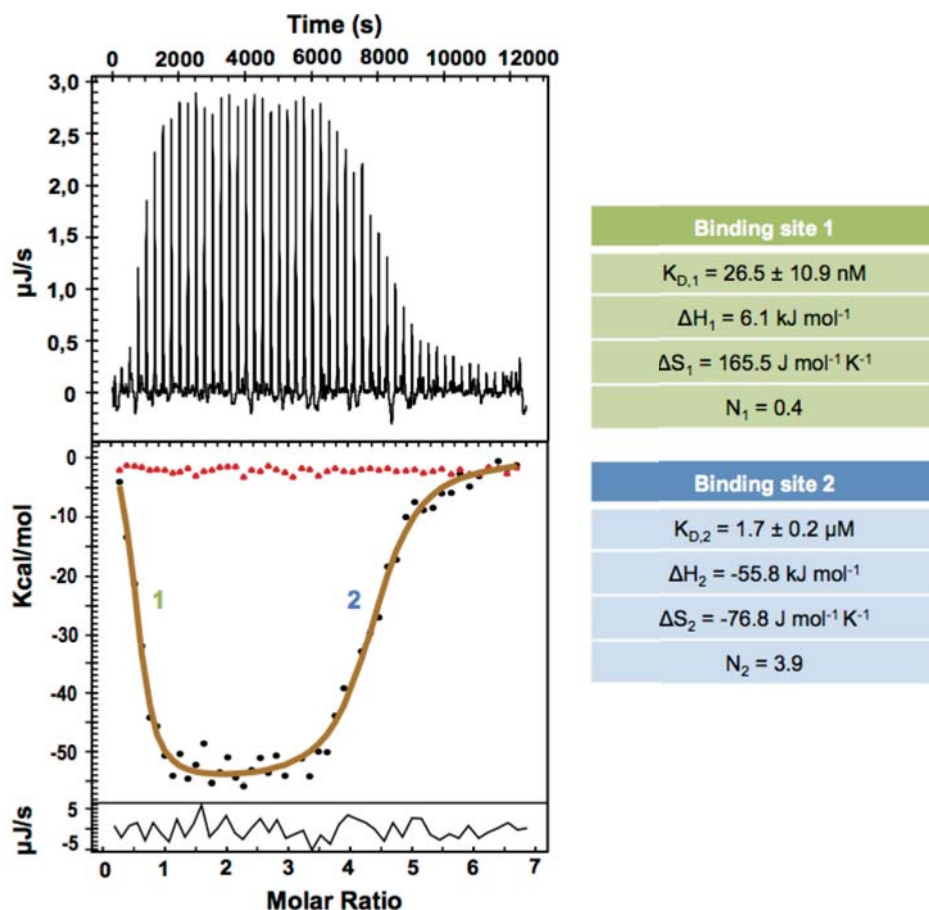


Figure 5.10: Ca²⁺ binding to the C2 domain as monitored by ITC. In the top panel, the raw measured differential heat per injection is displayed in function of time. Below, the integrated peak areas (black dots) and the calculated residuals are shown in function of the ligand (Ca²⁺) molar ratio. Red dots correspond to the blank experiment performed with no C2 domain in the experimental cell. The fitted binding isotherm is displayed in brown. The derived thermodynamic parameters for both binding sites are shown next to the data.

5.2.5. Ca²⁺ hinders HECT binding on the C2 surface

The HECT domain and Ca²⁺ ions use overlapping binding sites to interact with the HECT domain. Thus, we next performed an experiment to directly test the effect that Ca²⁺ has on the HECT-C2 contacts reported in Section 5.2.3. According to the available biochemical data, we expected them not to be established when Ca²⁺ was present at saturating concentrations in the buffer.

In this experiment, we used the ¹H-¹⁵N labelled C2 domain sample equilibrated in a buffer containing 20 mM CaCl₂. We prepared a new unlabelled sample of the HECT domain equilibrated in the same buffer.

Opposite to what we observed when using a Ca²⁺-free buffer, the addition of 1 equivalent of the HECT domain to the C2 sample did not cause the disappearance of

most of its resonances. Instead, about 98% of the resonances were still visible. Interestingly, the missing peaks corresponded to Lys35 and Ala40, both of them located in the $\beta 1$ - $\beta 2$ loop, which is shared in both the HECT domain and the Ca^{2+} ions binding sites.

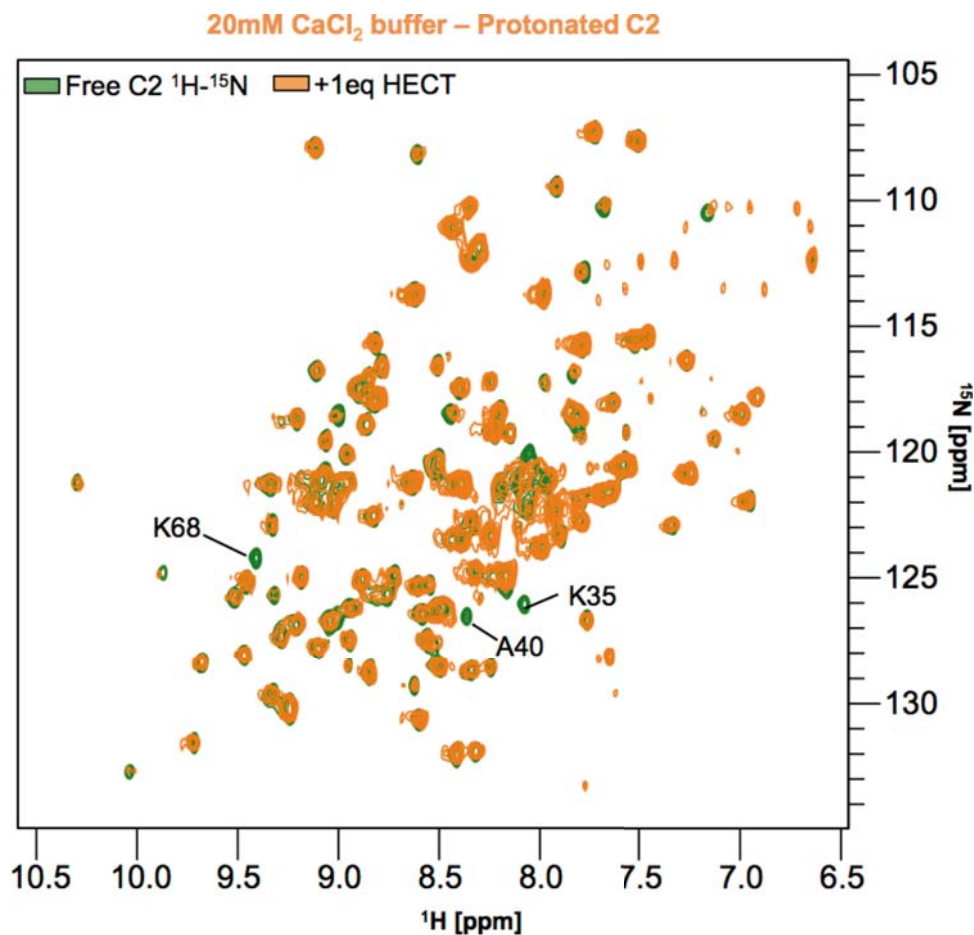


Figure 5.11: The C2-HECT complex is not established in Ca^{2+} -saturating conditions. HSQC titration consisting in the addition of 1 equivalent of HECT domain to a ^1H - ^{15}N labelled C2 domain sample, analogous to the one shown in **Figure 5.6** but with added 20 mM CaCl_2 in the buffer. In green, spectrum of the free, Ca^{2+} -bound C2 domain. In orange, spectrum of the same C2 domain sample after the addition of 1 equivalent of the HECT domain. Opposite to the case where Ca^{2+} is absent, 98% of the C2 domain signals are still present after the addition of the HECT domain, which indicates that the complex between the two protein domains is not established.

Indeed, the outcome of the experiment fits our interpretation of the experiments in Section 5.2.3 and supports our argumentation. In that case, when Ca^{2+} was not present in the buffer, we expected the establishment of a complex with high molecular weight relative to NMR standards. Due to unfavourable relaxation properties this yielded a C2 domain HSQC spectrum with most of the resonances missing. Instead, when Ca^{2+} was present in the buffer the C2 domain kept its favourable properties after the addition of the HECT domain, and we interpreted this

as an evidence that the big complex between the two proteins was not established any longer.

For comparability with experiments in Section 5.2.3, the titration was repeated adding 4 equivalents of the unlabelled HECT domain to a freshly prepared ^2H - ^{15}N C2 domain sample, both equilibrated in the Ca^{2+} containing buffer. After that, we observed small chemical shift perturbations of the resonances corresponding to residues Leu33, Lys 68 and Asp95, as well as the loss of the Gly39 resonance.

From these observations, we concluded that in the presence of a Ca^{2+} concentration able to saturate the interaction, the HECT domain still contacts its binding surface on the C2 domain but this interaction is not stable. Therefore, in this situation the ability of the C2 and the HECT domain to interact with each other is highly perturbed compared to Ca^{2+} -free conditions, ultimately resulting in the activation of the ligase.

5.2.6. Ca^{2+} enhances IP_3 binding to the C2 domain of Nedd4L

Localization experiments using confocal microscopy have revealed that the C2 domain of Nedd4L mediates the relocation of the ligase from the cytoplasm to the plasma membrane upon the increase of intracellular Ca^{2+} levels (Garrone et al., 2009). A polymorphism in the human gene codifying Nedd4L yields two isoforms containing or lacking the C2 domain. While the complete protein is able to relocate to the plasma membrane and target its substrate β -ENaC in response to an increase of the intracellular Ca^{2+} levels, the incomplete version is not.

Phosphatidylinositol-4,5-bisphosphate (PIP_2) is a membrane phospholipid that has been described to interact with C2 and also with PH domains (Corbalan-Garcia et al., 2003; Hyvonen et al., 1995). These domains have positively charged patches located on their surfaces that they use to interact with the inositol head of PIP_2 , which is exposed to the cytoplasm. In response to ligand activation of a G protein-coupled receptor (GPCR) or a receptor tyrosine kinase (RTK), the phospholipase C (PLC) hydrolyses the inositol head of PIP_2 , resulting in the cytoplasmic delivery of the second messenger molecule inositol 1,4,5-trisphosphate (IP_3) (Berridge, 2009). This results in the mobilization of intracellular Ca^{2+} reservoirs through the binding of IP_3 to its receptor (IP_3R) in ligand-gated Ca^{2+} channels located on the surface of the endoplasmic reticulum. The C2 and also the PH domains are also able to interact with the soluble IP_3 as it conserves the inositol head of PIP_2 .

We envisioned a dual role for IP₃ in the biology of Nedd4L: on the one side, as a part of PIP₂ it may contribute to the relocation of the ligase from the cytoplasm to the plasma membrane through its interaction with the C2 domain. On the other hand, its cytoplasmic delivery resulting in an increase of the intracellular Ca²⁺ levels constitutes a very likely trigger for the release of the HECT-C2 contacts reported in Section 5.2.3 and the subsequent activation of the ligase.

We used 1D ³¹P and 2D ¹H-¹⁵N HSQC titrations to investigate whether the C2 domain of Nedd4L binds IP₃ and to assess the role of Ca²⁺ in this interaction. First, we added progressively increasing amounts of IP₃ to a ¹H-¹⁵N labelled C2 domain sample equilibrated in a Ca²⁺-free buffer and followed both the ³¹P resonances of IP₃ as well as the protein amide resonances. After the addition of up to 30 equivalents of IP₃, two of its resonances were slightly affected. Almost no perturbation of the protein resonances was detected. Thus, we interpreted that almost no interaction occurs between IP₃ and the C2 domain in the absence of Ca²⁺.

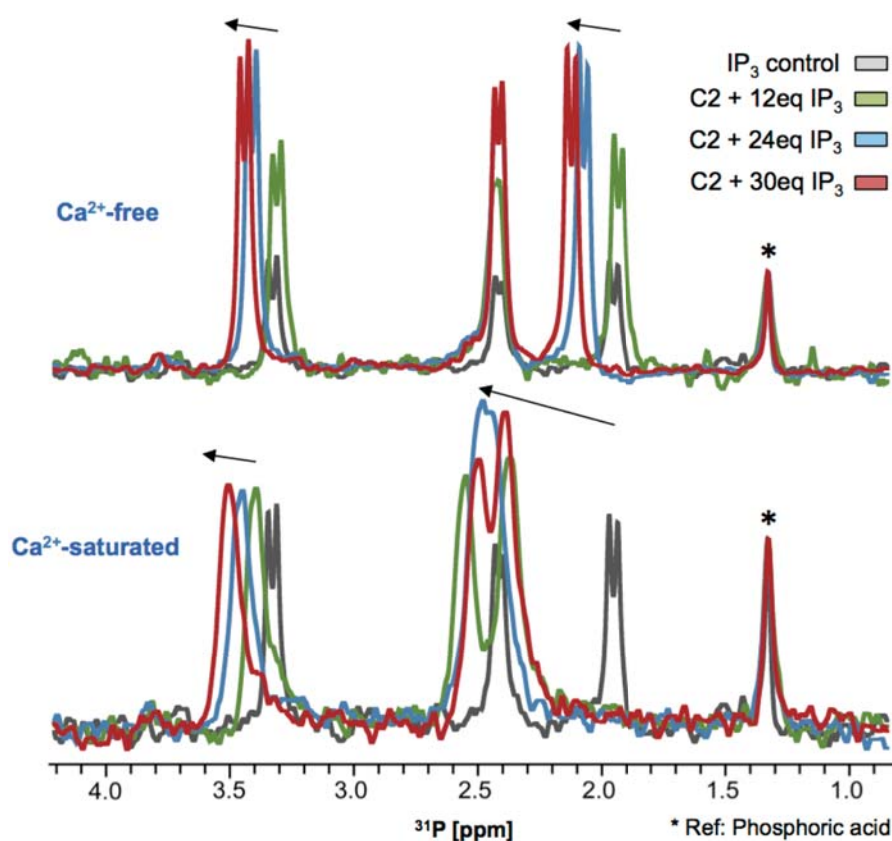


Figure 5.12: ³¹P titration monitoring the IP₃ resonances when binding to the C2 domain.

Top panel: titration performed using a C2 sample equilibrated in a Ca²⁺-free buffer. Bottom panel: same experiment repeated using a C2 sample equilibrated in a buffer containing 20 mM CaCl₂. Phosphoric acid was added to all samples as a reference. Displayed in grey, ³¹P spectrum of IP₃. Shown in green, blue and red are the ³¹P spectra of a C2 sample with added 12, 24 and 30 equivalents of IP₃, respectively. When Ca²⁺ is present in the buffer, the IP₃ resonance around 2 ppm experiences a sensibly higher perturbation in its chemical shift than when Ca²⁺ is absent.

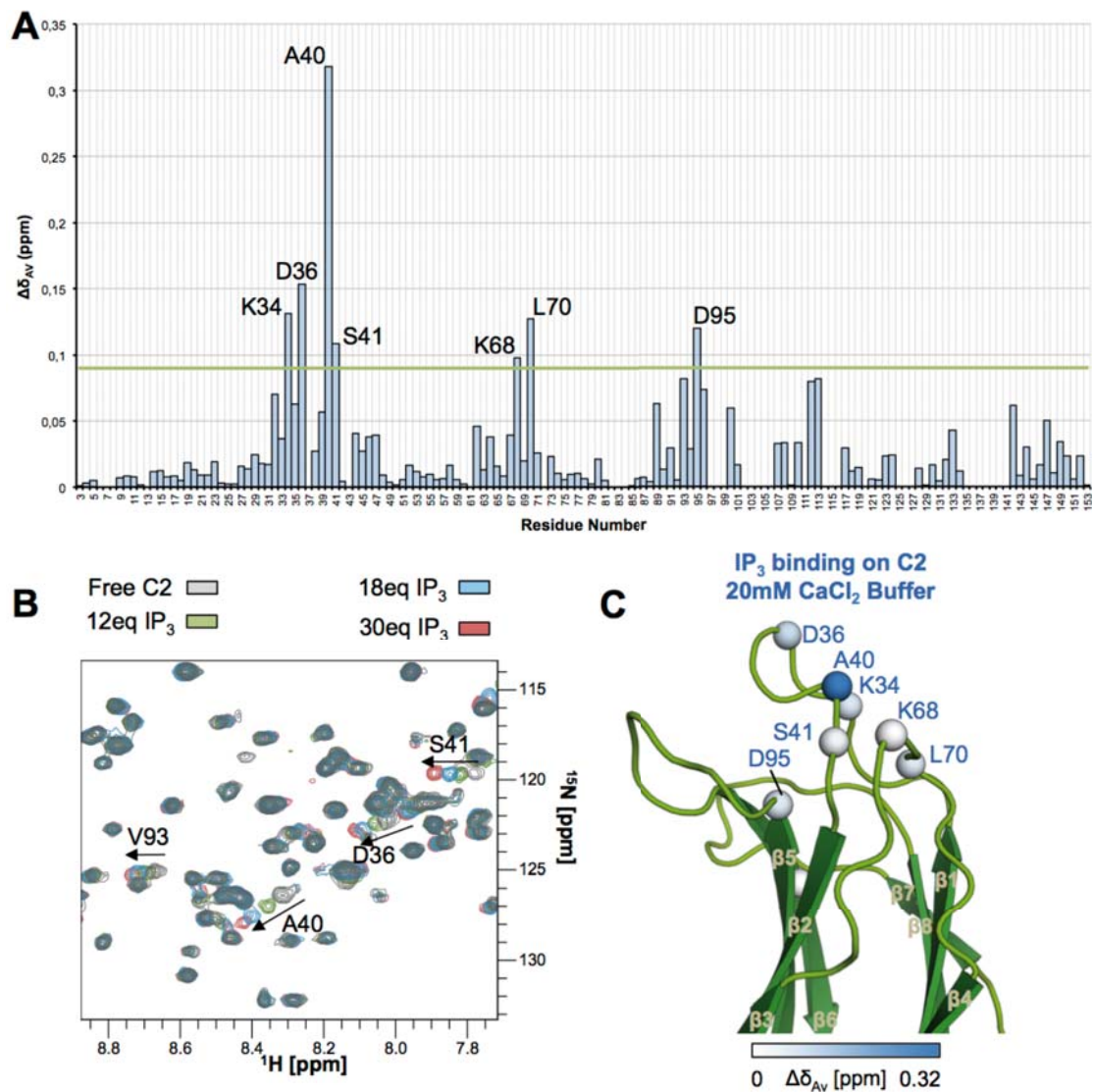


Figure 5.13: IP₃ binding to the C2 domain in Ca²⁺- saturating conditions. HSQC titration consisting in the addition of progressively increasing amounts (up to 30 equivalents) of IP₃ to a ¹H-¹⁵N labelled C2 sample prepared in a buffer containing 20 mM CaCl₂. (A) Histogram representation of the isotope weighted chemical shift perturbations ($\Delta\delta_{AV}$, ppm) per each individual residue in the C2 domain. (B) Close-up view of the overlaid HSQC spectra acquired during the titration. The spectrum of the free, Ca²⁺-bound C2 domain is shown in grey. The spectra resulting from the addition of 12, 18 and 30 equivalents of IP₃ are shown in green, blue and red, respectively. (C) Cartoon representation of the IP₃ binding site on the surface of the C2 domain. Residues with $\Delta\delta_{AV}$ values above 0.09 ppm (green line in (A)) are displayed as spheres, and coloured in a blue gradient according to the magnitude of their respective chemical shift perturbations.

Next, we repeated the titration using a fresh ¹H-¹⁵N labelled C2 domain sample equilibrated in a buffer containing 20 mM CaCl₂. In contrast to our previous experiment, one of the ³¹P resonances in IP₃ presented a major change in its chemical shift already after the addition of 12 equivalents of IP₃ (**Figure 5.12**). Major chemical shift perturbations were also observed in several protein resonances, with values up to 0.32 ppm ($\Delta\delta_{AV}$) after the addition of 30 equivalents of IP₃ reporting a stable interaction (**Figure 5.13A and B**).

This titration allowed us to map the IP₃ binding site on the surface of the C2 domain of Nedd4L. Again, many of the most affected residues were located in the β 1- β 2 loop, including Lys34, Asp36, Ala40 and S41. Lys68 and Leu70 in the β 3- β 4 loop, as well as Val93 and Asp95 in the β 5 strand were significantly affected too (**Figure 5.13C**).

The described binding sites for IP₃ and Ca²⁺ ions on the surface of the C2 domain overlap extensively. This is in agreement with the previously reported lipid binding mode of C2 domains where the Ca²⁺ ions bind to a negatively charged patch of the protein (**Figure 5.9**) switching its surface electrostatics and bridging the interaction with the negatively charged IP₃ (Cho and Stahelin, 2006). The C2 domain of PKC α has been described to bind phosphatidylserine in this mode, while it binds IP₃ using the cationic β -groove (a second lipid binding patch conserved in a variety of C2 domains, which is basic and does not bind Ca²⁺) (Guerrero-Valero et al., 2009). The IP₃ binding capabilities of the C2 domain in Smurf2 have been shown to be Ca²⁺-independent (Wiesner et al., 2007).

In addition, the IP₃ and the HECT domain binding sites on the C2 domain overlap too. Therefore, our observations suggest that an increase in the intracellular Ca²⁺ levels switches the binding preferences of the C2 domain in Nedd4L from interacting with the HECT domain to binding the inositol moieties attached to the membrane (PIP₂) and free in the cytosol (IP₃).

5.3. Summary

Prior to our study, data regarding the activation of Nedd4L by changing from a latent to an active conformation had been obtained using molecular biology-based methods and was available in the literature. *In vivo* Co-IP experiments had shown that the HECT and the C2 domains of the protein interact. Substrate and auto-ubiquitination assays revealed that the ligase is inactive in this situation (Wang et al., 2010). In a similar manner, it was shown that this interaction is impaired when intracellular Ca²⁺ levels rise and this increase results in the activation of the ligase. In addition, confocal microscopy results revealed that the C2 domain of the ligase plays a key role at localizing it to the plasma membrane where several of its substrates are located (Garrone et al., 2009).

The results reported in this chapter support these observations and provide structural information at the atomic level explaining the activation mechanism of the ligase. In addition, we propose that IP₃ is likely to play a key role in this mechanism.

First, we have optimized our NMR methods to characterize the HECT-C2 contacts in detail. We provide evidence supporting that the complex is established *in vitro* in Ca^{2+} -free conditions when both domains are produced as separate entities. We have been able to map the binding site of the HECT domain on the surface of the C2 domain.

Next, we have characterized the binding site of Ca^{2+} ions on the C2 domain too, and we have observed that it partially overlaps with the surface used by the HECT domain. We have found that Ca^{2+} ions tightly bind the C2 domain in two different acidic binding pockets, with affinities in the nanomolar and low micromolar ranges respectively. Given that the previously observed HECT-C2 complex is not established any longer when Ca^{2+} is present in our NMR buffer, we propose that Ca^{2+} impairs HECT contacts by displacing it from its binding site on the C2 domain, either through steric impedance or by an electrostatic switch.

The activation of different membrane receptors including GPCR and RTK leads to the cytoplasmic delivery of IP_3 , which is a central event in Ca^{2+} signalling resulting in the increase of intracellular Ca^{2+} levels. Therefore, we suggest that this is a likely trigger for the activation of Nedd4L.

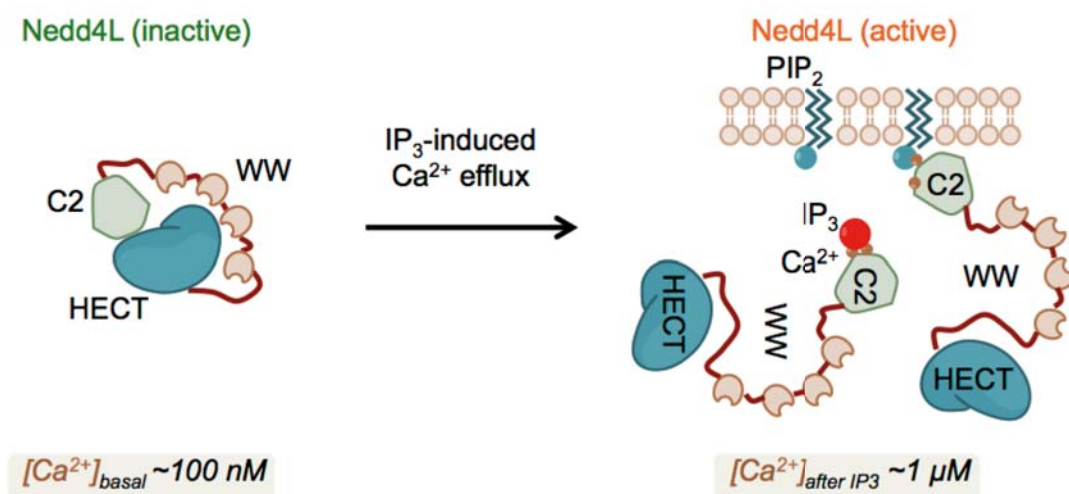


Figure 5.14: Proposed model for the activation of Nedd4L. At the left panel, Nedd4L is shown in its closed and inhibited conformation. Cytoplasmic release of IP_3 by PLC cleavage of PIP_2 results in a Ca^{2+} efflux to the cytoplasm, and Ca^{2+} concentration increases from about 100 nM to about 1 μM . This event triggers the transition from the closed to the open and active conformation of the ligase, which is now able to recognize and ubiquitinate its substrates.

We have observed a strong binding of IP_3 to the C2 domain of Nedd4L only when Ca^{2+} is present in solution. The binding site used by IP_3 overlaps with the surface used by both Ca^{2+} ions and the HECT domain to bind the C2 domain. We propose that Ca^{2+} switches the surface electrostatics in the loop area of the C2 domain from

negative to positive, thus bridging the interaction with IP₃. Therefore, upon Ca²⁺ coordination the C2 domain switches its binding preferences from the HECT domain to the IP₃ moieties. Cytosolic IP₃ molecules may contribute displacing the HECT domain from its binding site on the C2 domain, whereas the membrane phospholipid PIP₂ is a likely anchoring point for the ligase to the plasma membrane, given that its inositol phosphate head group is chemically identical to IP₃ and it retains its capabilities to interact with C2 domains.

As summarized in Figure 5.14, our results support the existence of a closed and inactive conformation of Nedd4L mediated by HECT-C2 contacts. IP₃ and the subsequent Ca²⁺ delivery into the cytoplasm trigger the transition to the open and active form of the ligase.

CHAPTER 6. Fully functional Nedd4L molecules target partially unstructured counterparts as a means to regulate the turnover rate of the ligase[‡]

6.1. Introduction

A relevant issue related to the regulation of the components of the protein ubiquitination system refers to how their turnover rates are regulated, especially in the case of the E3 ligases, which are the final effectors of the ubiquitination cascade.

Many E3 ligases have been shown to label themselves with ubiquitin in a process known as auto-ubiquitination. Although this event can label the ligase for several different fates, the most common one is proteasomal degradation (De Bie and Ciechanover, 2011; Lorick et al., 1999). Auto-ubiquitination may occur both in a substrate dependent (Ryan et al., 2006) or independent manner (Okamoto et al., 2009), and in some cases substrate binding may protect the ligase from self-degradation (Petroski and Deshaies, 2005).

The information currently available regarding the auto-ubiquitination mechanisms of the E3 ligases is limited. Indeed, the reaction has been observed to be intra-molecular in some cases (one molecule of the ligase labels itself) (Galan and Peter, 1999; Huibregtse et al., 1995) and inter-molecular in some others (a bi-molecular crossed reaction occurs where one molecule of ligase labels one of its counterparts) (Hu and Fearon, 1999; Nuber et al., 1998).

Regarding the HECT-type ligases belonging to the Nedd4 family, auto-ubiquitination has been observed in Itch, Nedd4, Nedd4L, Smurf1, Smurf2, WWP1 and WWP2 (Mund and Pelham, 2009). In several cases, the event is prevented by inter-domain contacts found in the latent conformations of these ligases (such as those described in CHAPTER 5 for Nedd4L). The transition to the active form is thus required for these ligases to auto-ubiquitinate. Sensible differences may exist in the mechanisms used: the reaction was shown to be exclusively intra-molecular in the case of Rsp5 (the only member in *S. cerevisiae*) but not in the case of p100 (a family member in rat) (Huibregtse et al., 1995).

[‡] Escobedo, A., Gomes, T., Aragón, E., Martín-Malpartida, P., Ruiz, L., Macias, M.J. (2014). *Structural basis of the activation and degradation mechanisms of the E3 ubiquitin ligase Nedd4L*. **Structure** 22, 1446-1457.

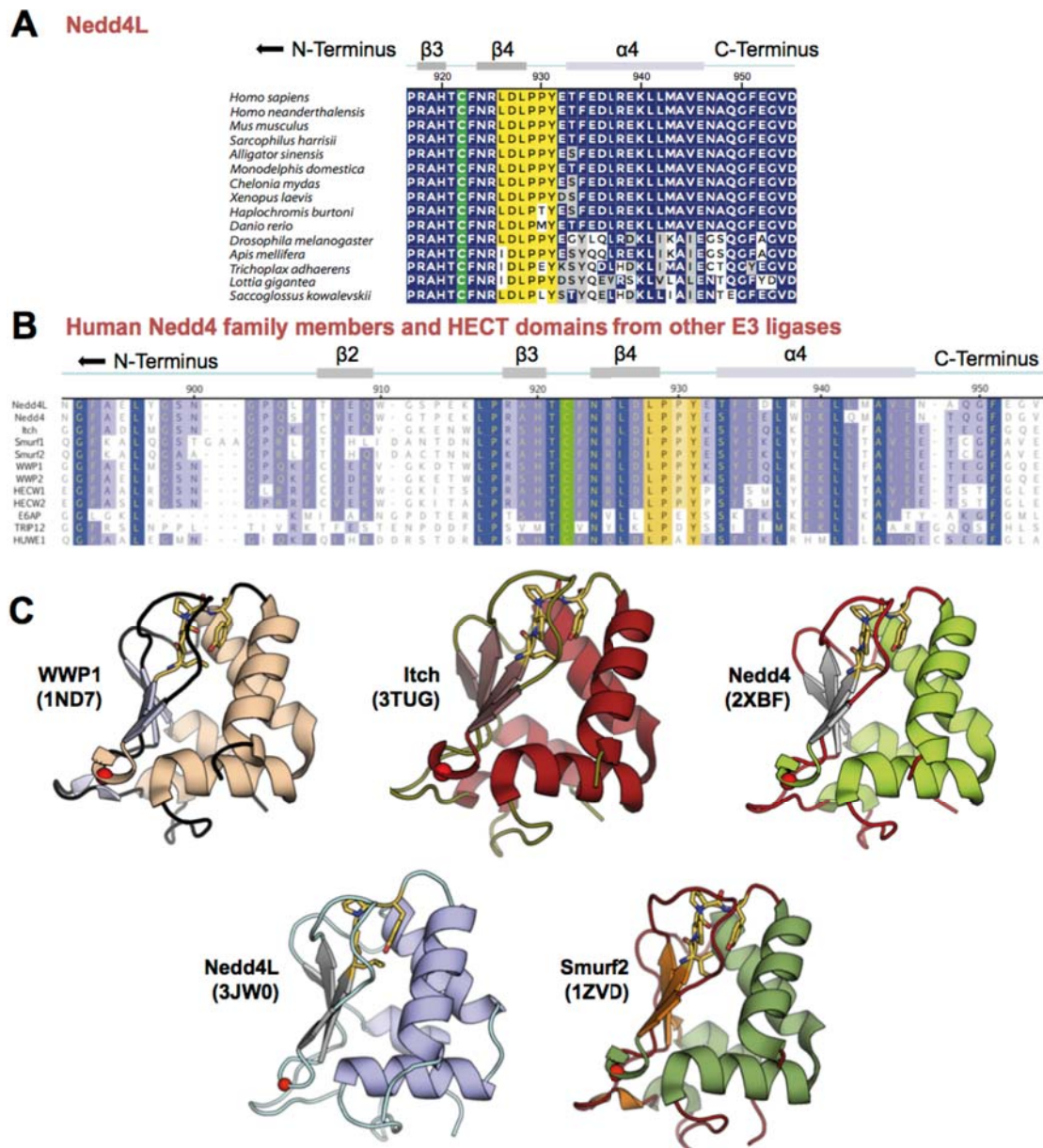


Figure 6.1: Conservation of the HECT PY motif. (A) Sequence alignment showing the conservation of the LPxY motif (in yellow) throughout evolution in Nedd4L. The strictly conserved catalytic cysteine residue is displayed in green. The 917 – 954 segment from 15 different metazoans is shown, including two hominids (sapiens and neanderthalensis), several mammals, a marsupial, an amphibian, several fishes and a reptile. (B) Sequence alignment illustrating the high degree of conservation of the PY motif in the HECT domains of Nedd4 family members and the HECT domains of other E3 ligases. The H920 – Y931 segment (Nedd4L, isoform 5 numbering) has 81.9% similarity that is raised to 93.5% for members of the Nedd4 family only (top 9 rows). (C) All available crystal structures of HECT C-terminal lobes belonging to members of the Nedd4 family. The catalytic cysteine is shown as a red sphere. The conserved PY motifs are shown in sticks representation and highlighted in yellow. In all cases, they are part of the hydrophobic core, apparently not available for WW recognition.

The HECT domain of Nedd4L presents a PY motif (LPxY) highly conserved throughout metazoan evolution located in the HECT C-terminal lobe, next to the strictly conserved catalytic cysteine residue (just 5 residues apart) (**Figure 6.1A** and

Section 9.4), which is recognized by WW domains despite being slightly different from the canonical PPxY sequence (Section 1.1.5.1) (Bruce et al., 2008). The motif is indeed conserved in all members of the Nedd4 family in humans and in the HECT domains of other E3 ligases (**Figure 6.1B**).

In general, Nedd4 family members recognize their substrates using the WW domains in their sequence to bind such motifs. The strict conservation of this HECT PY motif suggests that it may play a role in their biology, probably related to self-recognition and ubiquitination. Nonetheless, although part of the motif is located in an exposed loop, the tyrosine residue that is typically fundamental for WW recognition appears to be buried and not accessible in all the available crystal structures of the HECT domains of Nedd4 family members, including Nedd4 (PDB entry 2XBF), Nedd4L (3JW0), Itch (3TUG), Smurf2 (1ZVD) and WWP1 (1ND7) (**Figure 6.1C**).

Accordingly, the interaction between the WW domains of Nedd4L and the HECT PY motif has not been detected by molecular biology methods when the ligase is natively folded. Instead, the binding requires the prior denaturation of the HECT domain (Bruce et al., 2008). In addition, the ubiquitination of its substrate β -ENaC enhances Nedd4L auto-ubiquitination.

We focused our efforts on understanding the turnover rate regulation of Nedd4L by auto-ubiquitination. We hypothesized that under certain circumstances the WW domains of the ligase recognize its HECT PY motif in a manner similar to substrate recognition.

Using NMR, we found that the binding of the Nedd4L WW domains to the HECT PY motif is hampered when the HECT domain is natively folded. Relaxation measurements acquired on the folded HECT C-terminal lobe display low flexibility of the area where the PY motif is located, suggesting that its spontaneous exposure is unlikely and it needs to be triggered. *In silico* temperature-induced destabilization of the protein shows that the motif is exposed while the overall fold of the protein is maintained, suggesting that a naturally occurring unfolding process may include this exposure in its initial steps. When the motif is exposed to the solvent it is recognized by WW3 in a substrate-like manner, as supported by our ITC measurements and the NMR structure of the WW3-HECT PY complex that we solved. From these results we propose a bi-molecular crossed ubiquitination mechanism for Nedd4L.

6.2. Ubiquitination of the ubiquitin ligase: a bi-molecular reaction

6.2.1. Preparation of the WW domains and the HECT C-lobe protein samples and synthesis of the HECT-PY peptide

Due to the strict conservation of the HECT PY motif and the particular relevance of Nedd4L for our research interests, we set to investigate the potential recognition of this motif by the ligase's WW domains. All prepared samples are summarized in **Table 6.1**.

Molecule	MW ^a (kDa)	Cloning ^b	Samples	Production
HECT C-lobe	13.3	C922D mutant • His ₆ tag	• ¹⁵ N → CSP assays • ² H ¹³ C ¹⁵ N → bb. resonance assign.	• Expression in inclusion bodies • Fast refolding bound to Ni ²⁺
WW3	5.7	• Boundaries optimization • His ₆ tag • P489C and T500C mutants • His ₆ tag	• Unlabelled → CSP assays, 2D NOESY, ITC measurements • ¹⁵ N → CSP assays, 3D HSQC-NOESY • ² H ¹⁵ N → CSP assays • ² H ¹³ C ¹⁵ N → bb. resonance assign. • MTSL/MTS → PRE assays (mutants)	• Expression in inclusion bodies • Fast refolding bound to Ni ²⁺ • Slow refolding by dialysis
WW1	4.7	• His ₆ tag	• Unlabelled → CSP assays, ITC	• Expression in inclusion bodies
WW2	5.0			
WW4	4.6		measurements	• Fast refolding bound to Ni ²⁺
HECT-PY	2.1	-	• Unlabelled → CSP assays, ITC, structure resolution	• SPPS

Table 6.1: Summary of the prepared HECT C-lobe, WW domains and HECT-PY samples. (a) Molecular weight values as calculated for unlabelled samples. (b) The green dot next to the solubility tag indicates that the resulting constructs were soluble and obtained in good amounts.

Firstly, we tested the interaction in the context of natively folded domains (Section 6.2.2). Both the catalytic cysteine residue and the PY motif are located in the HECT C-terminal lobe. The preparation of soluble, stable recombinant samples of this subdomain is reported in detail in Section 5.2.1.

Each one of the four WW domains in Nedd4L was cloned in fusion with a His₆ tag in order to enable them for metal affinity purification. A TEV (Tobacco Etch Virus) protease cleavage site was included between the tag and the target protein.

Because NOESY experiments allowing structural resolution by NMR require protein samples at high concentration, the boundaries of the recombinant Nedd4L WW3 construct were optimized to increase the solubility of the protein.

All domains were expressed overnight at 37°C with 500 μM IPTG induction. Under these conditions, we obtained high expression yields in inclusion bodies. Isotope labelling was achieved as described (Marley et al., 2001) using the same expression conditions. The overexpressed proteins were recovered from the inclusion bodies by denaturing them using a 6 M guanidine hydrochloride buffer for cell lysis. Next, they were purified while denatured using a Ni²⁺ affinity resin. Before eluting them from the resin, they were refolded by thoroughly washing them with a denaturant-free buffer. The proteins were then eluted, the His₆ tag was removed by TEV protease digestion and a final size-exclusion chromatography purification step was performed.

To test the interaction of the WW domains with the unstructured HECT domain we prepared a synthetic short peptide including the PY motif (Arg925 – Lys 940) using solid-phase peptide synthesis (SPPS, CHAPTER 4). The peptide was synthesized using Fmoc/^tBu chemistry on the ChemMatrix[®] resin. The automatic microwave synthesis strategy was used at the 0.25 mmol scale. For residue coupling, we used 5 equivalents of the amino acids Fmoc derivatives activated with HATU/HOAt and DIEA in DMF. Fmoc deprotection was achieved using 20% piperidine in DMF. After completion of the sequence synthesis, the peptide was cleaved from the solid support using a 95:2.5:2.5 mixture of TFA, H₂O and TIS. After TFA evaporation, the cleaved crude was dissolved in 5% acetonitrile in H₂O and purified by reversed-phase chromatography. The target peptide was obtained with 98.97% purity and detected as the singly, doubly and triply charged ions by ESI-MS (*m/z* calculated: 2062.28, 1031.64 and 688.08; *m/z* observed: 2062.66, 1032.15 and 688.71; Section 9.1).

6.2.2. The HECT-PY motif is not accessible for WW domain binding in the context of the folded HECT C-lobe

Despite the tyrosine residue in the PY motif being packed inside the hydrophobic core of the HECT domain in all the available crystal structures of ligases belonging to the Nedd4 family, we wondered whether it could be eventually exposed in solution. This would allow the binding of WW domains to this region in the natively folded state of the HECT domain.

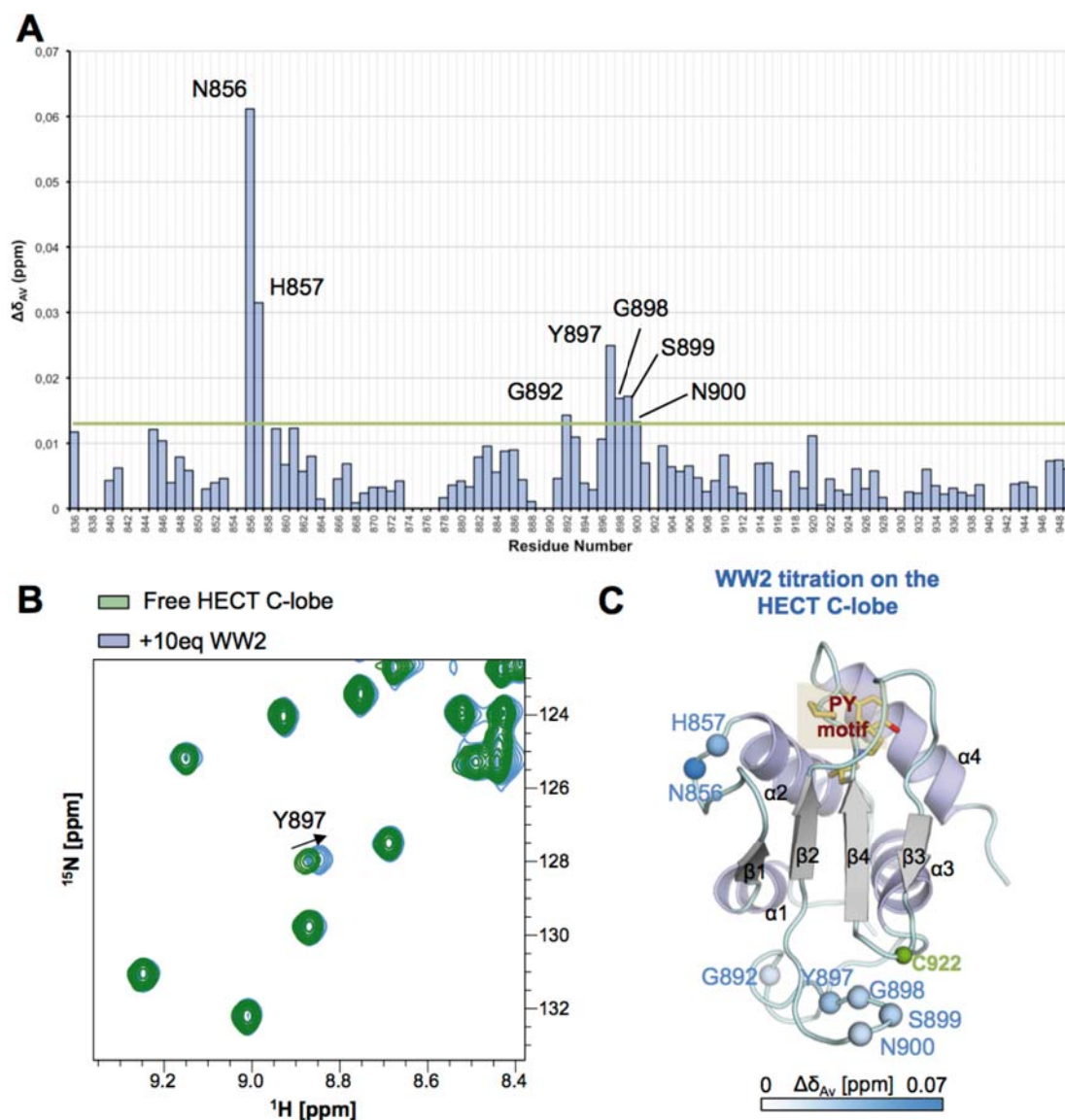


Figure 6.2: The WW2 domain does not bind the PY motif in the natively folded HECT C-lobe. Data resulting from the HSQC titration of progressively increasing amounts (up to 10 equivalents) of unlabelled WW2 domain on a ^1H - ^{15}N labelled HECT C-lobe sample. (A) Histogram representation of the chemical shift perturbations observed per residue ($\Delta\delta_{AV}$, ppm). Values for the residues located in and close to the PY motif indicate no binding in this area. All other values are small, reporting no stable binding. (B) Close-up view of two of the HSQC spectra resulting from the titration, illustrating the small CSPs observed. Shown in green is the spectrum of the free HECT C-lobe. The blue spectrum corresponds to the same sample after the addition of 10 equivalents of WW2 domain. (C) Representation of the affected residues on the X-Ray structure of the HECT C-lobe (3JW0), shown as spheres in a blue gradient. The PY motif is highlighted in yellow and the catalytic residue is shown as a green sphere.

To test that, we performed HSQC titrations consisting on the progressive addition of an unlabelled WW domain to a ^1H - ^{15}N labelled sample of the HECT C-lobe. We used the assignment of the backbone resonances of the HECT C-lobe presented in Section 5.2.2 to identify the resonances in the HSQC spectra of the domain. We repeated this experiment for the WW2 and the WW3 domains separately.

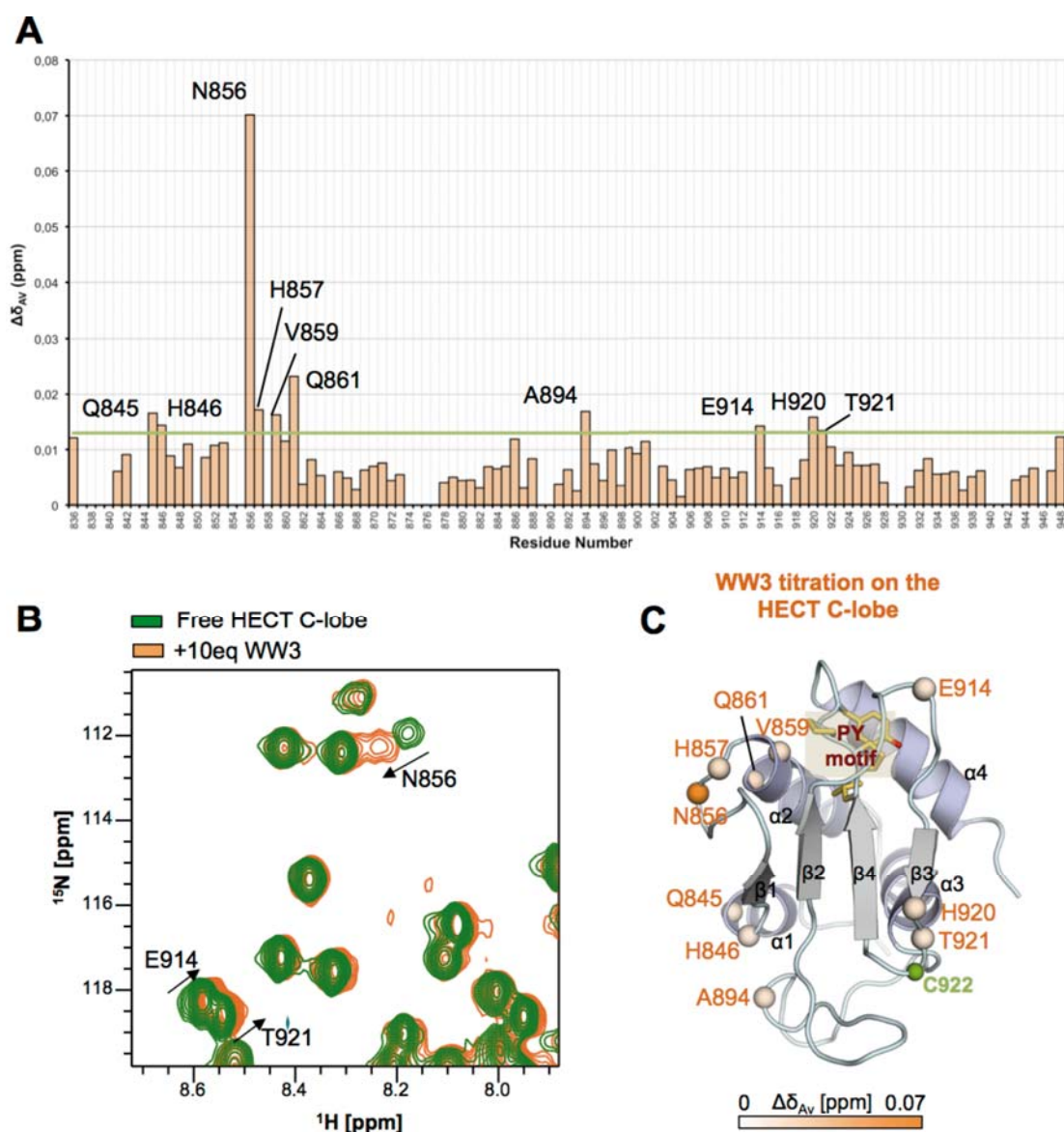


Figure 6.3: The WW3 domain does not bind the PY motif in the natively folded HECT C-lobe. The same experiment shown in **Figure 6.2** was repeated using WW3 instead of WW2. (A) Histogram representation of the observed chemical shift perturbations ($\Delta\delta_{AV}$, ppm). Again, residues in or next to the PY motif are unaffected. Low CSP values report no stable binding (B) Close-up view of the HSQC spectra of the free HECT C-lobe (green) and the same sample after the addition of 10 equivalents of WW3 (orange), illustrating the observed chemical shift perturbations. (C) Representation of the values shown in (A) on the crystal structure of the HECT C-lobe (3JW0). The affected residues are sparsely distributed and do not define a clear binding site. They are displayed as spheres coloured in an orange gradient depending on their value. Residues with $\Delta\delta_{AV} > 0.013$ ppm are represented (green line in (A)).

In both cases, after the addition of up to 10 equivalents of the WW domain just a few solvent-exposed residues displayed small perturbations ($\Delta\delta_{AV}$ values below 0.07 ppm), and these residues did not define clear binding sites. WW2 is the domain that recognizes the PY motifs on the substrate proteins Smad2/3 and 7. When titrating it on the folded HECT C-lobe, residues located in the $\beta 1$ - $\alpha 2$ and the $\alpha 3$ - $\beta 2$ loops were

slightly affected (**Figure 6.2**). As discussed below, the $\alpha 3$ - $\beta 2$ loop is the most flexible and exposed area of the domain.

In the case of WW3 the affected residues were more sparsely distributed around the structure: in $\alpha 1$, $\alpha 2$ and the $\beta 1$ - $\alpha 2$, $\alpha 3$ - $\beta 2$, $\beta 2$ - $\beta 3$ and $\beta 3$ - $\beta 4$ loop regions (**Figure 6.3**). Some of them are located in the vicinity of the PY, although none in the motif itself (the proline residues can't be followed in HSQC titrations due to their lack of amide protons).

Taken together, these results suggest that the accessibility to the conserved PY motif is hindered when the HECT domain is natively folded.

6.2.3. Relaxation measurements on the HECT C-lobe reveal packing of the PY motif in the protein core, hampering its spontaneous exposure

Prompted by these observations, we set to evaluate the packing of the HECT-PY motif in the structure of the domain in solution. We performed relaxation measurements to evaluate the dynamic properties of the protein, with the aim of defining its most flexible and exposed areas.

To evaluate local effects including loop motions, we characterized the dynamics of the protein at the fast timescale (ps-ns). Using a ^1H - ^{15}N labelled sample of the HECT C-lobe we acquired ^{15}N T_1 and T_2 values as well as heteronuclear ^1H - ^{15}N NOE measurements. These are HSQC-based experiments that provide information about the dynamic properties of each amino acid in the protein sequence.

Our measurements are in agreement with the protein fold in its crystal structure, suggesting that this fold is similar in solution. Values that are characteristic of compact areas were obtained for residues in secondary structure elements and short loops. These include the $\beta 4$ - $\alpha 4$ loop, where the PY motif is located. The observed T_1 and T_2 values oscillate around 0.7 s and 0.075 s respectively, while heteronuclear NOEs are centred at 0.8.

Residues located in the $\alpha 3$ - $\beta 2$ loop display values associated with higher flexibility and exposure. Many of these residues have heteronuclear NOE values around 0.6, with some of them displaying even lower values down to 0.4. T_1 and T_2 values show higher variability than in other regions of the molecule. NOE measurements for residues in the $\beta 1$ - $\alpha 2$ loop yielded ambiguous data. However, reduced T_1 values in the area suggest increased flexibility of this loop as well.

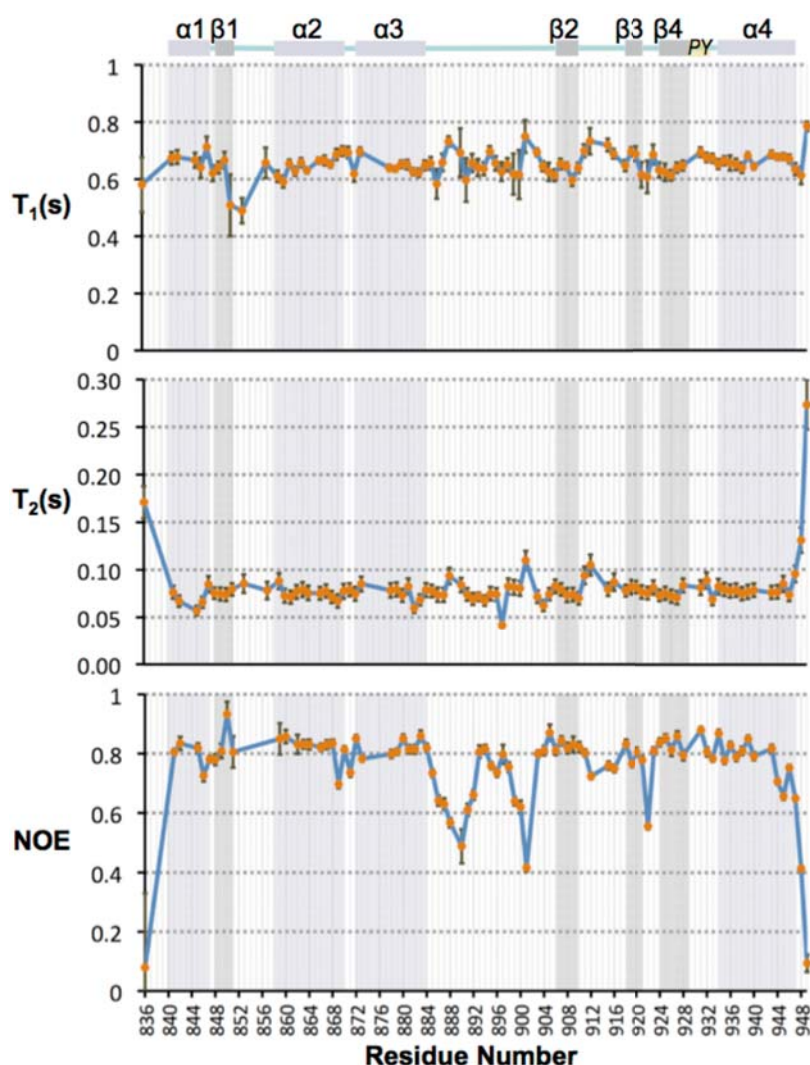


Figure 6.4: T_1 , T_2 and heteronuclear NOE measurements on the HECT C-lobe. Only unambiguously assigned residues were included in the analysis. The secondary structure elements are highlighted in blue (α helices) and grey (β sheets). Most residues show values typically found in compactly folded areas. Residues in the $\beta 1$ - $\alpha 2$ and the $\alpha 3$ - $\beta 2$ loops display values associated with higher flexibility. The $\beta 4$ - $\alpha 4$ loop, where the HECT-PY motif is located, is in a compactly folded area and its spontaneous exposure is hampered.

According to these measurements, the $\beta 1$ - $\alpha 2$ and the $\alpha 3$ - $\beta 2$ loops correspond to the most flexible and exposed areas of the protein. Remarkably, these areas are the most affected regions in the WW titrations presented in Section 6.2.2 (**Figure 6.2** and **Figure 6.3**), especially in the case of WW2. This supports our interpretation that the mild chemical shift perturbations observed do not correspond to the formation of any stable WW-HECT complex, but rather report transient contacts.

In addition, the $\beta 4$ - $\alpha 4$ loop and the areas close to it are highly compact, suggesting that the exposure of the PY motif that would allow WW recognition is hampered and will require a potential structural rearrangement.

6.2.4. MD simulations reveal the exposure of the HECT PY motif at the early stages of protein destabilization

The relaxation experiments provide information limited to events occurring in the fast timescale. The low flexibility observed in the area of the HECT PY motif suggests that its spontaneous exposure is improbable. Still, we considered the possibility of this event occurring at a longer timescale, with slower rates than those that can be observed with our previous measurements. To cover that, we performed a molecular dynamics (MD) simulation consisting in the calculation of a 120 ns trajectory at 310 K. Tiago Gomes performed this simulation in our laboratory.

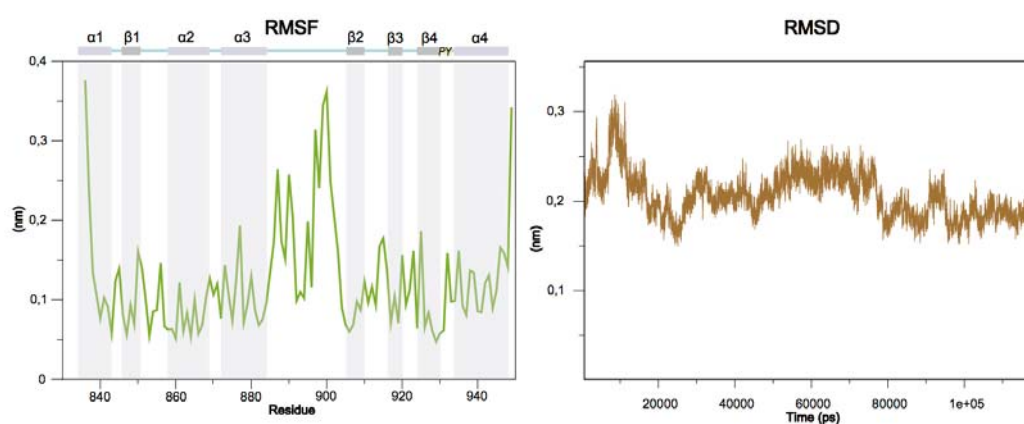


Figure 6.5: Molecular dynamics (MD) simulation of the HECT C-lobe along a 120 ns trajectory at 310 K. *Left panel:* root-mean-square fluctuation (RMSF) values per residue in the sequence in nm. Residues located at the $\alpha 3$ - $\beta 2$ loop show higher values, in agreement with the previously observed flexibility of the area (see **Figure 6.4**, lower panel). All other regions are compactly folded considering the length of the simulation. *Right panel:* root-mean-square deviation (RMSD) of the structure along the simulation. Its low variability centred at 0.2 nm indicates that no major conformational rearrangements occur and further supports the compactness of the structure.

The simulation was performed using the GROMACS package (Hess et al., 2008). The fragment corresponding to the HECT C-lobe construct used in all previous experiments was extracted from the crystal structure of the HECT domain (3JW0), solvated and equilibrated at 310 K. After that, it was simulated over 120 ns with no positional restraints. The resulting root-mean-square fluctuation (RMSF) and root-mean-square deviation (RMSD) values are shown in **Figure 6.5**.

Residues located at the $\alpha 3$ - $\beta 2$ loop show the highest RMSF values, reporting flexibility in this region. All other areas, including the $\beta 4$ - $\alpha 4$ loop where the PY motif is located, display lower RMSF values indicative of compactness and low mobility. The data is in good agreement with our relaxation measurements (compare **Figure 6.5** and **Figure 6.4**). RMSD values fluctuate around 0.2 nm along the simulation,

indicating that no major conformational arrangements occur. Therefore, the exposure of the tyrosine residue in the HECT PY motif is unlikely in the sampled conditions.

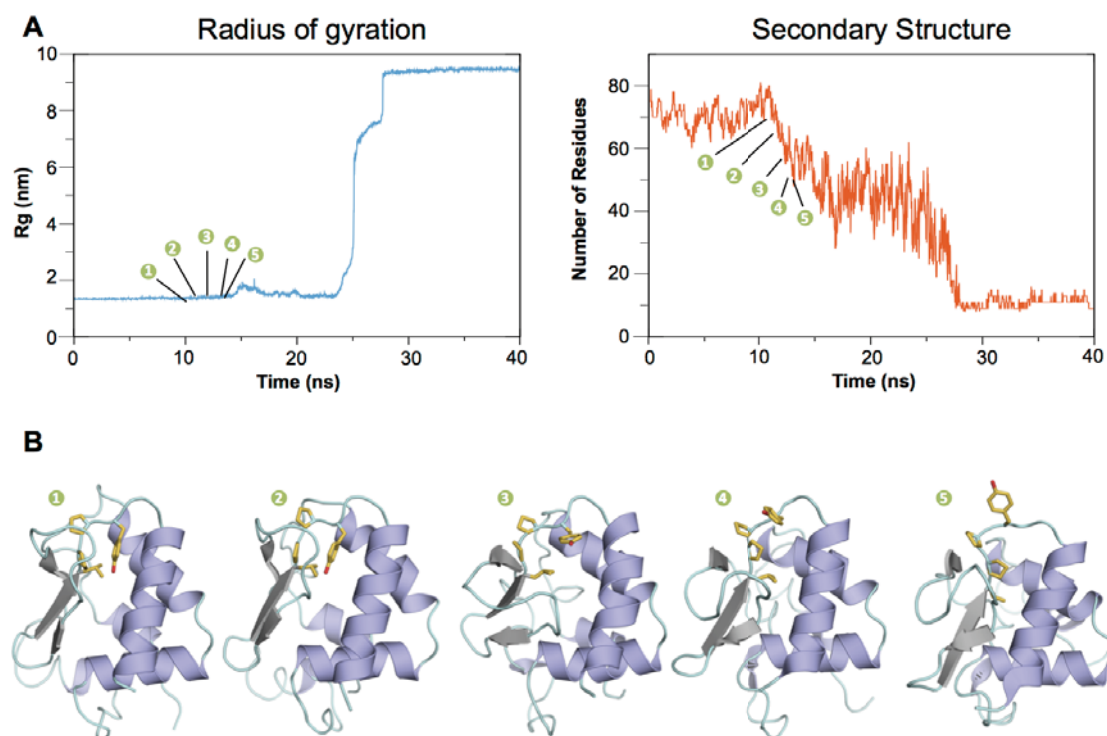


Figure 6.6: Simulated annealing MD simulation for the destabilization of the HECT C-lobe. A 40 ns trajectory was calculated with a temperature ramp from 310 to 600 K. (A) *Left panel:* radius of gyration (nm) of the structure along the simulation. The sudden increase at 24 ns corresponds to the unfolding of the protein. *Right panel:* number of residues in secondary structure elements, including α -helices, β -sheets, β -bridges and turns. (B) Snapshots extracted from the simulation, corresponding to 10 (1), 12.25 (2), 12.48 (3), 13.52 (4) and 13.59 (5) ns. Frame 5 displays the exposure of Tyr931, making the HECT PY motif accessible for WW recognition. The radius of gyration varies from 1.36 nm in frame 1 to 1.37 nm in frame 7, while its original value is 1.33 nm.

With these results in hand, we hypothesized that the exposure of the HECT PY motif needs to be triggered and is probably related to the destabilization of the protein. This may naturally occur for several reasons including ubiquitin loading of the close catalytic cysteine residue, a post-translational modification or simply ageing of the ligase.

To test the feasibility of our hypothesis, we performed a simulated annealing MD simulation where the protein was destabilized along a temperature ramp. A 40 ns trajectory was calculated with progressive heating from 310 to 600 K (Figure 6.6). Of note, the temperature values in such simulations are not representative of actual experimental values.

The protein is progressively destabilized as indicated by the increasing variability of the radius of gyration (r_g , nm) and the decrease of the number of residues in secondary structure elements (**Figure 6.6A**). The snapshots shown in **Figure 6.6B** illustrate the crescent instability of the area where the HECT PY is located. The motif is packed in the hydrophobic core during the initial steps. After 13.59 ns (snapshot 5) residue Tyr931 is exposed while the protein roughly maintains its fold. The internal destabilization continues until the unfolding of the protein, as indicated by a sudden dramatic increase of the radius of gyration accompanied by the loss of most of the residues in secondary structure elements.

6.2.5. Binding of Nedd4L WW domains to the HECT-PY peptide

Even though the binding of the WW domains to the HECT-PY motif is hampered when the HECT domain is completely folded (Section 6.2.2), its unfolding permits the establishment of these interactions (Bruce et al., 2008). In Sections 6.2.3 and 6.2.4 we suggest that the complete unfolding of the domain is not a must as the PY motif is exposed during the destabilization of the protein while it roughly maintains its native fold.

To characterize these bindings from a biophysical and structural point of view we synthesized a 15 residues peptide (Arg925 – Lys940) including the PY motif (HECT-PY peptide). The peptide was synthesized using SPPS as detailed in Section 6.2.1.

NMR HSQC titrations were performed consisting in the addition of progressively increasing amounts of the HECT-PY peptide to ^1H - ^{15}N labelled samples of each individual WW domain in Nedd4L (**Figure 6.7A**). My fellow colleagues Dr Nina Görner and Tiago Gomes had previously performed the assignment of the resonances in the HSQC spectra of these domains.

Figure 6.7 displays a comparison of the results obtained for each WW domain after the addition of 3 equivalents of the HECT PY peptide. WW3 shows the most intense chemical shift perturbations, with $\Delta\delta_{AV}$ values up to 0.75 ppm. The most affected residues are K501, T502, T503 and T504, located in the second loop and $\beta 3$ (**Figure 6.7B and C**).

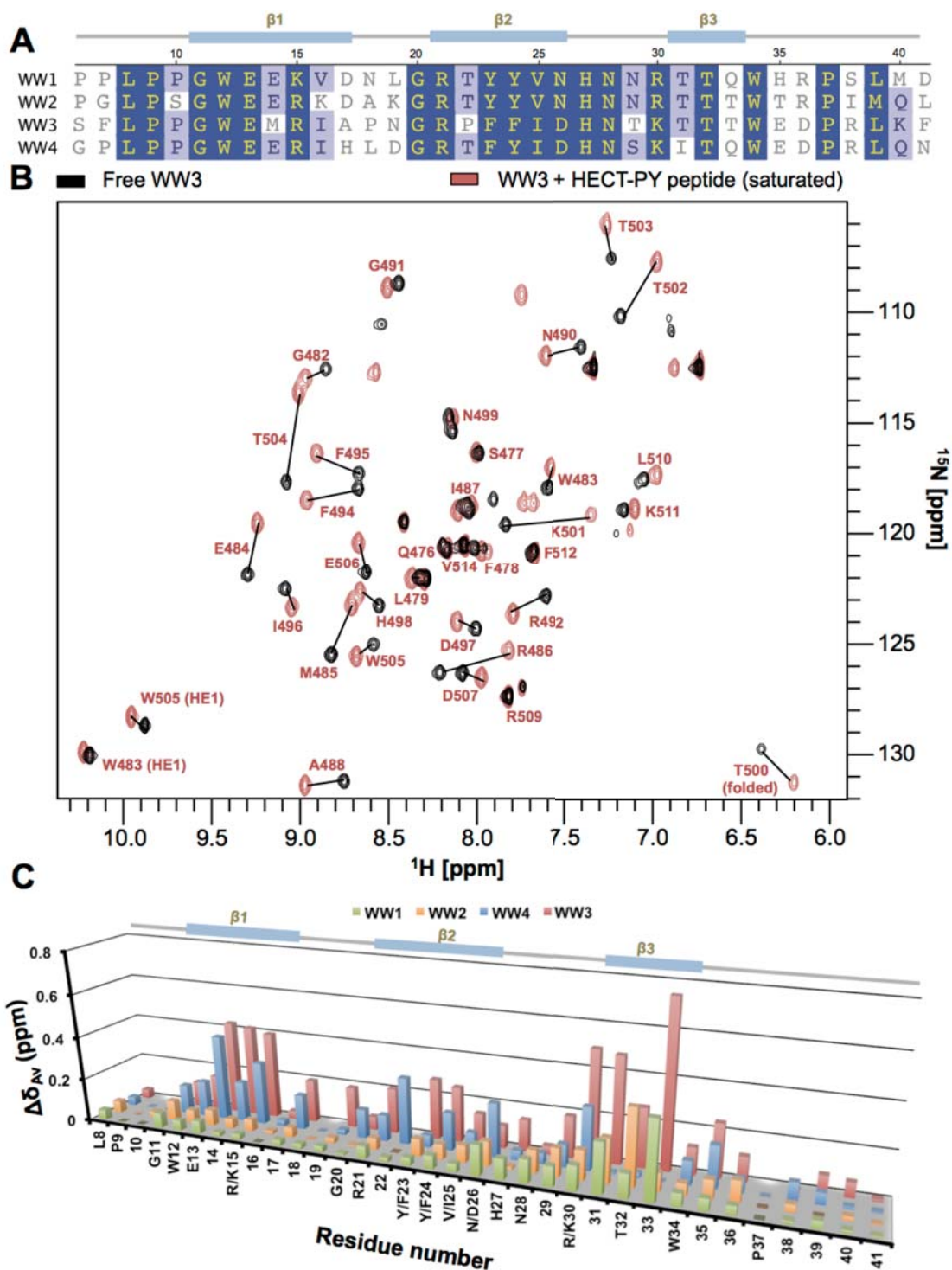


Figure 6.7: HSQC titrations of the HECT PY peptide on the ^1H - ^{15}N labelled WW domains of Nedd4L. Data obtained from the addition of progressively increasing amounts of the peptide to each individual domain. (A) Sequence alignment of the four WW domains in Nedd4L. Sequence numbers are arbitrary. (B) HSQC spectra acquired on a free WW3 sample (black) and on the same sample after the addition of 3 equivalents of the HECT PY peptide (orange). Chemical shift perturbations are indicated. Sequence numbers as in isoform 5. (C) Plot displaying the $\Delta\delta_{AV}$ values (ppm) per residue observed after the addition of 3 equivalents of the HECT PY peptide to each individual WW domain in Nedd4L. Sequence numbers are arbitrary.

All other WW domains also bind the HECT PY sequence, although the strength of these interactions is apparently lower as the observed chemical shift perturbations are smaller and the amount of peptide needed to saturate the interactions was bigger (**Figure 6.7C** and Section 9.3).

Next, we used Isothermal Titration Calorimetry (ITC) to obtain accurate measurements of the binding affinities of these interactions (**Figure 6.8**). The interaction between the WW3 domain and the HECT-PY peptide was confirmed as the strongest, with a $K_D = 4.4 \pm 0.8 \mu\text{M}$. This binding affinity is comparable to our previous measurements of Nedd4L WW domains binding to the Smad2/3 and Smad7, which are substrates of the ligase (Aragón et al., 2012; Aragón et al., 2011).

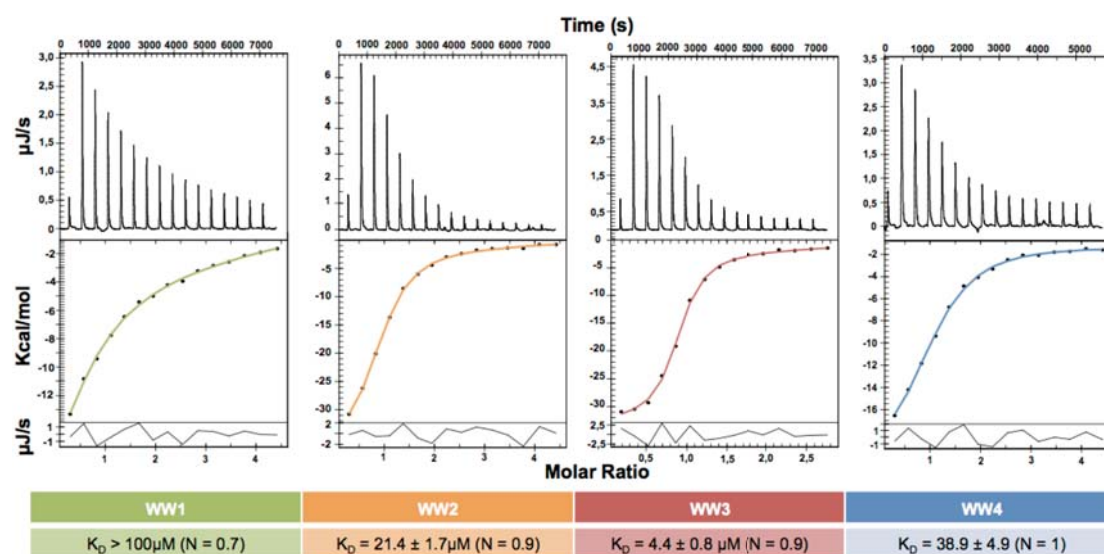


Figure 6.8: ITC titrations of the HECT PY peptide on the WW domains of Nedd4L. For each titration, the measured differential heat in function of time is shown in the top panel. The lower panel shows the integrated data and the calculated residuals in function of the ligand molar ratio, as well as the fitted binding isotherm. The derived K_D values are displayed below. The experiments were acquired at 5 °C in a phosphate buffer at pH7. The WW3 domain displayed the highest affinity for the peptide, with a K_D value in the range of WW domains substrate recognition.

The values obtained for all other domains were significantly lower. The WW2 domain binds the peptide with a $K_D = 21.4 \pm 1.7 \mu\text{M}$, a 5 fold weaker binding. The measured value for the WW4 domain was $K_D = 38.9 \pm 4.9 \mu\text{M}$, which is an order of magnitude weaker when compared to the WW3 domain. Finally, the titration with WW1 yielded a K_D value $> 100 \mu\text{M}$.

Taken together, the results from our NMR and ITC titrations reveal a preferential binding of the WW3 domain to the HECT PY motif when this is exposed. Moreover, the measured affinities are in agreement with the compared binding of the WW

domains relative to each other inferred from existing biochemical data, obtained in the context of the denatured HECT domain (Bruce et al., 2008).

6.2.6. Solution structure of the Nedd4L WW3 domain in complex with the HECT-PY peptide

The basis of the high affinity interaction between the Nedd4L WW3 domain and the HECT-PY peptide were further investigated by solving the structure of the complex in solution (PDB entry 2MPT, BMRB entry 25000). Most of the WW3 domain resonances were assigned from 3D NMR spectra. The CBCAcoNH/CBCANH pair was acquired on a ^{13}C - ^{15}N labelled sample of the domain in complex with the unlabelled peptide to assign the backbone resonances using the sequential assignment strategy (Section 3.4.1.3). ^{15}N -edited TOCSY and NOESY experiments were acquired to assign side-chain proton resonances as well as NOE restraints within the domain. Missing assignments were obtained from 2D homonuclear ^1H - ^1H TOCSY and NOESY spectra, which were also used to assign the resonances of the HECT-PY peptide as well as the intra-molecular NOE restraints defining the complex (detailed in **Table 6.2**). For clarity, residues in the WW3 domain are named using the three-letter amino acid code, whereas peptide residues are referred using the one-letter code.

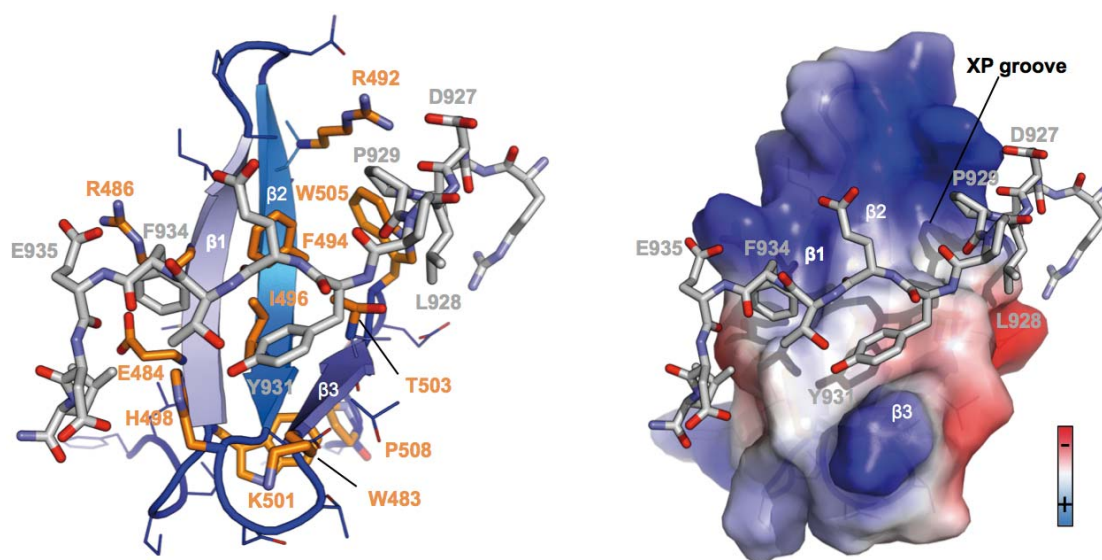


Figure 6.9: Solution structure of the Nedd4L WW3 - HECT-PY complex. The lowest energy structure is shown. In the left panel, the WW domain (residues 478 to 509) is shown in blue cartoon (main chain) and orange sticks and labels (side-chains taking part of the peptide coordination). The peptide (residues 925 to 937) is shown as grey sticks and labels. In the right panel, the surface charge distribution of the domain is shown from the same perspective as in the left panel.

Domain resonances in the vicinity of aromatic residues displayed atypical chemical shifts, reproducing values that are characteristic in WW domains (Macias et al., 2000; Macias et al., 1996). The indole ring in Trp483 contributed to the up-field shift of the resonances corresponding to Pro508 (HB2/3: 0.765 and 0.693 ppm; HG2/3: 0.536 and 0.162 ppm) and Asp497 (HB2/3: 1.908 and -0.035 ppm). Phe495 affected the HA resonance of Asp507 (2.521 ppm) whereas the proximity of His498 caused up-field shifts of the Lys501 resonances (HA: 2.122 ppm, QB: 1.446 ppm, QE: 1.830 ppm, QD: 0.966 ppm and HG2/3: 0.885 and 0.800 ppm).

Nedd4L WW3 - HECT PY				Nedd4L WW3 - HECT PY			
⁹²⁵ RLDLPPYETFE DLREK ⁹⁴⁰		Inter-molecular NOEs	Intra-molecular NOEs	⁹²⁵ RLDLPPYETFE DLREK ⁹⁴⁰		Inter-molecular NOEs	Intra-molecular NOEs
AA	Proton			AA	Proton		
D927	HN		D926 HN HA QD1		QD/E		T933 HA QG2 E935 HA HB2 HB3 HG2
	HA	TRP505	HD1				
	HB2/3	TRP505	HD1 HE1				
L928	QD1	PHE494	QD QE	F934		PHE494 QE ILE496 QG2 HG13	
		THR504	HN			HIS498 HE1	
		TRP505	HN HD1 HE1	HZ			E935 HG2
	HD2/3	PHE494	QD QE			PHE494 QE ILE496 QG2 HG13	
P929	HG2	ALA488	QB			HIS498 HE1	
		PHE494	QD QE	E935	HN		F934 HA
	HA	THR503	QG2		HG2	ARG486 HH22	
	HB2/3	THR503	QG2		HN		E935 HN HA
	QD/E	ILE496	QG2	D936			HB2 HB3 HG2
		HIS498	HE1				
Y931		LYS501	HA HB2 HD2 HE2 HG2 HG3				D936 HA HB2 HB3
		THR503	QG2		HN		
T933	QG2	HIS498	HE1	L937	HG	HIS489 HD2 HE1	
	HN		T933 HN HA		QD1	HIS489 HD2 HE1	
F934			HB QG2 E935 HN		QD2	GLU484 HN HIS489 HD2 HE1	

Table 6.2: Inter and intra-molecular NOE restraints for residues in the Nedd4L HECT-PY peptide in complex with the WW3 domain. Obtained through the assignment of homonuclear ¹H-¹H 2D NOESY spectra.

The domain displays the typical WW domain fold consisting of three anti-parallel β -sheets. Long-range restraints that are characteristic of WW domains define this fold, including the observed contacts between W483 and P508. The HECT-PY peptide is bound in a negative manner, from the C- to the N-terminus. The peptide binds the surface of the domain adopting an extended configuration.

PDB: 2MPT	
BMRB: 25000	
Restraints used for the calculation (120 structures ensemble)	
Inter-domain	68
Sequential ($ i - j = 1$)	164
Medium range ($1 < i - j \leq 4$)	70
Long range ($ i - j > 4$)	229
Dihedrals	74
Hydrogen bonds	10
All restraints (unambiguous)	615
Restraint ratio (45 residues)	13.7
RMSD (Å) from experimental	
NOE ($\times 10^{-3}$)	5.5 ± 0.8
Bonds ($\times 10^{-3}$) (Å)	5.3 ± 0.5
Angles ($^{\circ}$)	0.71 ± 0.05
Coordinate Precision (Å)	
Backbone, all residues in the complex (45 residues)	0.94 ± 0.19
CNS potential energy (kcal mol⁻¹)	
Total energy	-1263 ± 63.38
Electrostatic	-1569 ± 40.23
Van der Waals	-116.6 ± 24.96
Bonds	22.12 ± 4.55
Angles	108.2 ± 16.88
Structural quality (% residues) 20 best structures	
In most favored region of Ramachandran plot	88
In additionally allowed region	12

Table 6.3: PROCHECK statistics for the Nedd4L WW3 - HECT-PY complex.

As a group I WW domain, Nedd4L WW3 has both a hydrophobic pocket and an XP-groove, which respectively coordinate the Tyr and Pro residues in PY motifs. In this complex, P929 binds the XP-groove formed by Phe494 (β_2), Thr503 (β_3) and Trp505 (β_3). NOE signals were detected from the Phe494 aromatic ring to the P929 side-chain contributing to this positioning, which is supported by the additional contacts between L928 and Phe494, Thr504 and Trp505 as well as the electrostatic interaction between D927 and Arg492 (β_2). The latter interaction is analogous to the one we observed in the Nedd4L WW2-Smad7 PY complex (Aragón et al., 2012).

Y931 binds to the hydrophobic pocket centred in the second loop of the domain and formed by Ile496 (β_2), His498 (β_2 - β_3 loop), Lys501 (β_2 - β_3 loop) and Thr503 (β_3). Abundant contacts are observed from the side-chains of these residues to the aromatic ring of Y931.

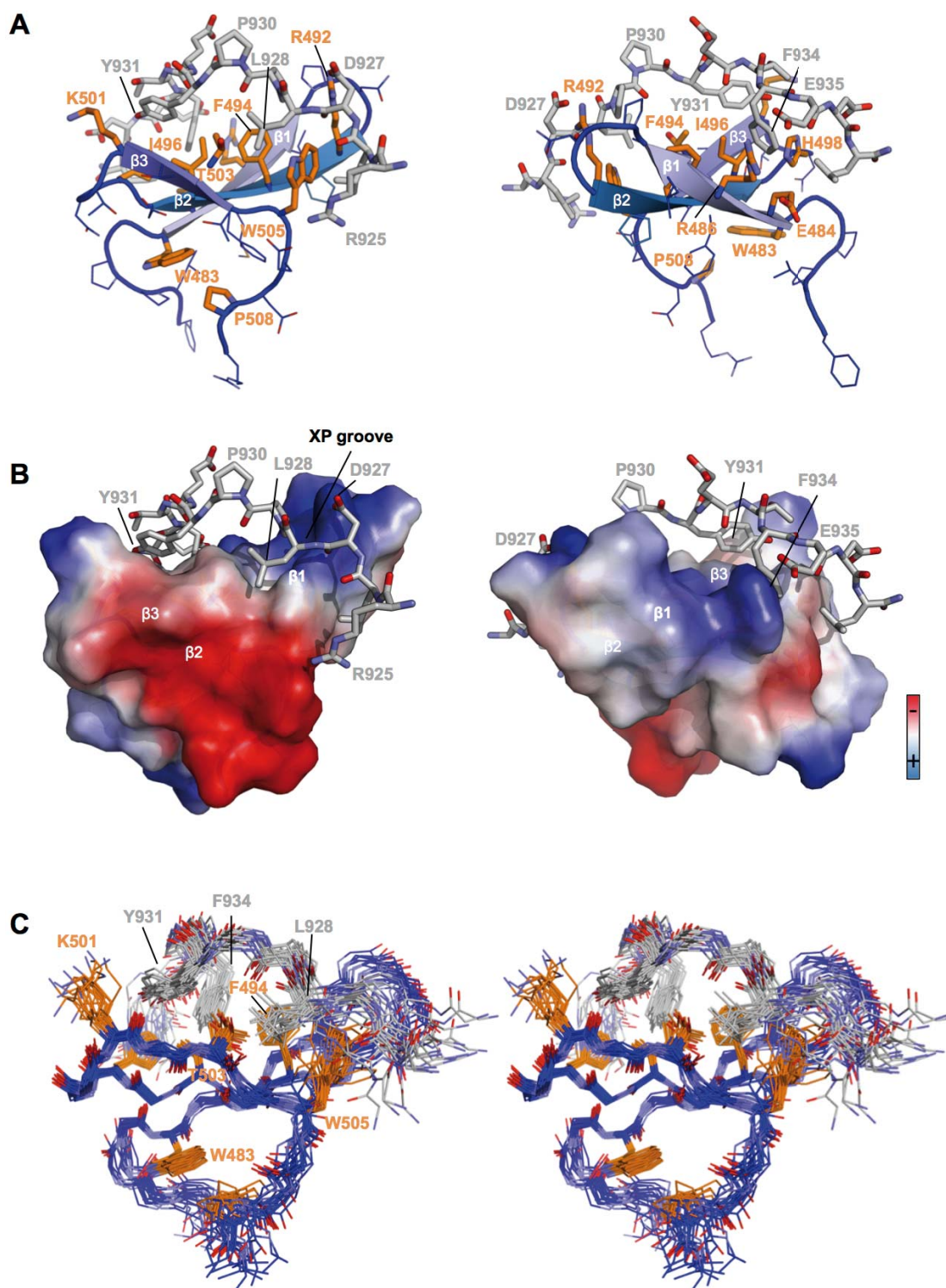


Figure 6.10: Side-views of the solution structure of the Nedd4L WW3 - HECT-PY complex. (A) As in Figure 6.9, left panel. (B) As in Figure 6.9, right panel. (C) Stereo view of the ensemble of 20 lowest energy structures derived from the calculation, shown in the same orientation as in (A), left panel. The alignment of the structures was performed using the software Theseus (Theobald and Steindel, 2012).

The aromatic ring of F934 is protected in an additional hydrophobic pocket formed by Phe494 ($\beta 2$), Ile496 ($\beta 2$) and His498 ($\beta 2$ - $\beta 3$ loop), as indicated by the detection of

abundant NOE signals correlating resonances in the side-chains of these residues. An electrostatic interaction between E935 and Arg486 ($\beta 1$) can be observed in the structure as well.

Both electrostatic and hydrophobic contacts from residues in the WW domain to the side-chains of L928, D927, F934 and E935 in the peptide contribute to the stable coordination of the PY motif and play a decisive role regarding the high affinity of this binding.

Regarding the quality of the structure, its analysis with PROCHECK-NMR (Laskowski et al., 1996) yielded a Ramachandran plot displaying 88% of the residues in optimal regions and 12% in additionally allowed regions. The RMSD of the 20 best structure ensemble was 0.94 ± 0.19 Å, reporting a good convergence of the ensemble. The complete statistics of the calculation are detailed in **Table 6.3**.

6.3. Summary

The experiments presented in this chapter were aimed at investigating the molecular mechanisms permitting the E3 ubiquitin ligase Nedd4L to control its turnover rate through ubiquitination. Using molecular biology methods, this auto-regulatory mechanism had previously been shown to play a relevant role in the biology not only of this ligase, but also of all other members of the Nedd4 family and many other E3 ubiquitin ligases structurally unrelated (De Bie and Ciechanover, 2011).

In general, E3 ligases in the Nedd4 family use the WW domains present in their modular structure to recognize the PY motifs in the target substrates that they ubiquitinate. The presence of a highly conserved PY motif in the catalytic HECT domain of these ligases suggests this is likely used for self-recognition and ubiquitination (**Figure 6.1A** and **B**). However, all crystal structures available for the HECT domains of ligases in the family display the tyrosine residue in the PY motif packed inside the hydrophobic core of the protein (**Figure 6.1C**).

Regarding Nedd4L, we considered the possibility of its HECT PY motif being exposed when the HECT domain is folded and stable in solution. This would permit its recognition by the WW domains of the ligase. However, HSQC titrations consisting in the addition of progressively increasing amounts of the unlabelled WW2 and WW3 domains to the ^1H - ^{15}N labelled HECT C-terminal lobe (containing both the HECT PY motif and the catalytic cysteine residue) showed that this binding does not occur in these circumstances (Section 6.2.2).

These observations were supported by our T_1 , T_2 and heteronuclear NOE relaxation measurements on the HECT C-lobe, that we used to sample possible rearrangements of the molecular structure in solution in the fast time-scale (ps to ns). These experiments reported a compact fold of the molecule, with little internal flexibility limited to the longest loop regions (Section 6.2.3). We also considered the possibility of these rearrangements occurring at longer time-scales, with slower rates. To investigate that, we calculated a 120 ns molecular dynamics simulation trajectory at 310 K. The fold of the molecule was compact along the simulation with no internal rearrangements, and the flexible areas correlated with the relaxation data (Section 6.2.4).

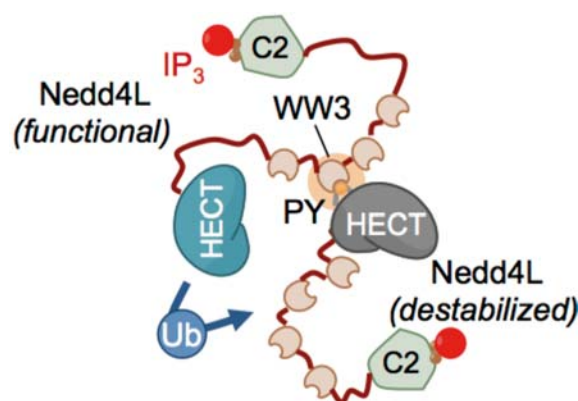


Figure 6.11: Bi-molecular model for the ubiquitination-mediated turnover regulation of Nedd4L. Destabilization of a Nedd4L molecule in the open, active conformation results in the exposure of the HECT-PY motif. This may naturally happen during the ageing or the catalytic action of the ligase, or in response to a post-translational modification. A functional Nedd4L molecule uses its WW3 domain to recognize the exposed HECT-PY motif in its destabilized counterpart and ubiquitinates it for proteasomal degradation.

Prompted by these observations we hypothesized that the exposure of the HECT PY motif is triggered, probably by an event related to its natural destabilization such as ageing, ubiquitin loading or in response to a post-translational modification. We tested that by destabilizing the protein *in silico*, using a simulated annealing MD simulation. A 40 ns trajectory was calculated with a 310 to 600 K temperature ramp. The exposure of the Y931 in the HECT PY motif was observed while the overall structure of the HECT C-lobe was roughly maintained (Section 6.2.4).

Next, we mimicked this solvent exposure *in vitro* by synthesizing a short peptide including the HECT PY motif. We studied the binding of this sequence to all four WW domains in Nedd4L using both NMR and ITC titration methods. These revealed that WW3 preferentially binds the sequence with an affinity value close to that observed for the binding of this domain to several targets of the ligase ($4.4 \pm 0.8 \mu\text{M}$) (Section 6.2.5). We investigated the structural basis of this high affinity interaction

by solving the solution structure of the WW3 – HECT-PY peptide complex (Section 6.2.6).

According to these results, we propose a bi-molecular model for the ubiquitination-mediated self-regulation of this ligase (**Figure 6.11**). In this model a destabilized Nedd4L molecule exposes its HECT PY motif to the solution. A fully functional Nedd4L molecule binds this motif using its WW3 domain and ubiquitinates its destabilized counterpart, labelling it for proteasomal degradation.

CHAPTER 7. Paramagnetic labelling of the Nedd4L WW3 – HECT-PY complex

7.1. Introduction

Spin labels causing paramagnetic effects measurable by NMR are currently of generalized use for the acquisition of long-range restraints. While classical NMR observables such as NOEs, which rely on the through-space transfer of magnetization, are limited to distances below 5 Å, spin labels can be used to measure twelve-fold longer distances up to 60 Å (Lian and Roberts, 2011). Such measurements are extremely valuable in the structural characterization of big macromolecules as well as in studies involving large macromolecular complexes where they can provide valuable information on docking and relative inter-domain orientation.

The generalized use of paramagnetic tools responds to the interest on the characterization of large macromolecular assemblies in solution. Our research interests have progressed towards this direction and therefore we set to undertake the first steps for the incorporation of paramagnetism to the set of tools we utilize in our NMR studies. As an example, the characterization of the C2-HECT contacts found in the latent conformation of Nedd4L (Section 5.2.3) would benefit from the application of this technique.

Nitroxide labels are widely used paramagnetic tags, and they provide distance measurements of up to 25Å. These tags are organic molecules with an unpaired electron in a sterically protected nitroxide group that are designed to attach to surface-exposed cysteine residues. They enhance T1 and T2 relaxation rates of nearby nuclei in a distance dependent manner. As a consequence, the volume of the spectral peaks corresponding to nuclei within the tag's effective range is decreased. This effect is known as paramagnetic resonance enhancement (PRE). Other paramagnetic effects such as residual dipolar couplings (RDC) or pseudo-contact shifts (PCS) have different consequences on the NMR signals and can be induced using other tags (Section 3.4.1.6).

Usually the presence of only one tag on the protein's surface is required for the measurements. The positioning of the tag on the surface of the protein can be engineered using site-directed mutagenesis. Residues in positions of interest can be

mutated into cysteine to label them and naturally occurring cysteines should be removed to avoid undesired labelling.

We have used the detailed structural characterization of the Nedd4L WW3-HECT PY complex (Section 6.2.6) to work on the incorporation of spin labelling methods for the acquisition of paramagnetic relaxation enhancement (PRE) restraints. In this chapter the engineering of two WW3 mutants and their labelling with the nitroxide spin label MTSL and its diamagnetic counterpart MTS are presented. This has settled the bases for the application of this powerful NMR technique to our particular research interests.

7.2. MTSL derivatization of Nedd4L WW3 cysteine mutants

As it is the case for most nitroxide spin labels, MTSL/MTS are covalently attached to the target protein through a covalent bond with the side-chain thiol group of an exposed cysteine (**Figure 7.1**). In order to incorporate the tag to our system, we used site-directed mutagenesis to engineer two mutants of the Nedd4L WW3 domain, each containing a cysteine residue. Because the wild-type sequence of the domain does not present any cysteine residues no depletion mutations were needed.

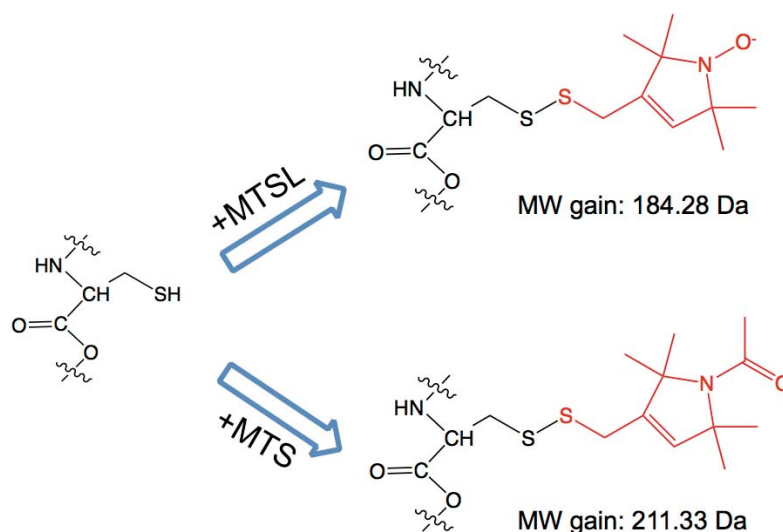


Figure 7.1: MTSL/MTS derivatization of cysteine residues. The added moieties are shown in red.

To avoid undesired effects affecting the domain structure and PY binding capabilities, both mutations were incorporated in non-conserved positions located in the loop regions (**Figure 7.2**). One cysteine was introduced in the $\beta 1$ - $\beta 2$ loop (P489C), which is not directly implicated in the coordination of the PY motif. The other mutation was located in the $\beta 2$ - $\beta 3$ loop (T500C) very close to the hydrophobic

pocket coordinating Y931 in the HECT-PY peptide. ^1H - ^{15}N labelled samples of both mutants were prepared.

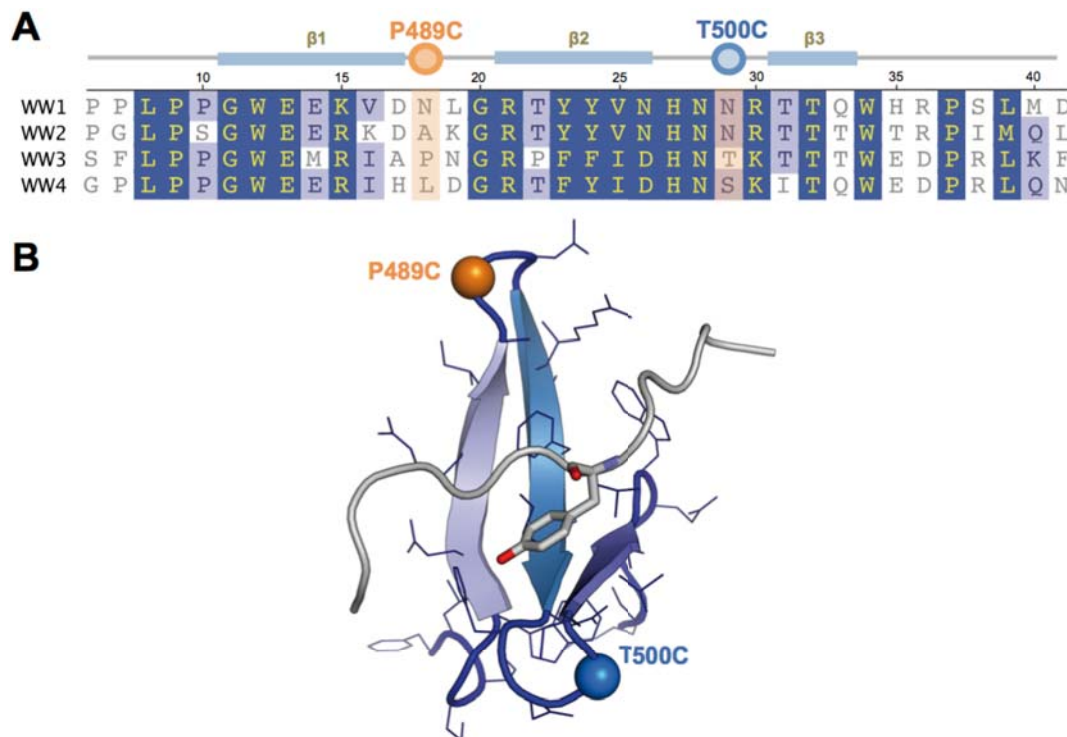


Figure 7.2: Nedd4L WW3 mutants engineered for spin labelling. Two separate mutants were created, each of them containing one of the mutations represented in the figure. The mutated positions are labelled and highlighted with orange (P489C) and blue (T500C) circles (A) or spheres (B).

The incorporation of the cysteine residues partially compromised the solubility of the domain during its purification. Similar to the production protocol used for the wild-type version of the protein (Section 6.2.1) the mutants were expressed in inclusion bodies and solubilized using a 6 M guanidine hydrochloride buffer. Instead of the fast refolding step used for the wild-type during Ni^{2+} affinity purification, the mutants were eluted from the Ni^{2+} resin in their denatured states and refolded in a buffer containing 2 mM DTT by dialysis (Figure 3.5).

The His_6 tag was removed by TEV digestion and the protein was incubated with 10 mM DTT while concentrating it for 2h from an initial volume of approximately 30 mL to a final volume of 0.5 mL. This was done in order to reduce all the molecules in the sample. After that, it was diluted to 3.5 mL in a DTT-free buffer and re-concentrated to 0.5 mL (three times) in order to remove the DTT. At that point, 7-10 equivalents of the spin label (MTSL/MTS) were added to the solution. The reaction proceeded overnight with orbital agitation at room temperature and protected from the light. The reagent excess together with other impurities, were removed with a

final size exclusion chromatography purification step using an analytical Superdex™ Peptide 10/300 GL column.

The success of the labelling reaction was evaluated using MALDI-TOF mass spectrometry. Small aliquots of the sample were taken before and after incubation with the spin label. To obtain optimal signals, the buffer salts were removed with a small scale reversed-phase chromatography step performed with a ZipTip®. These are pipette tips with a small amount of C₁₈ RPC stationary phase that in aqueous media retain peptidic samples but not salts. The sample is eluted in a solution of 80% acetonitrile. This permits a much more accurate detection of the ions by MS, but the fold of the protein is destroyed. **Table 7.1** summarizes the results obtained using this methodology.

Mutant	MTS (m/z)			MTSL (m/z)		
	-	+	$\Delta m/z$	-	+	$\Delta m/z$
P489C	5694.75	5907.16	212.41	5691.35	5877.49	186.14
T500C	5664.69	5877.53	212.84	5664.39	5848.74	184.35

Table 7.1: Monitoring of MTSL/MTS derivatization with MALDI-TOF MS spectrometry. The experimental values obtained for each sample are shown (m/z). In all cases, the singly ionized specie was detected. The expected $\Delta m/z$ were: MTS -> 211.33, MTSL -> 184.28.

In the HSQC spectra of both mutants, the resonances corresponding to a few residues next to the mutated position were slightly affected, but all others presented the same chemical shift as in the wild-type version of the WW3 domain. No effects in the chemical shifts were observed due to derivatization with MTSL and MTS. Therefore, neither the mutation nor the derivatization affected the fold of the protein.

The addition of 3 equivalents of the HECT-PY peptide caused the same chemical shift perturbations observed for the wild-type version of the protein, indicating that the presence of the spin label did not affect the PY binding capabilities of the domain either (**Figure 7.3**).

For each WW3 mutant – HECT-PY complex, the peaks in the HSQC spectra of both samples derivatized with MTSL and MTS were integrated using nmrPipe (Delaglio et al., 1995). The resonance corresponding to Asn515 was selected to normalize all the signals in each spectrum, given that it is located far from the labelling positions and is unaffected by the binding of the HECT-PY peptide. The I_{MTSL}/I_{MTS} ratios of each peak's normalized intensities were calculated and are shown in **Figure 7.4**.

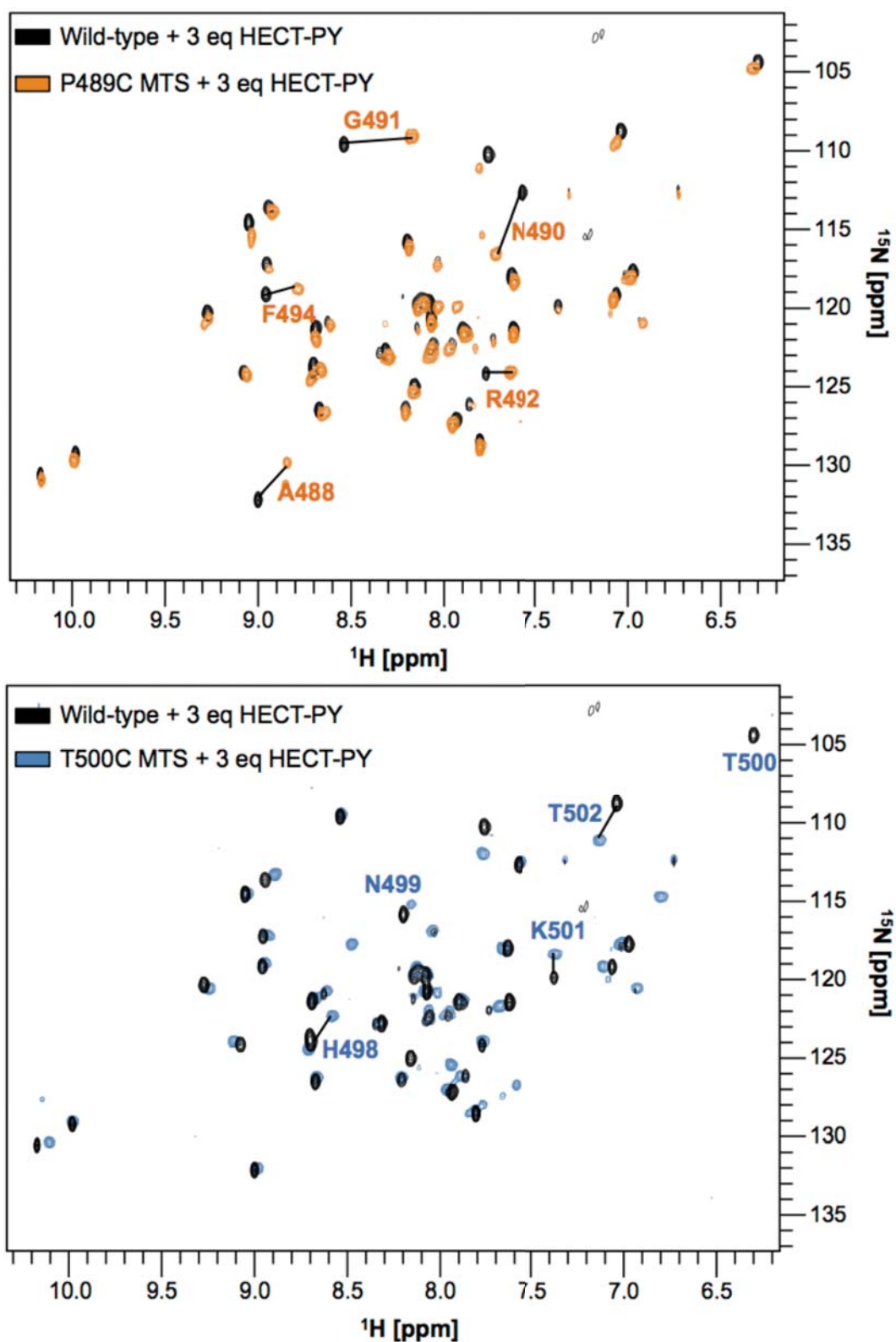


Figure 7.3: HSQC spectra of the MTS-derivatized WW3 mutants bound to the HECT-PY peptide. The chemical shifts of the MTS-derivatized mutants bound to the HECT-PY peptide are equivalent to the WT version of the experiment. Slight perturbations next to the mutated positions are labelled.

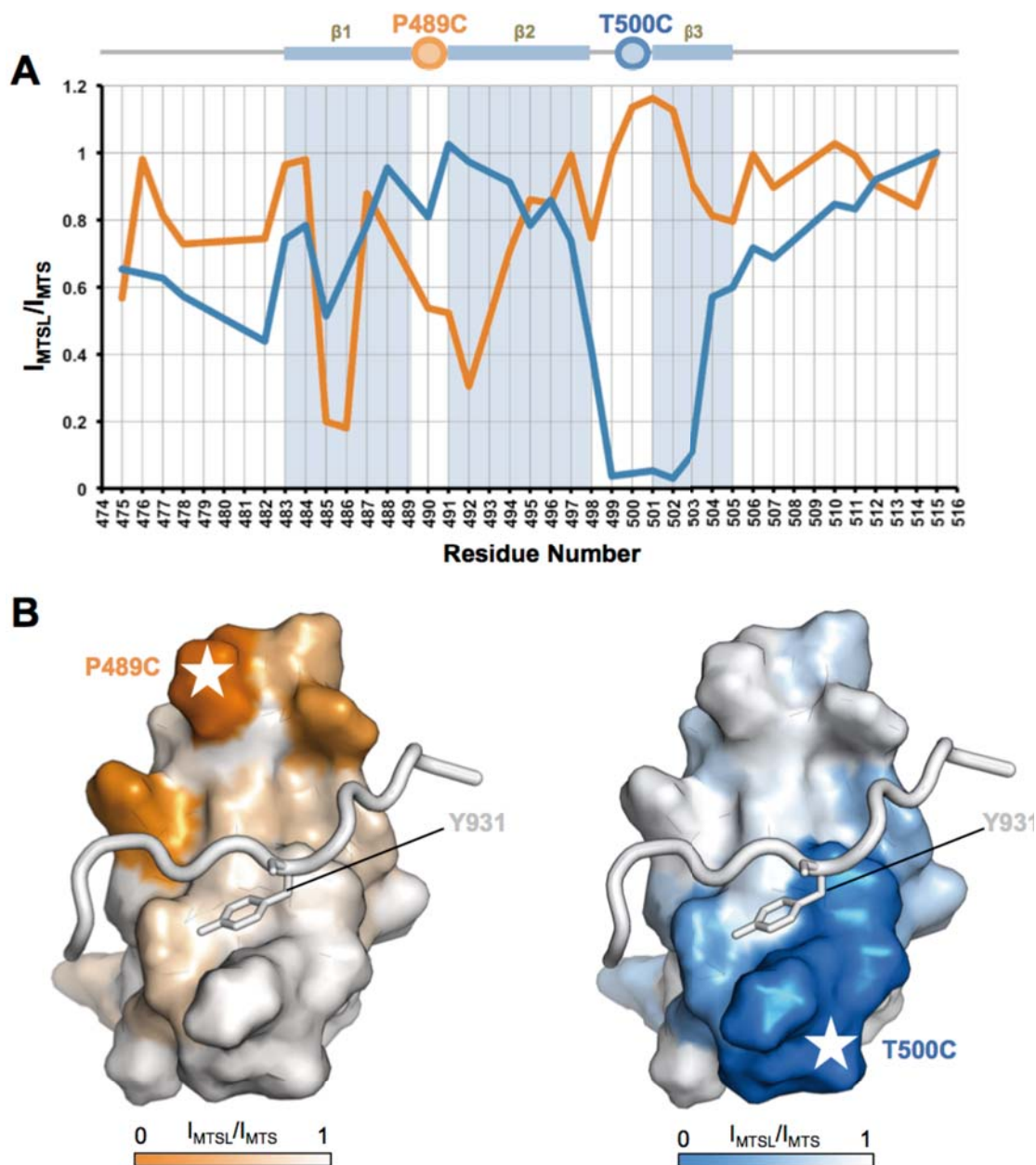


Figure 7.4: PRE data obtained from the MTSL/MTS labelled Nedd4L WW3 cysteine mutants. A) I_{MTSL}/I_{MTS} ratios obtained per each residue in the sequence, both for the P489C (orange) and the T500C (blue) mutants. (B) Surface representation of the values displayed in (A). The residues are coloured in an orange (P489C) or blue (T500C) gradient depending on their I_{MTSL}/I_{MTS} values. The HECT PY peptide is shown as a white cartoon, with the residue Y931 highlighted in sticks representation.

In the case of the P489C mutant, the resonances of most of the residues located in the Glu484 – Phe494 segment displayed I_{MTSL}/I_{MTS} values below 0.8, indicating that they were significantly affected by the presence of the paramagnetic tag. This defines an approximate effective radius of 15 Å, as measured to the furthest spatially located affected residue (Met485). However, not all the resonances corresponding to residues within this range were affected as expected, as judged by their position in the structure of the complex.

The data obtained from the T500C mutant was more consistent. All resonances corresponding to residues located in the Asp497 – Asp507 segment displayed $I_{\text{MTSL}}/I_{\text{MTS}}$ values below 0.8, with a clear minimum displaying values close to 0 centred in position 500. Low $I_{\text{MTSL}}/I_{\text{MTS}}$ values of the residues in the Phe478 – Met485 segment reflect the spatial proximity of the tag location to the N-terminus of the protein. In this case, the observed approximate effective radius of the tag is 19 Å and most of the residues within this range are affected in a consistent manner regarding their distance to the tag.

The obtained $I_{\text{MTSL}}/I_{\text{MTS}}$ values can be translated into distance restraints to be used in structural calculations. These preliminary experiments we performed in a system that we had previously characterized enable us to use PRE based measurements in other systems studied in the lab. This is of special relevance for the study of the interactions between folded proteins.

7.3. Summary

Paramagnetic tagging widens the potential of NMR at studying protein-protein interactions, as it permits the measurement of much longer distance restraints than conventional NMR observables. The application of this technique to our research interests has the potential to improve studies such as the Nedd4L C2 – HECT domain interactions reported in this thesis.

In this chapter we present a preliminary paramagnetic relaxation enhancement (PRE) study performed on the extensively characterized WW3 – HECT-PY complex. Two WW3 cysteine mutants were engineered and labelled in ^{15}N and with both paramagnetic (MTSL) and diamagnetic (MTS) tags. The fold and substrate binding capacities of the domain were not affected in either case. PRE effects within an effective range up to 19 Å were observed by comparing the HSQC peaks intensity of the mutants in the presence of the paramagnetic (I_{MTSL}) and the diamagnetic (I_{MTS}) tags.

Along the procedure, the protocols have been optimized for the generation of WW domain mutants with intact fold and function, the derivatization of the cysteine residues with the spin labels, the MALDI-TOF-based detection of the derivatized samples and the measurement of the PRE restraints. With these preliminary experiments we now have the tools to apply this technique to the characterization of systems requiring the measurement of long-range restraints.

Part V.

Discussion

CHAPTER 8. DISCUSSION	183
8.1. Antecedents	183
8.2. Smad phosphorylation affects substrate binding in Nedd4L and related proteins	184
8.3. Nedd4L changes from a latent to an active conformation in response to IP₃/Ca²⁺	187
8.4. A crossed mechanism for Nedd4L ubiquitination	190
8.5. Concluding remarks	193

CHAPTER 8. Discussion

8.1. Antecedents

Ubiquitination is a major post-translational modification that serves eukaryotic cells in the control of protein function and fate. It affects many fundamental aspects of cell biology including cell cycle and division, embryo development and tissue homeostasis, and its malfunction has been related to a variety of life-threatening diseases (Petroski, 2008). Most commonly proteins are labelled with Lys48 poly-ubiquitin chains, which results in their proteasome-mediated degradation. This constitutes the ubiquitin-proteasome pathway, the major cellular mechanism to control the turnover of damaged proteins and functional ones that are involved in temporary processes requiring termination. Other ubiquitination patterns label proteins for a variety of functions.

The final effectors of ubiquitination are a group of enzymes known as the E3 ubiquitin ligases. There are several hundreds of such enzymes in humans, each of them being highly specialized in a small subset of targets. Many of these ligases are tightly regulated at different levels in order to protect their substrates from unspecific and undesired ubiquitination. In some cases, phosphorylation of the ligases or their targets affects their substrate recognition capabilities. In some others, the ligases exist in a latent conformation that is activated in response to well-defined events. In addition, most of them are able to auto-ubiquitinate and label themselves for degradation, thus providing a negative feedback control mechanism on their activity.

Nedd4L is a HECT-type E3 ubiquitin ligase belonging to the Nedd4 family that regulates the internalization and degradation of a variety of plasma membrane ion channels, including the epithelial sodium channel ENaC (Staub et al., 1996), and is therefore involved in the homeostasis of sodium and other ions in the body fluids. Defects on the Nedd4L-mediated regulation of some of these channels have been linked to diseases such as Liddle's syndrome and cystic fibrosis. In addition, it has recently been discovered to regulate the TGF- β cytokine membrane receptor as well as the receptor activated Smad2/3 and the inhibitory Smad7 (Aragón et al., 2011; Gao et al., 2009). Deregulation of the TGF- β signalling pathway has been related to several types of cancer and metastasis.

Bearing in mind the biomedical relevance of Nedd4L, the experimental work presented in this thesis has been directed towards contributing to the understanding from a structural and biophysical perspective of the regulatory mechanisms of this ubiquitin ligase, including substrate selectivity, transition from the latent to the active conformation and auto-ubiquitination.

8.2. Smad phosphorylation affects substrate binding in Nedd4L and related proteins

Prior to the work presented hereby, *in vivo* evidence available in the literature and expanded with data discovered by our collaborators in the group of Dr Joan Massagué suggested that the ability of Nedd4L to bind the Smad2/3 inter-domain linker region depends on its phosphorylation state (Gao et al., 2009). Indeed, phosphorylation only by CDK8/9 favours the association of Smad2/3 with Pin1, and Nedd4L is the preferred binding partner only after GSK3- β phosphorylates the sequence too. An analogous scenario is found in the case of Smad1, whose CDK8/9 phosphorylation favours its association with YAP and the subsequent GSK3- β action tips the scale in favour of Smurf1 (Alarcón et al., 2009). This switch permits the R-Smads to transduce the TGF- β signal in association with Pin1 and YAP before Nedd4L and Smurf1 terminate it labelling them for destruction.

Therefore, in this case substrate phosphorylation acts as a regulatory mechanism of the activity of Nedd4L and the closely related ligase Smurf1. We set to investigate its biophysical and structural fundamentals with the aim of contributing to its detailed understanding.

Two sites are available for WW domain binding on the R-Smad linker: a PPxY motif that is the consensus sequence for group I WW domain recognition and a phospho-S/TP cluster located next to it, which is recognized by group IV WW domains. The whole segment containing both sites is about 35 amino acids long. We hypothesized that proteins containing more than one WW domain (Nedd4L, Smurf1 and YAP) may bind to both sites simultaneously using tandem combinations of these domains, even though all WW domains in these proteins had so far been classified as group I domains. In order to test this hypothesis, we planned to perform ITC and NMR binding experiments involving single (Pin1) or tandem WW domain constructions and peptides representing different phosphorylation states of the R-Smad linker region containing both the PY and the phospho-S/TP sites.

The work presented in this thesis (CHAPTER 4) corresponds to the synthesis of two sets of phosphopeptides representing different phosphorylation states of Smad1 and Smad3. The syntheses were performed by SPPS using Fmoc chemistry, which was used to obtain peptides up to 36 residues long and including up to 4 phosphorylated residues with purities above 90% in most cases and in sufficient amounts for our binding experiments.

The synthesis of these peptides was not trivial. Most of the sequences were considerably long thus limiting the yield of the final product. In addition, the coupling efficiency of phosphorylated amino acids is often low since their big size hinders the coupling and the amounts used are limited due to their high cost. Taking this into account, syntheses were performed using the 0.2 – 0.25 mmol scales in order to generate enough amounts of peptides for our experiments. Moreover, coupling and deprotection steps were systematically checked in a careful manner through colorimetric and mini-cleavage tests, and alternative strategies were applied on residues requiring re-coupling.

Some of the difficulties were particularly related to phosphopeptide synthesis. We observed the β -elimination of phosphate groups caused by the nucleophilic attack by piperidine during microwave-assisted backbone amide deprotection, which was solved using DBU instead of piperidine. During the cleavage step, we encountered the alkylation of methionine residues by the phosphate protecting group benzyl (Bzl). Bzl is a much more common protecting group in Boc than in Fmoc chemistry, and 2-PySH is used as a scavenger for its reactive species. We successfully avoided Bzl methionine alkylation by adding 2-PySH to our TFA cleavage mixtures.

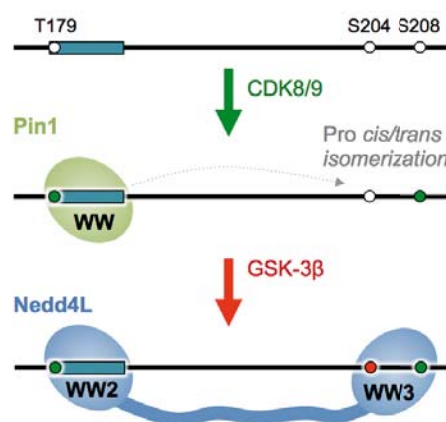
Related to the low coupling efficiency of phospho-amino acids, we eventually obtained considerable amounts of truncated by-products lacking one or more of them that were seriously compromising the purity of the final product. Substituting TFA by formic acid in the mobile phase in combination with a Waters SunFire™ PREP C18 OBD™ column as the stationary phase solved this problem.

After obtaining the peptides in sufficient amounts and purity, they were used in binding and structural studies with the WW domains of Nedd4L and Pin1 (Smad3 peptides) as well as Smurf1 and YAP (Smad1 peptides). Dr Nina Görner and Dr Eric Aragón in our laboratory carried out these experiments, which are reported in their respective doctoral theses.

In the case of Smad3, both Nedd4L WW2 and Pin1 WW domains bind the pT[PY] motif with similar affinities. In addition, Nedd4L binds the phospho-S/TP cluster

using its WW3 domain with highest affinity when both CDK8/9 and GSK3- β sites are phosphorylated. Remarkably, Nedd4L binds Smad3 with all proline residues in the phospho-S/TP cluster in the *trans* configuration. Therefore, upon CDK8/9 phosphorylation both proteins compete for the binding of the PY, and Pin1 may condition the binding of Nedd4L through *cis/trans* isomerisation of the proline residues in the S/TP cluster. GSK3- β favours the interaction of Smad3 with Nedd4L, which becomes the preferred binding partner in this situation (**Figure 8.1A**). SGK1 phosphorylation of the WW2-WW3 linker in Nedd4L inhibits its interaction with Smad3 (Gao et al., 2009), probably due to a change on the substrate binding capabilities of the WW domain pair.

A TGF- β cytokine-activated pathway – Smad3 linker



B BMP cytokine-activated pathway – Smad1 linker

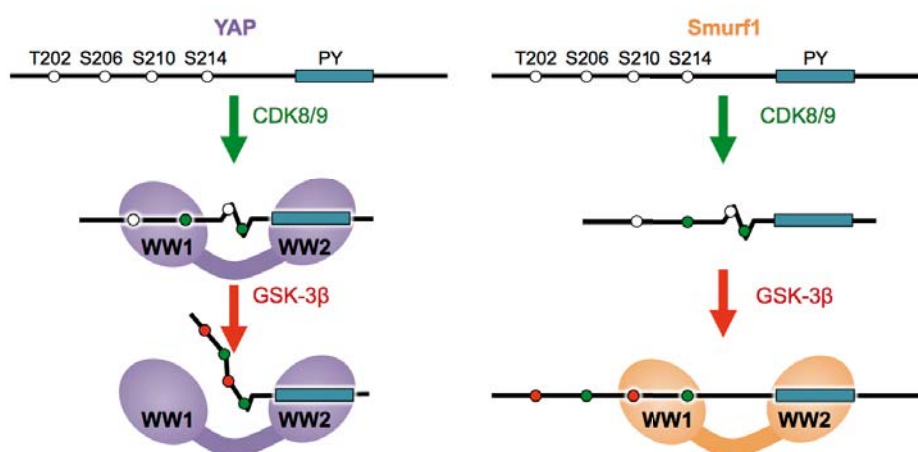


Figure 8.1: Structural basis of Smad regulation through linker phosphorylation. (A) Phosphorylation of position T179 in the linker region of Smad3 by CDK8/9 creates a docking site for Pin1 WW domain, which contributes to BMP signal transduction and conditions Nedd4L binding to Smad3 through proline *cis/trans* isomerization in the S/TP cluster. GSK3- β phosphorylation of position S204 enhances Nedd4L binding. (B) YAP binds the Smad1 linker when both CDK8/9 positions are phosphorylated and the linker adopts a conformation characterized by two β -turns. Smurf1 binds the sequence after GSK3- β phosphorylation, which locks the linker in an extended form.

YAP binds Smad1 with the highest affinity when both CDK8/9 sites are phosphorylated. Smad1 adopts a configuration characterized by two β -turns, which enables the simultaneous interaction of the WW2 domain with the PY motif and of the WW1 domain with the distal phosphorylated position (**Figure 8.1B**). Phosphorylation of GSK3- β positions locks the peptide in an extended conformation and generates electrostatic repulsion with YAP. Differently, Smurf1 coordinates Smad1 in this extended conformation, using its WW2 domain to interact with the PY motif and its WW1 domain to coordinate the proximal phosphorylated positions in the S/TP cluster (**Figure 8.1B**). CKIP-1 favours the catalytic activity of Smurf1 by binding the WW1-WW2 linker (Lu et al., 2008), probably inducing an optimal orientation of the domains that would enhance their substrate binding capabilities.

WW domain cooperativity in substrate recognition had been previously reported for a few E3 ligases of the Nedd4 family, including Nedd4L (Bhattacharya et al., 2008; Jennings et al., 2007; Wiemuth et al., 2010). However, the Pin1 WW domain was so far the only one known to bind phospho-S/TP motifs, and its phosphate coordinating residues are not conserved in type I WW domains such as those found in Nedd4L, Smurf1 and YAP. Our results show that the regulation of the R-Smads is concentrated in the short linker region depending on its phosphorylation state. The detailed structural and geometrical characterization of the coordinate recognition of the PY motif and the phospho-S/TP cluster located at the R-Smad linker by WW domain tandems represents a breakthrough in the understanding of substrate recognition by Nedd4L and related proteins, which is an actual mechanism regulating the activity of these molecules.

8.3. Nedd4L changes from a latent to an active conformation in response to IP₃/Ca²⁺

The transient adoption of latent, self-inhibited conformations that can be activated in response to defined intracellular events is a common activity regulatory mechanism among E3 ligases of the Nedd4 family. Most of its human members (6 out of 9) have been shown to protect their substrates and themselves from undesired ubiquitination using this strategy, including Itch (Gallagher et al., 2006), Smurf2 (Wiesner et al., 2007), Smurf1 (Wan et al., 2011), WWP2 (Soond and Chantry, 2011), Nedd4 and Nedd4L (Wang et al., 2010).

Even though some similarities can be noted, each ligase has apparently developed its own particular strategy to attain transient inactivity. Apart from Itch, whose latent conformation is achieved through contacts between the HECT domain and the

central area of the ligase where the WW domains are located, contacts between the HECT and the C2 domains have been found in all other cases. Some of these are intra-molecular as it is the case for Smurf2, and others are inter-molecular (Smurf1 and WWP2), while this remains uncertain in the case of Nedd4 and Nedd4L.

Different mechanisms have been found to trigger the transition from these closed conformations to the active forms of the ligases including phosphorylation, substrate binding and cofactor binding. Some of them may in fact co-exist and complement each other. For instance, Itch is activated by phosphorylation (Gallagher et al., 2006) and also binding to the accessory proteins NDFIP1 and 2 (Nedd4 family interacting proteins 1 and 2) (Mund and Pelham, 2009). Nedd4 is activated by NDFIP1/2 binding as well, but cytoplasmic increases of the Ca^{2+} concentration also render the ligase active (Wang et al., 2010).

In the case of Nedd4L, the C2-HECT contacts were detected *in vivo* at basal cytoplasmic Ca^{2+} levels using co-IP experiments (Wang et al., 2010). Low ubiquitinating activity of the ligase was observed in these conditions, both of its substrate ENaC and itself (auto-ubiquitination). After an induced increase of the cytoplasmic Ca^{2+} concentration, the self-inhibitory C2-HECT contacts were not detected any more and the ligase was able to ubiquitinate its substrates and itself. This is accompanied by the re-location of the ligase to the plasma membrane, which is mediated by the C2 domain (Garrone et al., 2009).

Using recombinant constructs of the C2 and the HECT domains, we could detect by NMR their interaction in Ca^{2+} -free conditions through the establishment of a high molecular weight complex with unfavourable relaxation properties. In contrast, the interaction was not detected in the absence of Ca^{2+} . In order to gain further structural detail, we used a deuterated C2 domain sample with better relaxation properties to characterize the contact surface with the HECT domain (**Figure 8.3A**).

Ca^{2+} binds very tightly to the C2 domain, occupying two binding sites with affinities in the nanomolar and the low micromolar ranges respectively. These sites are indeed two acidic pockets in the surface of the domain, so Ca^{2+} binding has an effect on its surface charge distribution.

Notably, the HECT domain and Ca^{2+} ions bind to overlapping areas on the surface of the C2 domain (**Figure 8.3B**). Based on these observations, we suggest that Ca^{2+} displaces the HECT domain from its binding site by tightly binding the C2 domain. Both the change in C2 surface electrostatics as well as steric effects may contribute to this mechanism.

Cytoplasmic increases of Ca²⁺ concentration are often initiated by external stimuli, which activate plasma membrane receptors resulting in the release of second messengers able to mobilize the intracellular Ca²⁺ reserves mainly localized in the endoplasmic reticulum. The most extensively characterized Ca²⁺-mobilizing second messenger is IP₃, which results from the hydrolysis of the membrane phospholipid PIP₂ by the G-protein coupled receptor (GPCR)-activated phospholipase C (PLC). C2 domains act as membrane-localizing modules in many proteins through interactions with membrane phospholipids including PIP₂, which are often Ca²⁺-dependent as in the case of the C2 domain of Nedd4 (Plant et al., 1997). The PIP₂ moiety interacting with C2 domains is chemically identical to the soluble IP₃. Therefore, we hypothesized that PIP₂ may play a role both in the activation of the ligase and contribute to its relocation to the plasma membrane.

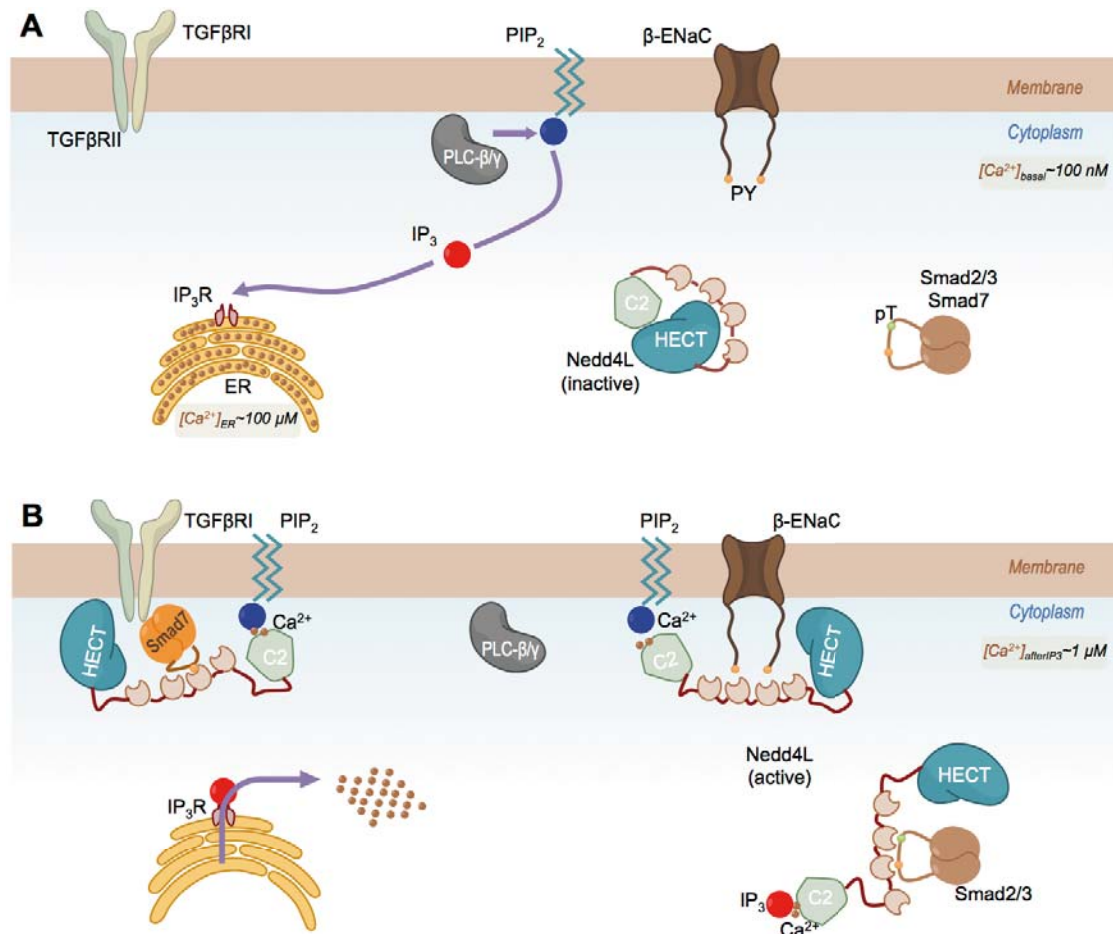


Figure 8.2: Model for the role of IP₃/Ca²⁺ signalling in Nedd4L activity regulation. (A) The basal Ca²⁺ concentration in the cytoplasm is about 100 nM. In this situation Nedd4L exists in its latent conformation, which is mediated by HECT-C2 contacts. In response to external stimuli PLC-β/γ initiates the Ca²⁺ signalling cascade by cleaving PIP₂ and releasing IP₃ into the cytoplasm, which diffuses to the ER and binds its receptors (IP₃R). (B) This provokes an increase in Ca²⁺ concentration to about 1 μM, resulting in the activation of Nedd4L. Depending on the IP₃/PIP₂ ratio the ligase can be localized and target its substrates in the cytoplasm or the plasma membrane.

Our data reveals that IP₃ binds the C2 domain of Nedd4L in a Ca²⁺-dependent manner. In addition, IP₃ binds an area on the C2 domain that overlaps with the binding sites of Ca²⁺. In the absence of Ca²⁺, the IP₃ binding surface on the C2 domain is a negatively charged patch. Because IP₃ is a negatively charged molecule at physiological pH, this results in electrostatic repulsion. Ca²⁺ binding to the C2 domain switches the surface electrostatics of this area from negative to positive, thus bridging the interaction with IP₃.

In addition, the IP₃ binding site on the C2 domain also overlaps extensively with that of the HECT domain (**Figure 8.3B**). Therefore, Ca²⁺ changes the preferences of the C2 domain from interacting with the HECT domain to binding IP₃/PIP₂. In the cell, IP₃ binding may contribute to the stable activation of Nedd4L by further interfering in the C2-HECT contacts.

Based on these data, we propose that increments of the cytoplasmic Ca²⁺ concentration act as a switch modulating the population of active and inactive conformations of the ligase as well as its intracellular localization (Escobedo et al., 2014). Through C2 domain-mediated interactions and depending on the PIP₂/IP₃ ratio, the ligase can be localized and target its substrates in the cytoplasm (through interactions with the soluble, PLC released IP₃) where it targets Smad2/3, or in the membrane's proximity (bound to PIP₂) where it targets ENaC and other ion channels as well as the TGF-β cytokine receptor (with Smad7 scaffolding the interaction) (**Figure 8.2**). Ca²⁺-dependent binding to other membrane phospholipids such as phosphatidylcholine and phosphatidylserine also contribute to the membrane localization of the ligase.

All other ligases in the Nedd4 family have a conserved C2 domain with Ca²⁺ and phospholipid binding capabilities. Therefore, the activating and membrane relocating effects of IP₃/Ca²⁺ signalling that we describe for Nedd4L may be common to some of these ligases, and can complement and co-exist with other mechanisms triggering the transition from their latent to their active conformations.

8.4. A crossed mechanism for Nedd4L ubiquitination

E3 ubiquitin ligases of different classes and types (including both HECT-type and RING ligases) have been found to auto-ubiquitinate (De Bie and Ciechanover, 2011). In most cases, this phenomenon constitutes a negative feedback control mechanism, as the ligases label themselves for proteasome-mediated degradation. Different mechanisms have been described, which can involve a single molecule that labels itself in an intra-molecular manner (Galan and Peter, 1999; Huibregtse et al., 1995),

or rely in the crossed labelling of one molecule by another one (inter-molecular) (Hu and Fearon, 1999; Nuber et al., 1998). Both examples have been found in orthologous proteins of the Nedd4 family in yeast (Rsp5 auto-ubiquitinates in an intra-molecular way) and rat (p100 uses an inter-molecular strategy) (Huibregtse et al., 1995).

Nedd4L has an LPxY motif located in its HECT domain that is highly conserved along evolution and in other members of the Nedd4 family (HECT-PY motif). This sequence is similar to the canonical PPxY motif bound by type I WW domains. Indeed, the WW domains of Nedd4L bind this motif when the HECT domain is unfolded (Bruce et al., 2008). Several Nedd4L and Nedd4 targets contain uniquely an LPxY motif (Persaud et al., 2009). There are other examples of WW domains binding to LPxY motifs, including the WW3 domain of Nedd4 orthologous in drosophila (Kanelis et al., 2006), the WW domains of Rsp5 (Hesselberth et al., 2006) and YAP WW1 domain (Pires et al., 2001).

Nedd4L has been reported to auto-ubiquitinate and label itself for proteasome-mediated degradation (Mund and Pelham, 2009). We hypothesized that the LPxY motif in Nedd4L may play a role in the auto-ubiquitination mechanism of the ligase. This would be attained by the recognition of this motif on one ligase by a WW domain of another ligase.

Fundamental to PY recognition by type I WW domains is the coordination of the tyrosine residue in the motif by a hydrophobic pocket in the second loop of the domain. However, the tyrosine residue in the HECT-PY motif appears buried in the hydrophobic core of the HECT domain in its available crystal structure. This is also observed in the crystal structures of Itch, Nedd4, WWP1 and Smurf2. Thus, the putative recognition of the HECT-PY motif by a WW domain would require the eventual exposure of this residue to the solvent.

Using NMR, we tested the accessibility for WW domain binding to the Nedd4L HECT-PY motif when it is part of the folded HECT domain, which would rely on the spontaneous exposure of the tyrosine residue to the solvent. We prepared a recombinant construct of the HECT C-lobe (which contains the motif) and we titrated the WW2 and WW3 domains separately. In this situation, we could not observe the specific recognition of the motif. The dynamics of the HECT C-lobe in the fast time-scale were characterized through relaxation measurements, which revealed low flexibility of the area where the HECT-PY motif is located. Even though no signs of slower exchange, such as peak broadening, were observed in our NMR experiments suggesting that the exposure of the tyrosine residue could happen in a

slower time-scale, we tested that by performing a molecular dynamics simulation at 310 K. The HECT-PY tyrosine residue was not exposed along the calculated 120 ns trajectory.

Therefore, we hypothesized that the HECT-PY tyrosine residue may indeed be exposed when the protein's fold is destabilized, yielding a defective protein. This could occur naturally when the protein ages in the cell and needs to be degraded. We tried to experimentally induce the destabilization of the HECT domain, but we could not trap the protein in a semi-destabilized form while avoiding its precipitation. Instead, we performed a simulated annealing molecular dynamics simulation where the structure of the domain was destabilized using a temperature ramp from 310 to 600 K along a 40 ns trajectory. In this calculation we could observe the exposure of the HECT-PY tyrosine residue while the HECT C-lobe roughly maintains its pseudo-native fold.

In order to study the interaction of the HECT-PY motif with the individual WW domains of Nedd4L, we synthesized a peptide with its sequence and a few surrounding residues by SPPS. The ITC interaction experiments revealed that the WW3 domain binds the motif with an affinity in the low micromolar range, which is comparable to the binding of Nedd4L to some of its substrates such as Smad3 and Smad7 (Aragón et al., 2012; Aragón et al., 2011). As discussed in Section 8.2, Nedd4L utilizes its WW2 domain to bind conventional PPxY sequences whereas it binds phosphorylated motifs using its WW3 domain. The presence of negatively charged amino acids surrounding the LPxY sequence of the HECT-PY motif may explain the preference of the WW3 domain for this site.

Further structural insights into the WW3-HECT-PY interaction were gained by solving the solution structure of the complex using multidimensional NMR (**Figure 8.3C**). Numerous contacts between the peptide and the protein define the structure, where both hydrophobic and electrostatic interactions can be observed. This supports our interpretation that negatively charged amino acids surrounding the LPxY motif play a role in the interaction.

Based on these results, we suggest that the HECT-PY motif may take part in the auto-ubiquitination mechanism of Nedd4L, which is driven by a crossed intermolecular ubiquitination process (**Figure 8.3C**). Protection of the PY site in a natively folded, functional protein will prevent its degradation. Destabilization of the protein possibly caused by aging (or some other event such as a PTM) would lead to a natural unfolding process. The HECT-PY motif is exposed at the early stages of the unfolding thus turning the destabilized protein into a target of a fully functional

Nedd4L molecule, which would selectively ubiquitinate its damaged counterpart for its disposal in the ubiquitin-proteasome pathway (Escobedo et al., 2014).

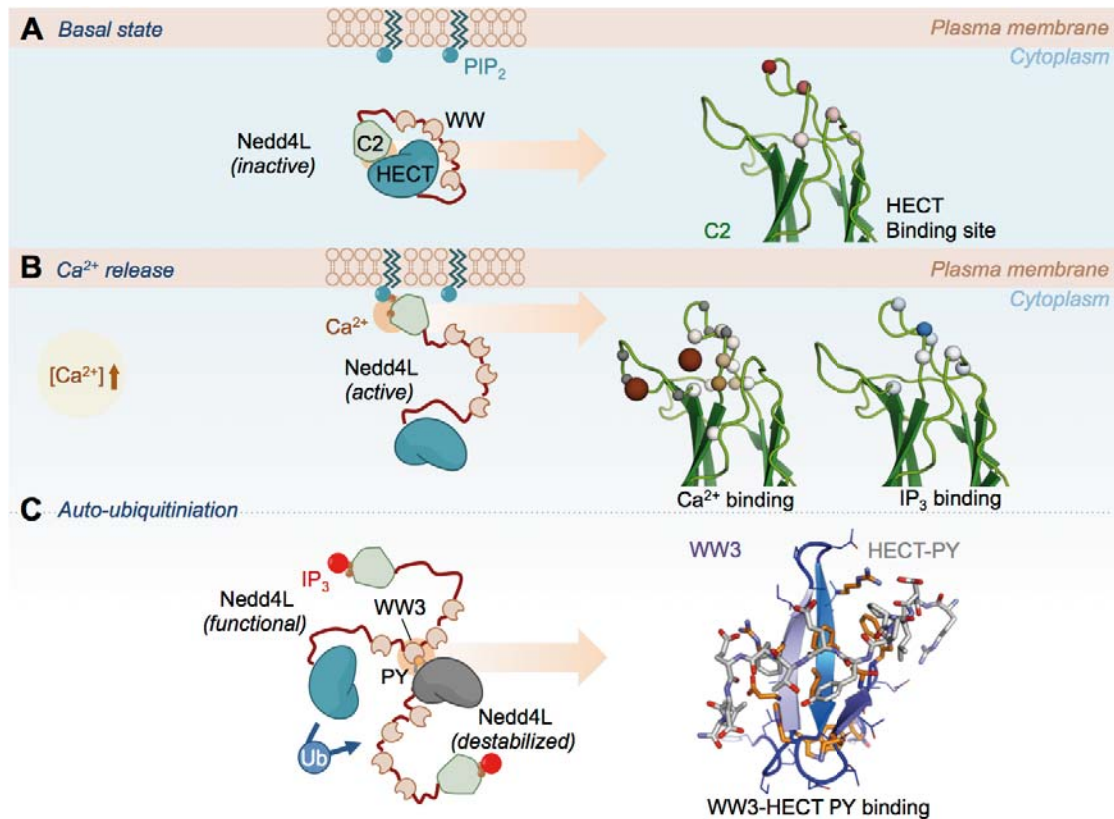


Figure 8.3: Proposed activation and auto-ubiquitination mechanisms for Nedd4L. (A) The latent conformation of Nedd4L is mediated by HECT-C2 contacts. (B) When the cytoplasmic Ca²⁺ concentration increases, Ca²⁺ ions tightly bind to the C2 domain, displace the HECT domain and mediate the interaction with IP₃/PIP₂. (C) Destabilization of Nedd4L is accompanied by the exposure of the HECT-PY motif, which is recognized by the WW3 domain of a functional counterpart, defining a crossed auto-ubiquitination mechanism.

8.5. Concluding remarks

The experimental work reported in this thesis provides a structural and biophysical perspective that contributes at expanding the current knowledge on different mechanisms regulating the HECT-type E3 ubiquitin ligase Nedd4L. These include the external regulation by protein kinases of its substrate recognition capabilities as well as the auto-regulation of its activity and turnover. Bearing in mind the similarities with the other eight human members of the Nedd4 family of ubiquitin ligases, some of the findings reported for Nedd4L may also be applicable to other closely related proteins. The general conclusions arising from these findings can be summarized as follows:

1. **The binding of Nedd4L and related ligases to their targets can be modulated by substrate phosphorylation.** CDK8/9 phosphorylation of Smad3 permits its association with Pin1, which cooperates with TGF- β signal transduction. Subsequent GSK3- β phosphorylation is required for Nedd4L to bind Smad3 using a PY motif and phosphate-coordinating WW domain tandem arrangement. A parallel situation is observed for Smad1, which cooperates with YAP and is targeted by Smurf1.
2. Ca^{2+} and the HECT domain bind to overlapping areas on the surface of the C2 domain. The tight binding of Ca^{2+} displaces the HECT domain from its binding site by steric and electrostatic effects, thus triggering the **transition from the latent to the active conformation** of Nedd4L.
3. IP_3 is chemically identical to the cytoplasmic moiety of the membrane phospholipid PIP_2 . Ca^{2+} binding to the C2 domain creates a positively charged patch that allows its interaction with IP_3 (and thus PIP_2), which occupies the HECT binding site and hampers the auto-inhibitory HECT-C2 contacts. **The IP_3/PIP_2 ratio controls the location of the ligase**, distributing it between the cytoplasm and the plasma membrane respectively.
4. A conserved PY motif in the HECT domain is protected in functional molecules of Nedd4L. Aging of the protein within the cell could initiate its natural unfolding. At the early stages of fold destabilization, the exposure of the PY motif allows its recognition by a functional counterpart, which labels the damaged protein for proteasomal degradation using a **crossed auto-ubiquitination mechanism**.

Part VI.

Appendix

CHAPTER 9. APPENDIX	197
9.1. Chromatographic and MS characterization of the peptides	197
9.2. Reconstruction of the β 1- β 2 loop in the crystal structure of the C2 domain	203
9.3. HSQC spectra of the free and HECT-PY bound WW1, WW2 and WW4 domains	204
9.4. Sequence alignment of the HECT domain of Nedd4L in metazoans	206

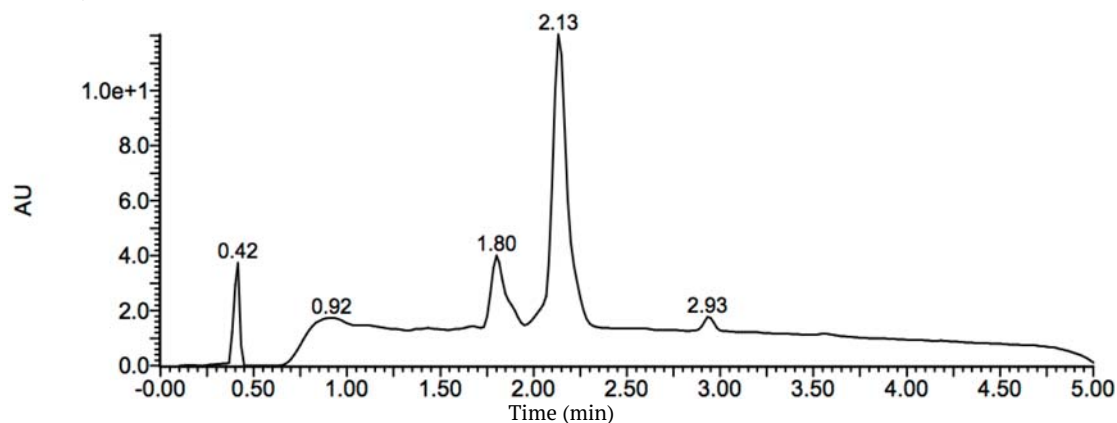
CHAPTER 9. Appendix

9.1. Chromatographic and MS characterization of the peptides

Smad1 pS214 (Section 4.4.1)

HPLC details: column XSelect C₁₈ 4.6x50 3.5µm; gradient 0-100% ACN over 6.5min

Purity: 77.62%

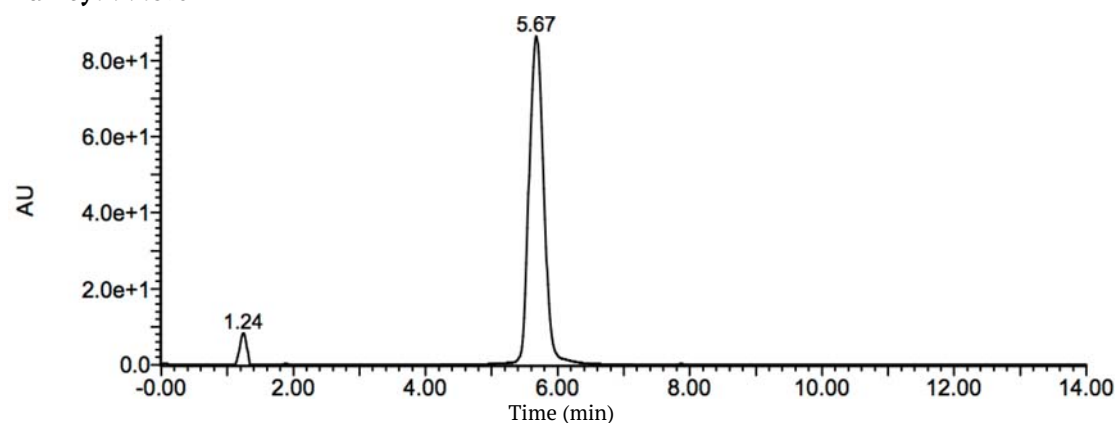


MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
ESI-Q-TOF	Double	1416.97	1417.78
	Triple	944.98	945.23

Smad1 pS210-pS214 (Section 4.4.2)

HPLC details: column SunFire C₁₈ 4.6x100 3.5µm; gradient 0-100% ACN over 14min

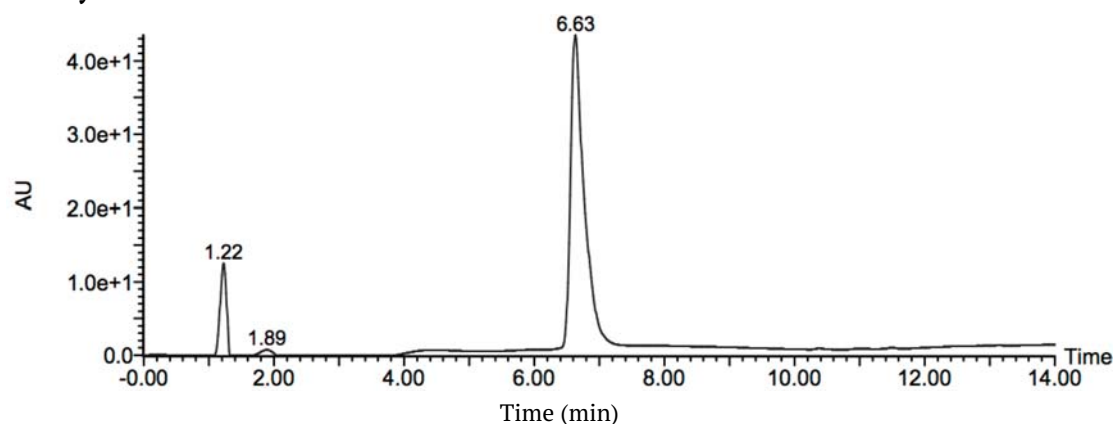
Purity: 99.3%



MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
MALDI-TOF	Single	2898.89	2894.80
ESI-Q-TOF	Double	1449.95	1450.63
	Triple	966.96	967.11

Smad1 pS206-pS214 (Section 4.4.3)HPLC details: column SunFire C₁₈ 4.6x100 3.5µm; gradient 0-100% ACN over 14min

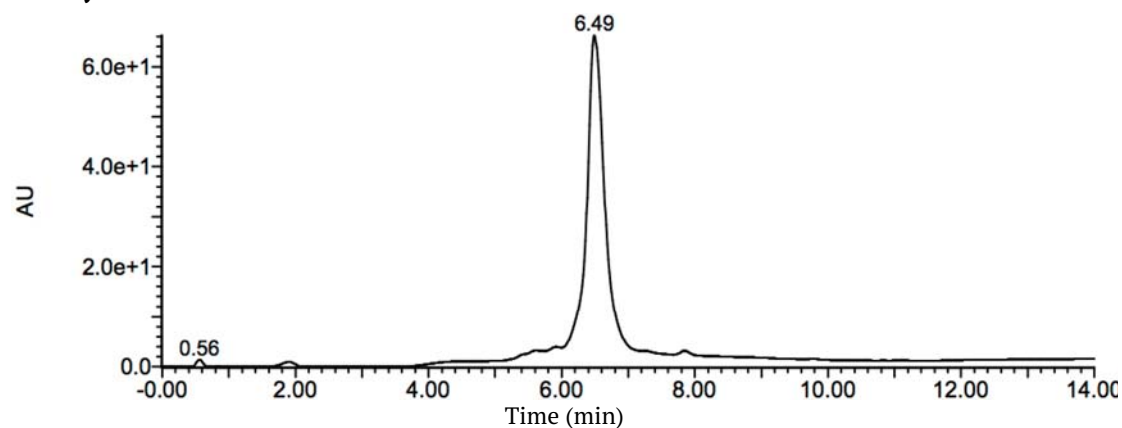
Purity: 98.9%



MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
MALDI-TOF	Single	3769.80	3769.70
ESI-Q-TOF	Double	1885.40	1886.20
	Triple	1257.27	1257.81

Smad1 pT202-pS206-pS210-pS214 (Section 4.4.4)HPLC details: column SunFire C₁₈ 4.6x100 3.5µm; gradient 0-100% ACN over 14min

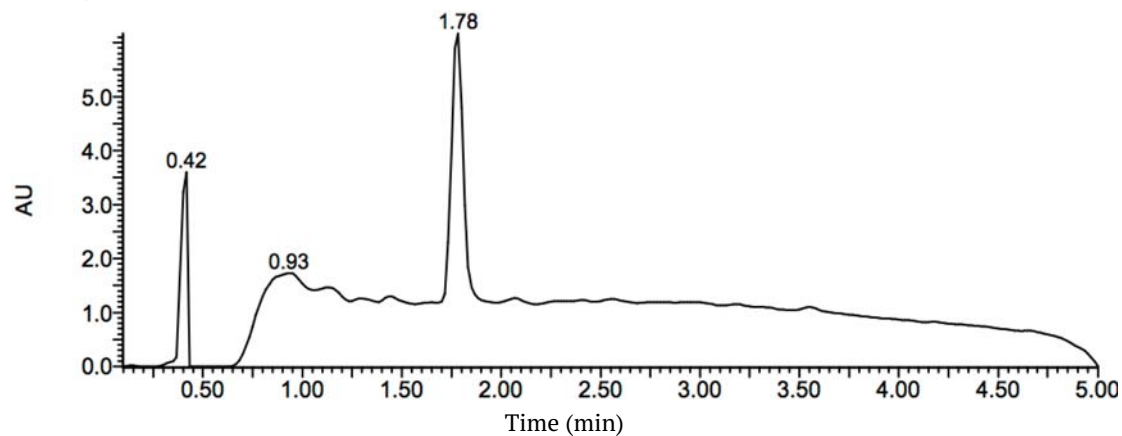
Purity: 97.5%



MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
ESI-TOF	Triple	1399.99	1399.55

Smad3 NP (Section 4.5.1)HPLC details: column XSelect C₁₈ 4.6x50 3.5µm; gradient 0-100% ACN over 6.5min

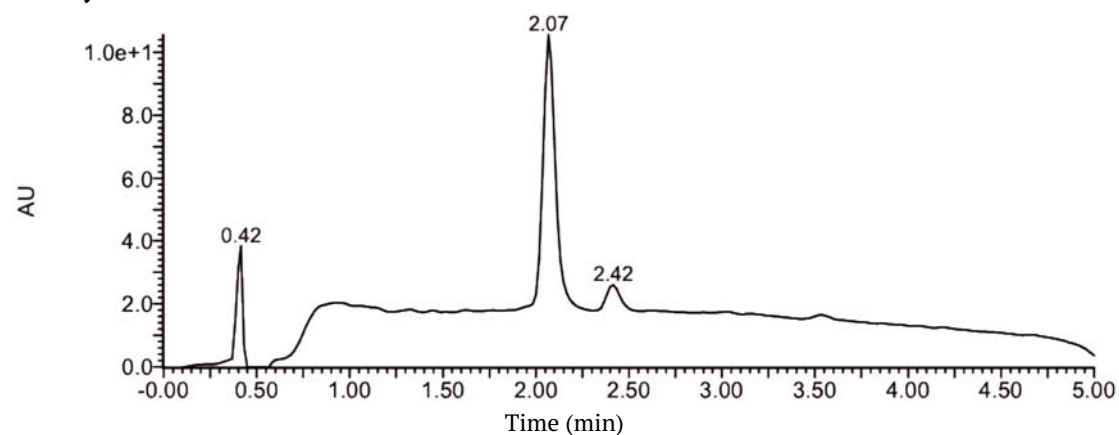
Purity: 97.7%



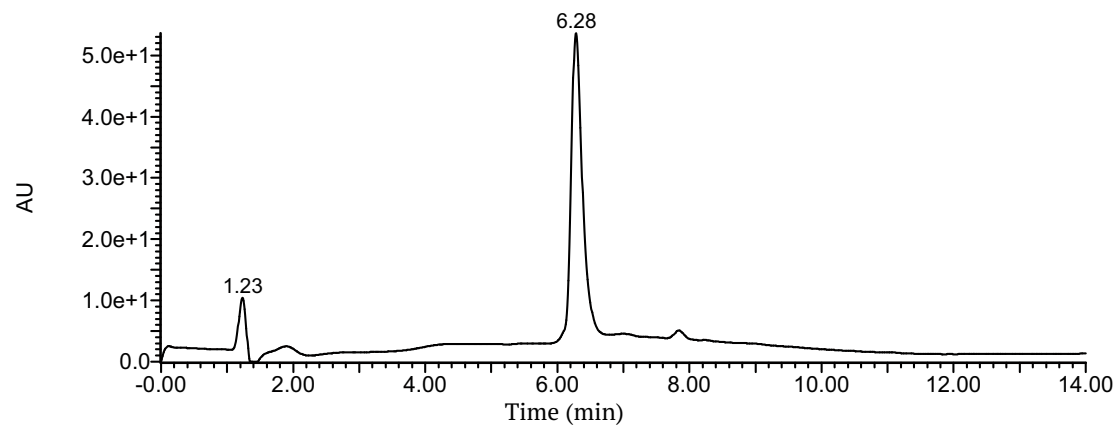
MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
MALDI-TOF	Single + Na ⁺	1478.57	1477.90
	Single + K ⁺	1493.57	1493.80
ESI-Q-TOF	Single	1455.57	1455.88
	Double	728.29	728.61

Smad3 pT179 (Section 4.5.1)HPLC details: column XSelect C₁₈ 4.6x50 3.5µm; gradient 0-100% ACN over 6.5min

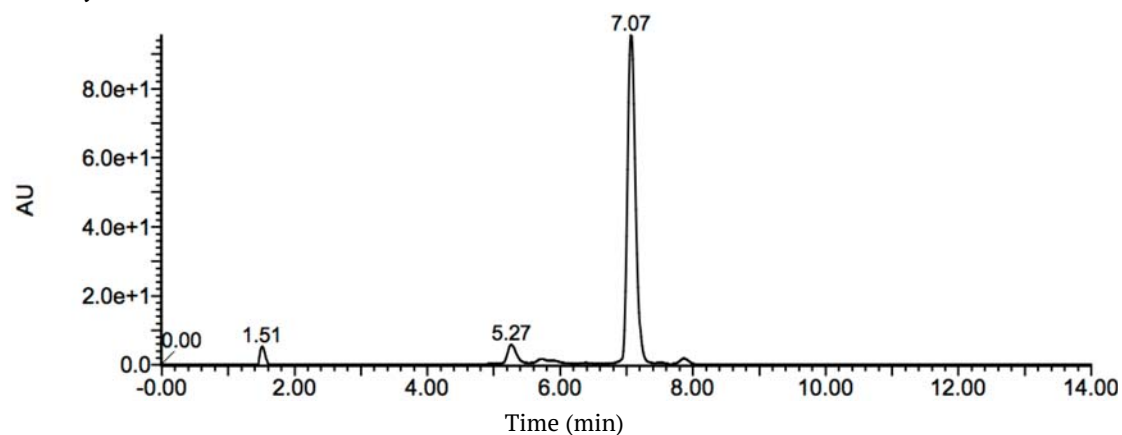
Purity: 90.3%



MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
ESI-Q-TOF	Single	1536.58	1536.70
	Double	768.79	768.67

Smad3 pT179-pS204-pS208 (Section 4.5.2)**HPLC details:** column SunFire C₁₈ 4.6x100 3.5 μ m; gradient 0-100% ACN over 14min**Purity:** 97.5%

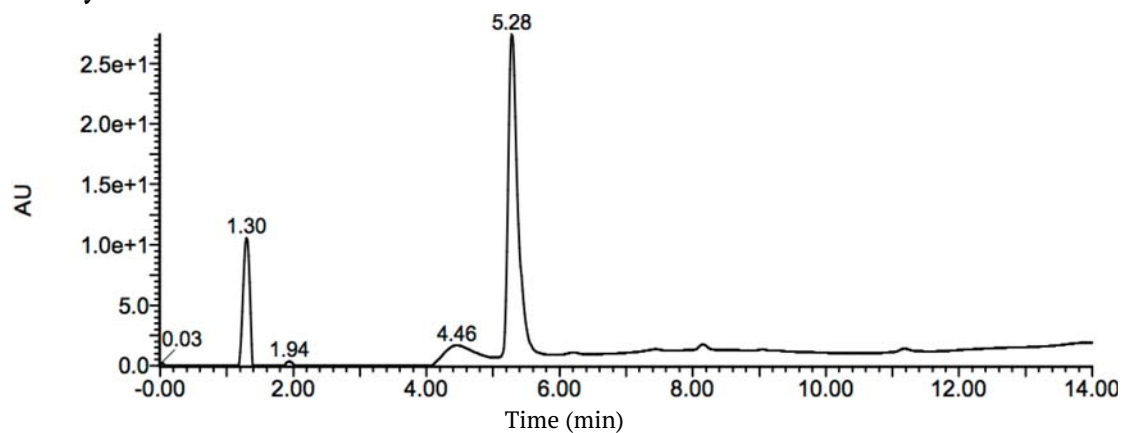
MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
MALDI-TOF	Single	4102.11	4099.60
ESI-Q-TOF	Triple	1368.04	1368.18
	Quadruple	1026.28	1026.43

GYPGA-COS-Bzl (Section 4.6.1)**HPLC details:** column SunFire C₁₈ 4.6x100 3.5 μ m; gradient 0-100% ACN over 14min**Purity:** 89.6%

MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
ESI-TOF	Single	611.71	612.21

Smad3 pS204-pS208 (Section 4.6.2)HPLC details: column SunFire C₁₈ 4.6x100 3.5 μ m; gradient 0-100% ACN over 14min

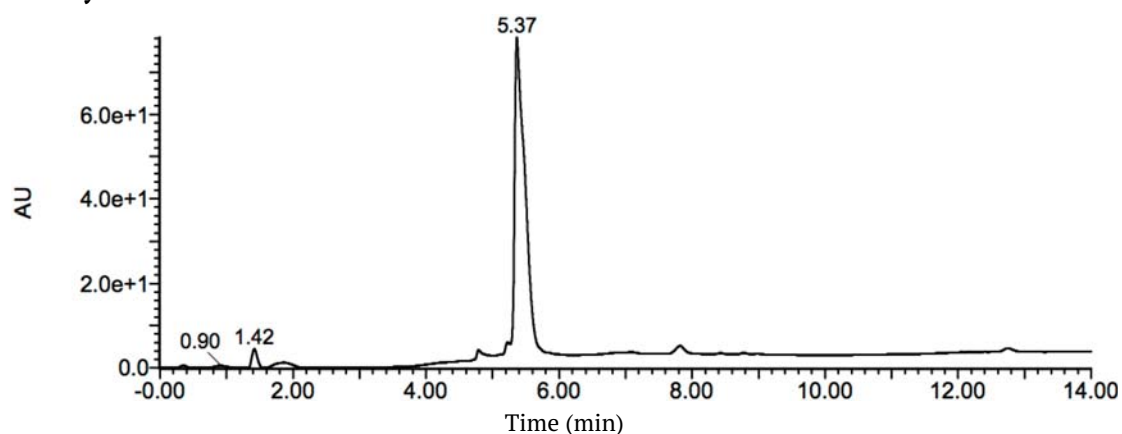
Purity: 77.2%



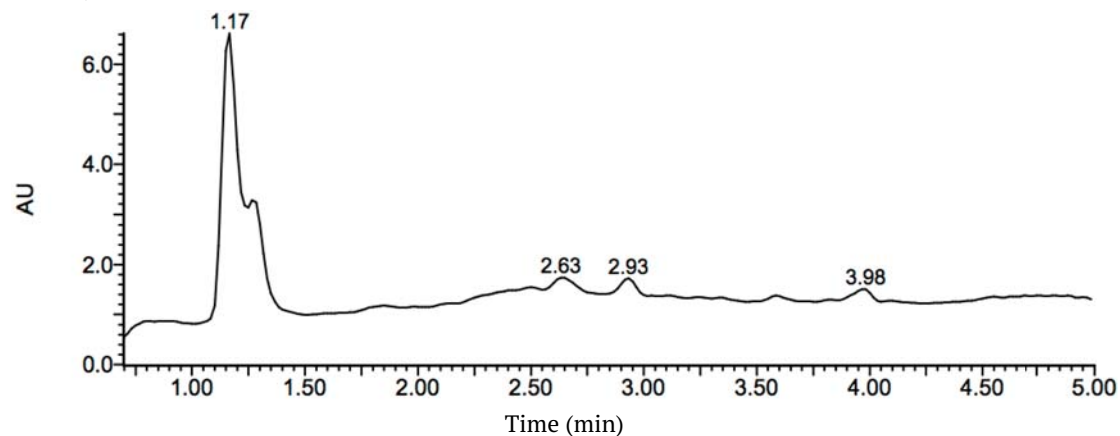
MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
ESI-Q-TOF	Double	1156.14	1156.36
	Triple	771.09	771.67

GYPGA-Smad3 pS204-pS208 (Section 4.6.2)HPLC details: column SunFire C₁₈ 4.6x100 3.5 μ m; gradient 0-100% ACN over 14min

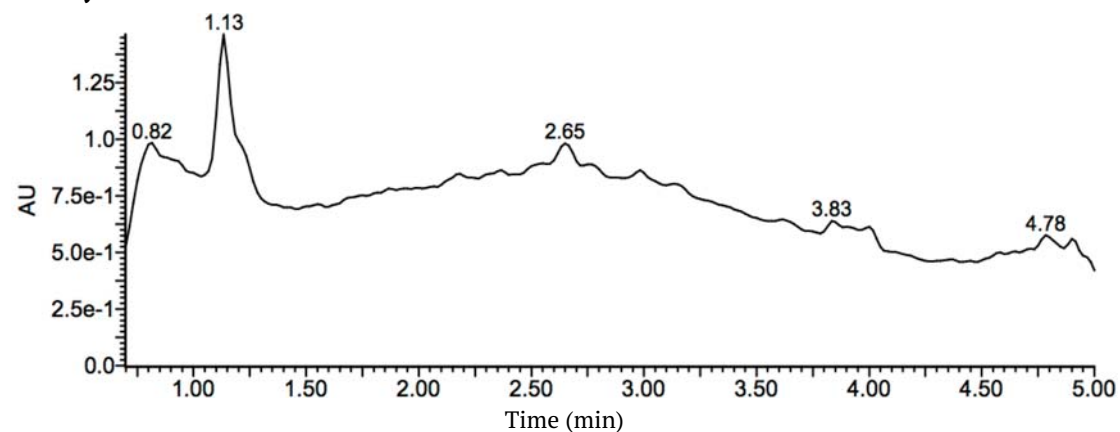
Purity: 92.1%



MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
ESI-Q-TOF	Double	1399.90	1400.58
	Triple	933.60	934.16

Cys-Smad3 pS208 (Section 4.6.3)**HPLC details:** column XSelect C₁₈ 4.6x50 3.5µm; gradient 0-100% ACN over 6.5min**Purity:** 53.1%

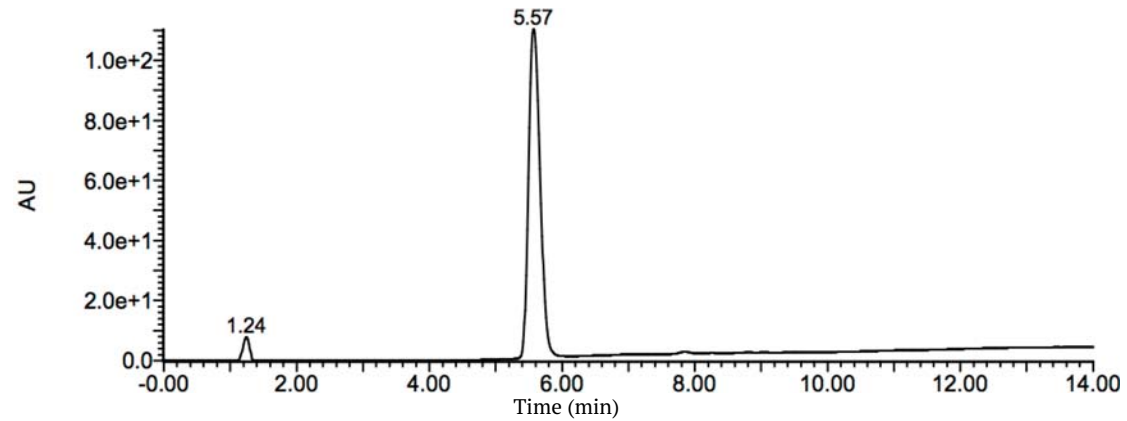
MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
ESI-Q-TOF	Single	976.99	977.61
	Double	489.00	489.42

Ala-Smad3 pS208 (Section 4.6.3)**HPLC details:** column XSelect C₁₈ 4.6x50 3.5µm; gradient 0-100% ACN over 6.5min**Purity:** 45.0%

MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
ESI-Q-TOF	Single	944.93	945.42
	Double	472.97	473.23

Nedd4L HECT-PY (Section 6.2.1)HPLC details: column SunFire C₁₈ 4.6x100 3.5 μ m; gradient 0-100% ACN over 14min

Purity: 98.97%



MS method	Ionization	<i>m/z</i> Calculated	<i>m/z</i> Observed
ESI-Q-TOF	Single	2062.28	2062.66
	Double	1031.64	1032.15
	Triple	688.08	688.71

9.2. Reconstruction of the $\beta 1$ - $\beta 2$ loop in the crystal structure of the C2 domain

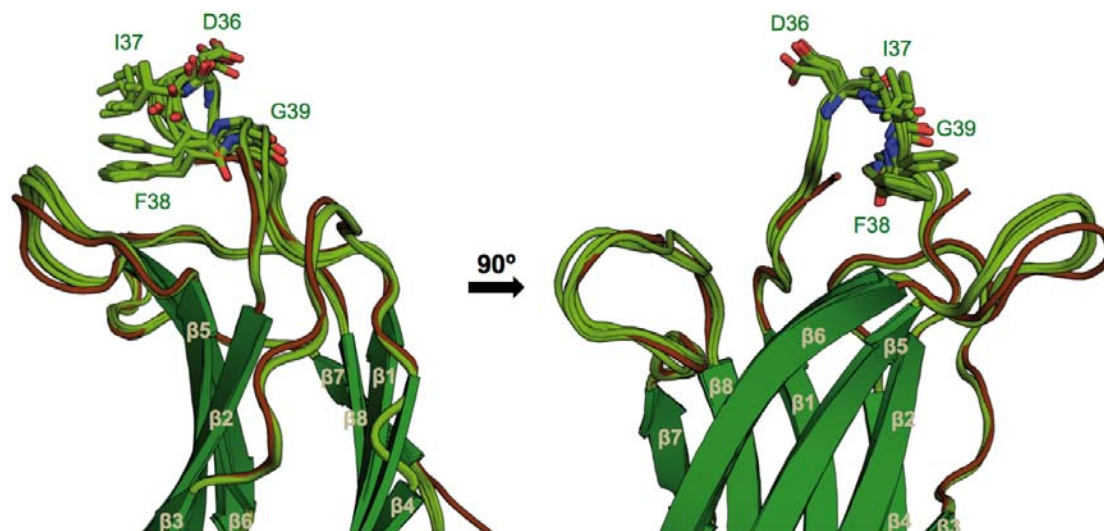
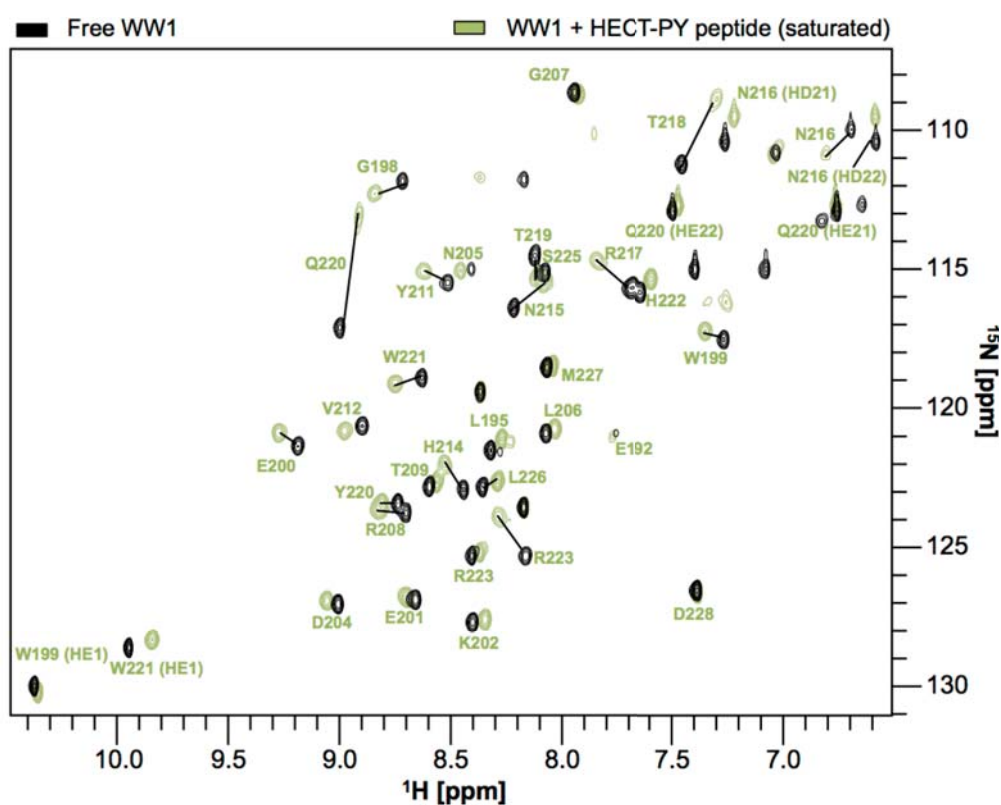


Figure 9.1: Modelling of the C2 domain $\beta 1$ - $\beta 2$ loop using Rosetta. Ensemble of the five lowest energy decoys of the largest cluster selected for the analysis (light green loops) structurally aligned to the experimentally determined crystal structure of the domain (brown loops). Left panel: three-dimensional orientation as in all figures presented above. Right panel: 90° turn around the z-axis to the right.

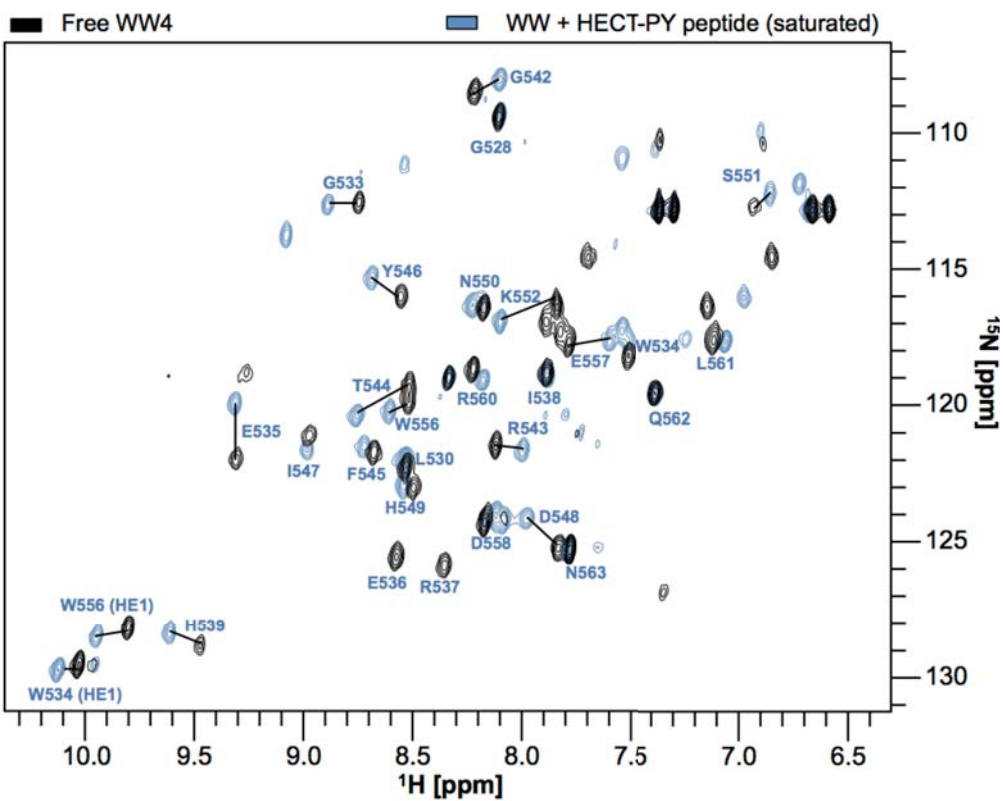
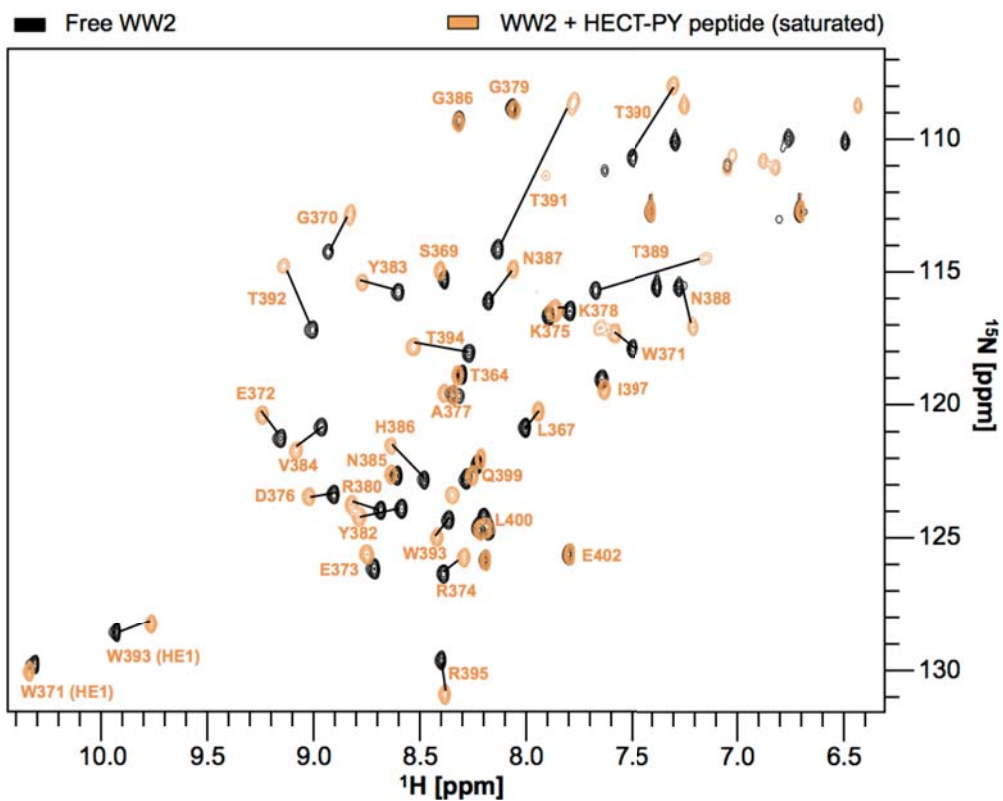
As mentioned in Section 5.2.2, the Asp36 – Gly39 segment in the $\beta 1$ - $\beta 2$ loop is missing in the available crystal structure of the C2 domain (2NSQ). Residues located in this segment are part of the binding sites of the HECT domain, the Ca^{2+} ions and IP_3 . In order to present these binding sites in a clear way, Tiago Gomes in our laboratory reconstructed the segment using Rosetta.

He used the Rosetta molecular modelling suite, version 3.4 (Leaver-Fay et al., 2011). 1000 decoys were generated by the kinematic loop modelling protocol (Mandell et al., 2009), and then subjected to clustering by the Rosetta cluster application. The five lowest energy decoys of the largest cluster –consisting of 160 decoys– were selected for further analysis (**Figure 9.1**). Final total RMSD was 0.474.

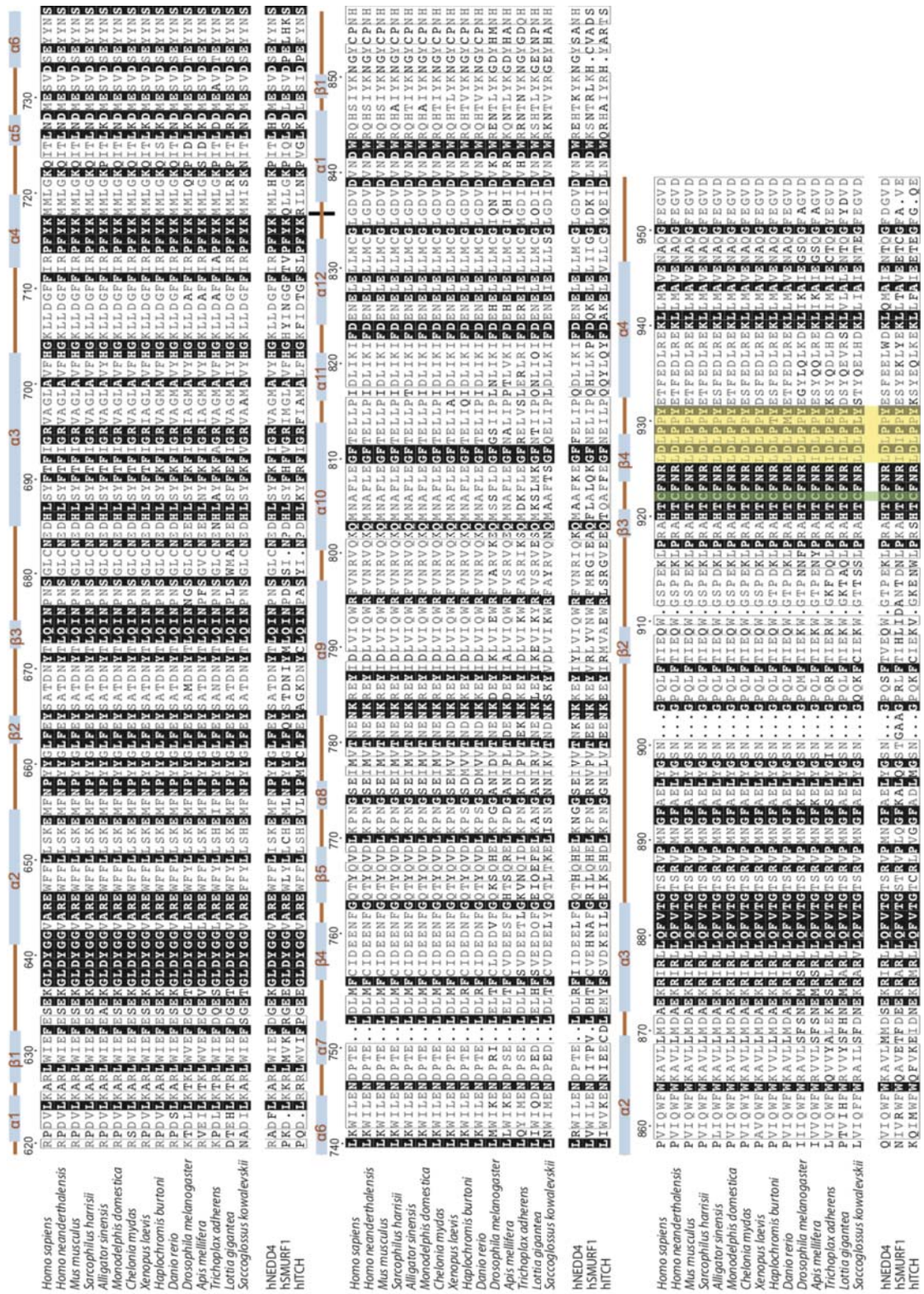
9.3. HSQC spectra of the free and HECT-PY bound WW1, WW2 and WW4 domains



9.3 HSQC spectra of the free and HECT-PY bound WW1, WW2 and WW4 domains



9.4. Sequence alignment of the HECT domain of Nedd4L in metazoans



The human sequence is aligned to its counterparts in 14 different metazoans including several mammals, a marsupial, an amphibian, several fishes and a reptile. Below, the sequences corresponding to the human Nedd4 family E3 ubiquitin ligases Nedd4, Smurf2 and Itch are aligned as well. The strictly conserved catalytic cysteine residue (Cys922) is highlighted in green whereas the PY motif is shaded in yellow. Residue numbering corresponds to isoform 5 of the human protein (Uniprot: Q96PU5-5). The secondary structure elements are numbered independently in the N-terminal lobe (residues 620 – 835) and the C-terminal lobe (residues 836 – 955).

Part VII.

Resumen en castellano

CHAPTER 10. RESUMEN EN CASTELLANO	211
10.1. Introducción	211
10.2. Objetivos	218
10.3. Resultados	220
10.4. Discusión	228

CHAPTER 10. Resumen en castellano

10.1. Introducción

A nivel molecular, las células organizan sus procesos bioquímicos fundamentales en complejas redes de agentes efectores, principalmente proteínas. La abundancia de cada proteína está controlada en primera instancia mediante la transcripción a mRNA, el splicing alternativo y la traducción en cadenas polipeptídicas ejecutada por los ribosomas. La complejidad de las células eucariotas les permite organizar su desarrollo en varias fases del ciclo celular, adaptarse rápidamente a los cambios en su entorno y coordinarse entre sí en tejidos y poblaciones. El comportamiento y el destino de una proteína en particular se regulan generalmente mediante modificaciones post-translacionales (PTMs), que consisten en la adición covalente de uno o más grupos químicos o polipéptidos a la secuencia nativa.

La ubiquitinación es una PTM conservada a lo largo de la evolución eucariota que consiste en la adición covalente de una o más moléculas de la proteína ubiquitina, y es determinante para el destino de la proteína sustrato. La ubiquitina es una proteína altamente conservada de 76 aminoácidos cuyo grupo carboxilo en el extremo C-terminal se liga al grupo amino ϵ de un residuo de lisina en la proteína sustrato (**Figura 10.1A**).

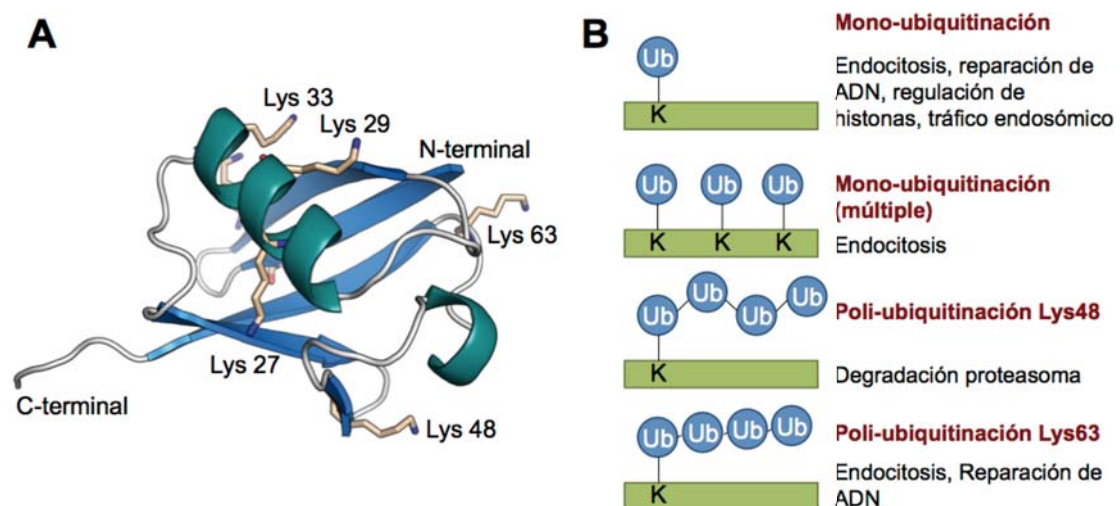


Figura 10.1: La ubiquitinación tiene varios efectos en el destino de la proteína sustrato. (A) Estructura cristalográfica de la ubiquitina (entrada en el PDB 1UBQ). (B) Diferentes patrones de ubiquitinación tienen consecuencias distintas en el destino de la proteína sustrato.

La ubiquitina contiene siete lisinas que con frecuencia son usadas para enlazar varias unidades y formar cadenas de poli-ubiquitina. El marcaje de la proteína sustrato con una sola unidad de ubiquitina se conoce como mono-ubiquitinación, y se puede dar en uno o varios residuos de lisina. La adición de una cadena de poli-ubiquitina se denomina poli-ubiquitinación y presenta variantes según el residuo de lisina usado para enlazar las distintas unidades de ubiquitina. Los efectos de cada una de estas modificaciones son variados (**Figura 10.1B**). La primera en ser descubierta y la más relevante consiste en la adición de cadenas de ubiquitina enlazadas por su residuo Lys48 y marca a la proteína sustrato para su degradación en el proteasoma. Este es de hecho el principal mecanismo celular de regulación de la vida media de las proteínas y juega un papel fundamental en procesos de señalización y metabolismo intracelular así como en la degradación de proteínas en mal estado.

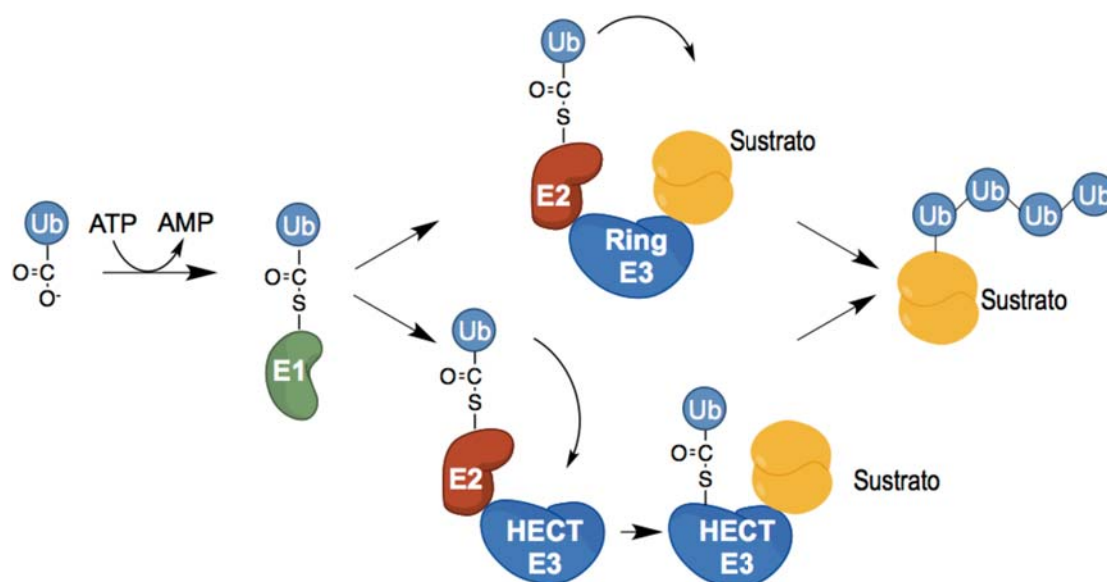


Figura 10.2: La ubiquitinación requiere la acción coordinada de tres tipos de enzimas. Las enzimas E1 (activadora), E2 (conjugadora) y E3 (ligasa) intervienen en el proceso. Hay dos tipos principales de ligasas E3, dependiendo de si establecen un intermediario covalentemente unido a ubiquitina (HECT) o no (RING).

El marcaje con ubiquitina requiere la acción coordinada de tres tipos de enzimas (**Figura 10.2**). Las enzimas E1 activan las moléculas de ubiquitina en el citoplasma uniéndolas covalentemente al grupo tiol de un residuo de cisteína en una reacción dependiente de ATP. A continuación, la ubiquitina es transferida a otra cisteína en la enzima E2 y unida finalmente a la proteína diana mediante la intervención de una ligasa E3. Existen dos tipos principales de ligasas E3, dependiendo de si establecen (HECT) o no (RING) una unión covalente con la ubiquitina antes de transferirla al sustrato. Así, las ligasas E3 son las enzimas responsables finales de transferir la ubiquitina al sustrato, y la regulación de su actividad y vida media es fundamental

para el correcto funcionamiento celular. Muchas de estas ligasas regulan su propia vida media mediante un proceso conocido como auto-ubiquitinación.

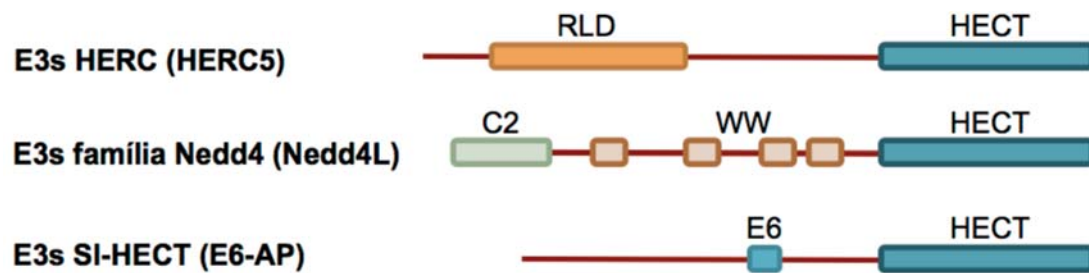


Figura 10.3: Ubiquitin-ligasas del tipo HECT en humanos. Las ligasas E3 del tipo HECT se clasifican en tres clases principales en función de su arquitectura modular. Las pertenecientes a la familia de Nedd4 (9 en humanos) son las mejor caracterizadas dada su relevancia.

En humanos se han identificado 28 proteínas distintas con un dominio HECT (Scheffner and Staub, 2007). Todas las que han sido caracterizadas hasta la fecha son proteínas modulares con el dominio HECT localizado en el extremo C-terminal. Dependiendo de los dominios que le acompañan, se clasifican en tres clases principales (**Figura 10.3**). Las ligasas de la **familia Nedd4** son proteínas multi-dominio con una arquitectura similar que cuenta con:

- Un dominio **C2** de unión a Ca^{2+} y fosfolípidos, incluyendo el fosfatidilinositol, la fosfatidilserina y la fosfatidilcolina (Plant et al., 1997). Es el responsable de la localización de estas ligasas en la membrana plasmática y otras membranas intracelulares.
- Entre dos y cuatro dominios **WW** responsables del reconocimiento del sustrato a través de la unión a secuencias ricas en prolina. Estos son los dominios de proteína más pequeños conocidos hasta la fecha y deben su nombre a dos residuos de triptófano altamente conservados. Se dividen en cuatro grupos según el tipo de secuencia que reconocen. Los dominios de tipo I (56%) unen motivos PY (secuencias PPxY). Los de tipo IV unen secuencias del tipo fosfo-S/TP (serina o treonina fosforiladas seguidas de prolina).
- El dominio catalítico **HECT**, que está estructuralmente dividido en dos sub-dominios independientemente plegados (los lóbulos N- y C-terminal). Las enzimas E2 se unen al lóbulo N-terminal, mientras que el C-terminal contiene la cisteína catalítica. Entre ambos sub-dominios existe una región flexible que permite la reorganización espacial de un lóbulo respecto al otro esencial para la actividad catalítica del dominio (Kamadurai et al., 2009). Cabe destacar que, cercano a la cisteína catalítica, el lóbulo C-terminal presenta un motivo LPxY altamente conservado a lo largo de la

evolución que difiere ligeramente de la secuencia PPxY canónica de un motivo PY pero que es reconocida por dominios WW cuando el dominio HECT está desplegado (Bruce et al., 2008).

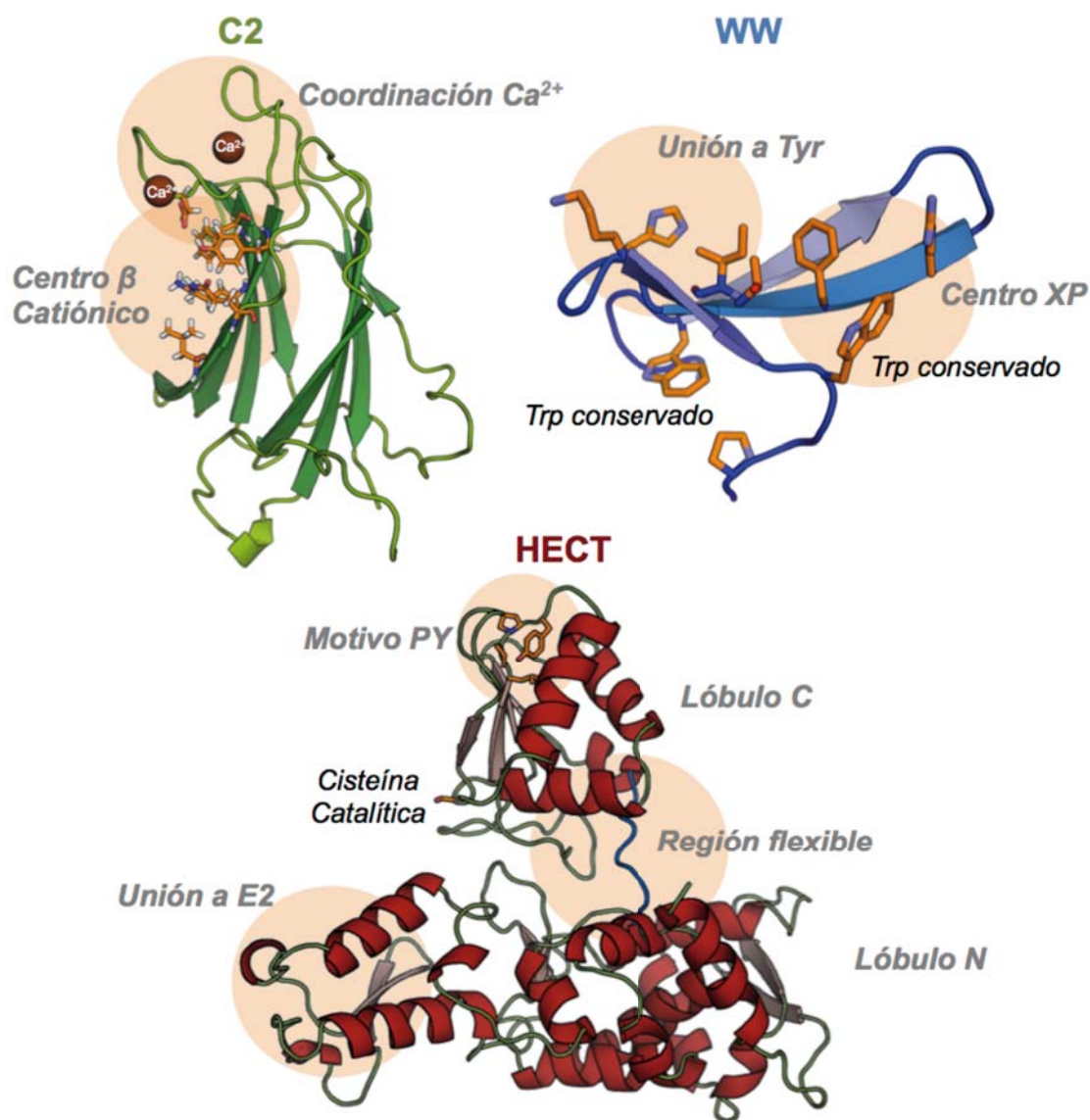


Figura 10.4: Dominios que integran la arquitectura modular de las ligasas E3 de la familia de Nedd4. Todas las estructuras mostradas pertenecen a Nedd4L. La estructura de su dominio C2 (2NSQ) se muestra en verde, con sus bucles de coordinación de Ca^{2+} y su centro β catiónico (ambos implicados en la unión a lípidos) destacados. La estructura por RMN del dominio WW3 (2MPT), un dominio WW de tipo I, se muestra en azul con su sitio de unión a Tyr y su centro XP (unión a prolinas) destacados. La estructura cristalográfica de su dominio HECT (3JW0) se muestra en rojo. La región flexible que une el lóbulo N- con el C-terminal se muestra en azul. El motivo PY y la región de unión a la enzima E2 se muestran destacados. En esta estructura, la posición correspondiente a la cisteína catalítica está mutada a serina.

La familia Nedd4 consta de nueve miembros distintos en humanos y se ha caracterizado extensivamente dada la relevancia de las proteínas cuya vida media regulan. Estas ligasas poli-ubiquitinan receptores y canales iónicos de membrana así

como proteínas intracelulares (incluyendo supresores de tumores, agentes implicados en procesos de diferenciación, proteínas encargadas de la transducción de señal, etc.) para su degradación en el proteasoma, y regulan el tráfico y la endocitosis de proteínas en la membrana plasmática mediante mono-ubiquitinación.

El trabajo experimental de esta tesis se centra en el estudio de **Nedd4L** (**Figura 10.5**), que fue inicialmente relacionada con la regulación de la vida media de ENaC, el principal canal de sodio del epitelio que juega un papel fundamental en la reabsorción de líquido en el riñón y en la homeostasis del sodio (Kamynina et al., 2001). De hecho, mutaciones en los motivos PPxY citoplasmáticos de ENaC son la causa de la forma poco común de hipertensión hereditaria conocida como **síndrome de Liddle** (Staub et al., 1996). La ligasa también ha sido asociada con la **fibrosis quística** (Caohuy et al., 2009).

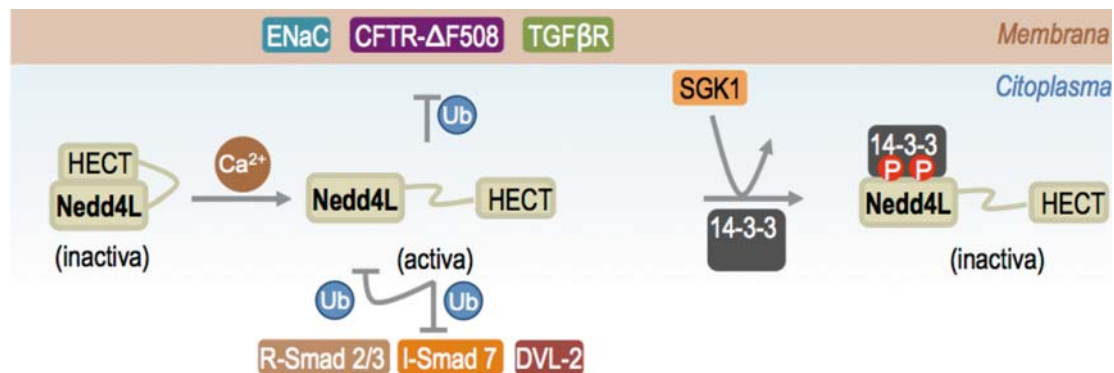


Figura 10.5: Función y regulación de Nedd4L. La conformación cerrada e inactiva de Nedd4L se establece mediante contactos HECT-C2. El Ca²⁺ activa la ligasa y facilita su relocalización a la membrana celular, donde ubiquitina varios canales iónicos y el receptor de citoquinas TGF-β. En el citoplasma, la ligasa activa ubiquitina las proteínas Smads y además se auto-ubiquitina. La fosforilación de SGK-1 y la unión de 14-3-3 inactivan la ligasa.

Recientemente se ha relacionado la ligasa con la regulación de proteínas de la ruta de señalización activada por citoquinas de la familia de **TGF-β**. Esta ruta gobierna varios aspectos de la diferenciación celular, la morfogénesis, la homeostasis y la regeneración en tejidos durante el desarrollo y también en la etapa adulta. Disfunciones en esta ruta se han relacionado con distintos tipos de **cáncer** y **metástasis** entre otros (Massagué, 2012). La unión de un dímero de citoquinas de la familia de TGF-β a su receptor en la membrana celular provoca la activación de las R-Smads (Smad1, 2, 3, 5 y 8), que se asocian con la Co-Smad (Smad4) y son translocadas al núcleo celular (**Figura 10.6**). Allí regulan la expresión de los genes diana de esta ruta. Las I-Smads (Smad6 y 7) inhiben el proceso a distintos niveles como mecanismo de regulación negativa.

Las R-Smads son proteínas modulares que presentan dos dominios MH conectados a través de una región flexible (Shi and Massagué, 2003). En la mayoría de casos el dominio MH1 une a ADN y el dominio MH2 (altamente conservado) interacciona con el receptor de membrana, con otros Smads e interviene en la relocalización del citoplasma al núcleo. La región flexible que conecta los dominios MH contiene un clúster de sitios de fosforilación S/TP y un motivo PPxY adyacente que habilitan su interacción con dominios WW en proteínas asociadas.

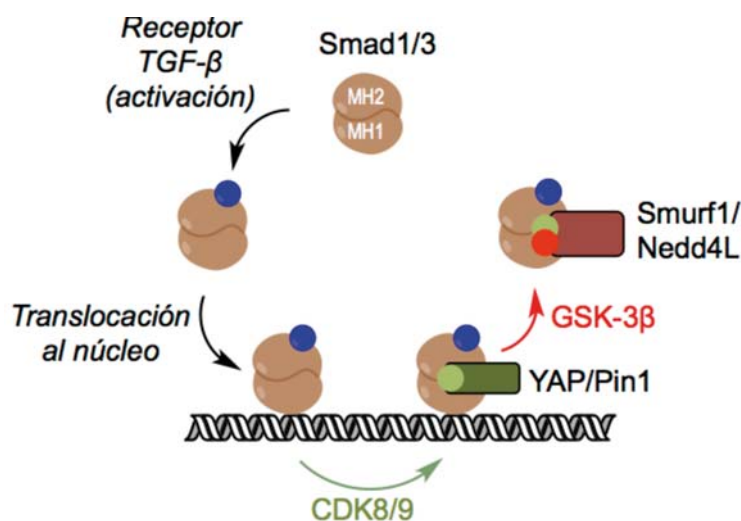


Figura 10.6: Las fosforilaciones nucleares en la región desordenada de Smad1 y 3 gobiernan su destino. CDK8/9 favorecen la interacción con YAP y Pin1, que cooperan en la transducción de señal. CDK8/9 también inducen la acción de GSK3- β , que favorece la interacción con Smurf1 y Nedd4L resultando en la degradación de las Smads.

Cuando el complejo R-Smads/Co-Smad se encuentra en el núcleo, las quinasas CDK8/9 fosforilan el clúster S/TP (Matsuura et al., 2004). En esta situación, Pin1 y YAP reconocen con sus dominios WW a Smad3 y Smad1 respectivamente y cooperan en la transducción de la señal (Alarcón et al., 2009; Matsuura et al., 2010). No obstante, la acción de CDK8/9 induce la subsiguiente fosforilación del clúster S/TP por parte de la quinasa GSK3- β , que actúa favoreciendo la asociación de las R-Smads con las ubiquitín ligasas Nedd4L (Smad3) y Smurf1 (Smad1), dando paso a su degradación (Gao et al., 2009; Sapkota et al., 2007).

Muchas de las proteínas de la familia de Nedd4 (Itch (Gallagher et al., 2006), Smurf2 (Wiesner et al., 2007), Smurf1 (Wan et al., 2011), WWP2 (Soond and Chantry, 2011), Nedd4 y Nedd4L (Wang et al., 2010)) optimizan su actividad y vida media cambiando de una conformación cerrada e inactiva a una abierta y activa. Los mecanismos que gobiernan este proceso no se conocen por completo. Datos bioquímicos anteriores a la realización de la presente tesis doctoral sugerían que Nedd4L existe en una conformación latente auto-inhibida mediada por contactos

entre los dominios C2 y HECT que es incapaz de ubiquitinar a sus sustratos y de auto-ubiquitinars. Un incremento intracelular de los niveles de Ca^{2+} resulta en la activación de la ligasa (ya que los contactos HECT-C2 no pueden ser detectados a la vez que los sustratos de la ligasa son ubiquitinados) y además facilita su re-localización a la membrana celular.

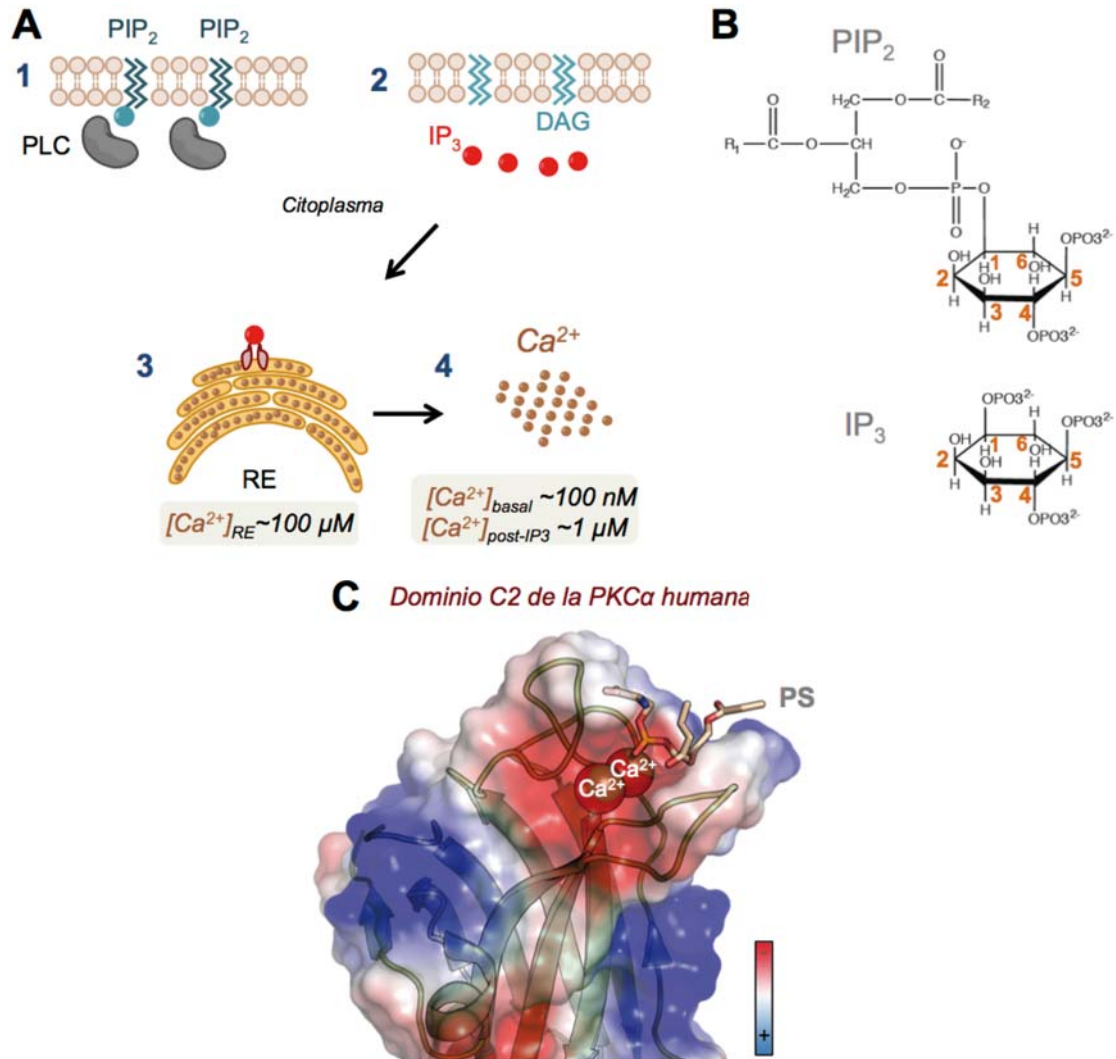


Figura 10.7: Inter-relación entre la señalización por Ca^{2+} y la unión de dominios C2 a la membrana celular. (A) La hidrólisis del fosfolípido de membrana PIP₂ por parte de la enzima PLC (1) libera IP₃ al citoplasma (2), cuya unión a sus receptores en el retículo endoplásmico (3) resulta en un incremento en la concentración citoplasmática de Ca^{2+} (4). (B) Fórmula química del PIP₂ y del IP₃. (C) Los dominios C2 de la familia de la PKCα humana unen a Ca^{2+} , que desempeña un papel fundamental en la interacción con fosfolípidos de membrana como la fosfatidilserina (PS) y el PIP₂.

La señalización intracelular por Ca^{2+} es a su vez un mecanismo fundamental en la biología celular implicado en la regulación de gran variedad de procesos. La concentración basal de Ca^{2+} en el citoplasma está alrededor de 100 nM (Clapham, 2007). Los incrementos transitorios en los niveles de Ca^{2+} son inducidos por

estímulos externos, a los que la célula responde mediante distintos mecanismos. El mas caracterizado de ellos es iniciado por la fosfolipasa C (PLC), que hidroliza el fosfolípido de membrana PIP_2 (fosfatidilinositol 4,5-bifosfato) y libera al citoplasma su cabeza soluble, el IP_3 (inositol 1,4,5-trifosfato). Esta molécula es capaz de movilizar las reservas intracelulares de Ca^{2+} en el retículo endoplásmico, consiguiendo llevar la concentración de Ca^{2+} en el citoplasma hasta $1 \mu M$ (Berridge et al., 2000). Los dominios C2 como los que se encuentran en las ubiquitín ligasas de la familia Nedd4 se anclan a la membrana celular mediante la interacción con varios fosfolípidos de membrana incluyendo el PIP_2 , y varios de ellos incluyendo el de Nedd4L requieren que el Ca^{2+} medie dicha interacción (Murray and Honig, 2002; Verdaguer et al., 1999). Los contactos se establecen con la cabeza citoplasmática del PIP_2 , químicamente idéntica al IP_3 .

10.2. Objetivos

La caracterización funcional de Nedd4L constituye la motivación central para su estudio en profundidad. Los modelos para su activación conformacional, su auto-ubiquitinación así como la selectividad de sustratos en función de su estado de fosforilación han sido inferidos a partir de experimentos bioquímicos y de biología molecular. La presente tesis pretende contribuir a la comprensión de dichos procesos desde una perspectiva estructural y biofísica que ayude a revelar sus bases mecánicas a resolución atómica y a entender sus implicaciones biomédicas.

Objetivo 1 – Síntesis de una familia de fosfopéptidos correspondientes a la región desordenada de Smad1 y 3 para estudiar la selectividad de sustrato en Nedd4L y Smurf1.

En el marco de un estudio centrado en revelar las bases estructurales y biofísicas del reconocimiento de sustratos por parte de los dominios WW de las proteínas asociadas a R-Smads (Pin1, YAP, Smurf1 y Nedd4L), planeamos la síntesis de una familia de fosfopéptidos representando diferentes estados de fosforilación de la región desordenada de Smad1 y Smad3 mediante síntesis peptídica en fase sólida (SPPS).

Objetivo 2 – Caracterización estructural de los contactos inter-dominio C2-HECT en la conformación latente de Nedd4L y del papel que los niveles inter-relacionados de IP_3/Ca^{2+} juegan en la activación de la ligasa.

Mediante resonancia magnética nuclear (RMN), planeamos mapear el sitio de unión del dominio HECT en la superficie del dominio C2 en el contexto de la

conformación latente de Nedd4L y evaluar el efecto adverso del Ca^{2+} en esta interacción. El Ca^{2+} también se une a la superficie del dominio C2, así que estudiamos dicha unión mediante RMN y calorimetría de titulación isotérmica (ITC) a la búsqueda de la base estructural para dicho efecto adverso. Además, planteamos la hipótesis de que el Ca^{2+} induce la re-localización de Nedd4L a la membrana celular favoreciendo la interacción con el IP_3 .

Objetivo 3 – Caracterización estructural y biofísica del reconocimiento de un motivo PY altamente conservado en el dominio HECT por parte de los dominios WW de Nedd4L en relación con la auto-ubiquitinación.

El motivo LPxY del lóbulo C-terminal del dominio HECT (HECT-PY) de Nedd4L se encuentra parcialmente enterrado en su estructura, especialmente el residuo de tirosina. La exposición de dicho residuo es esencial para que el motivo sea reconocido por dominios WW. Planteamos la hipótesis de que la desestabilización del dominio HECT haga accesible este residuo, resultando ser uno de los mecanismos implicados en la auto-ubiquitinación de la ligasa. Así, planeamos evaluar por RMN e ITC el reconocimiento por parte de los dominios WW de Nedd4L del motivo HECT-PY, ambos en el contexto del subdominio HECT-C plegado así como completamente expuesto (en forma de péptido sintético). Además, decidimos estudiar la flexibilidad interna del subdominio HECT-C mediante experimentos de relajación y simulaciones de dinámica molecular con el fin de evaluar la posible exposición de la tirosina del motivo HECT-PY durante la desestabilización de su plegamiento.

Objetivo 4 – Marcaje paramagnético de un dominio WW

Las etiquetas paramagnéticas incrementan el potencial de la RMN para el estudio de interacciones entre proteínas o con ligandos, ya que permiten medir distancias mucho mayores que los observables convencionales como los NOEs. Algunas de estas etiquetas dan lugar al efecto conocido como incremento paramagnético de la relajación (PRE), cuya conversión en medidas de distancias es relativamente sencilla. Aprovechando la amplia caracterización estructural del dominio WW3 de Nedd4L, planeamos usarlo para realizar un estudio preliminar con el fin de explorar la aplicación de esta técnica a nuestros intereses.

10.3. Resultados

Síntesis de una familia de fosfopéptidos para el estudio de la regulación de las Smads en señalización por citoquinas TGF- β [‡]

Se sintetizaron dos sets de péptidos representando distintos estados de fosforilación de la región desordenada inter-dominio de Smad1 y Smad3 con el fin de ser utilizados en experimentos de interacción con los dominios WW de proteínas asociadas a la ruta de señalización de las citoquinas TGF- β (**Figura 10.8A**). Para ello, se utilizó la síntesis de péptidos en fase sólida (SPPS) con química Fmoc. Debido a la incorporación de hasta cuatro residuos fosforilados a secuencias relativamente largas (hasta 36 residuos), algunas de las síntesis requirieron la optimización de distintos protocolos, incluyendo el acoplamiento de aminoácidos, el desanclaje de la resina y la purificación.

Para el acoplamiento de aminoácidos, se usaron tres estrategias distintas: manual o automática a temperatura ambiente y automática asistida por microondas.

Para el acoplamiento de residuos fosforilados de forma automática con microondas se utilizaron condiciones optimizadas, consistentes en un tiempo de reacción mayor (15 minutos en lugar de 5) y una temperatura ligeramente menor (72 °C en lugar de 75) comparado con los residuos no fosforilados. Además, la desprotección del grupo amino del esqueleto peptídico de los residuos fosforilados fue realizada con DBU en lugar de piperidina para evitar la β -eliminación del fosfato. En el modo manual, algunos residuos impedidos estéricamente requirieron ser re-acoplados, y la eficiencia de estas reacciones fue mejorada con el uso combinado de triazoles y carbodiimidas como agentes activadores del acoplamiento.

En péptidos que contenían el residuo metionina, el desanclaje del péptido de la resina fue realizado en atmósfera inerte para evitar su oxidación (**Figura 10.8B**). Si además contenían algún residuo fosforilado, se observó la alquilación de la metionina con el grupo protector del fosfato (Bzl). Para solucionar dicho problema se añadió 2-PySH a la solución de TFA usada en el desanclaje de estos péptidos. El 2-PySH es un agente recolector de grupos Bzl usado en química BOC cuyo uso en la síntesis de fosfopéptidos por química Fmoc no había sido descrito con anterioridad.

[‡] 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 and Development** 25, 1275-1288.

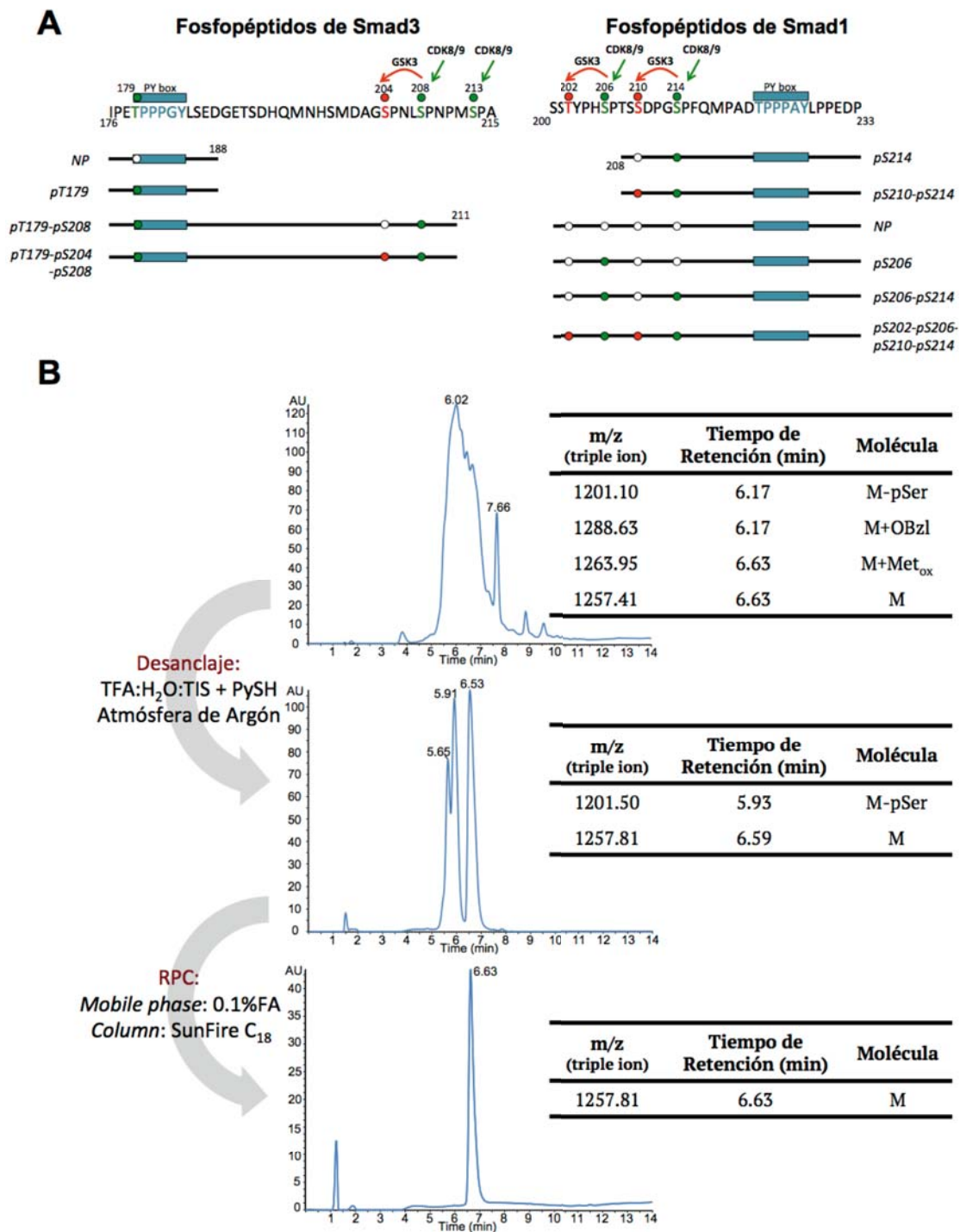


Figura 10.8: Fosfopéptidos de Smad3 y Smad1 sintetizados por SPPS. (A) Resumen de los fosfopéptidos sintetizados. (B) Optimización del desanclaje de la resina y de la purificación para el péptido Smad1 pS206-pS214.

Dado que el acoplamiento de residuos fosforilados puede presentar ocasionalmente un rendimiento bajo, la presencia de subproductos truncados que carecen de uno o más de estos residuos es un problema potencialmente intrínseco a la síntesis de fosfopéptidos. Se optimizaron las condiciones de purificación para separar el producto deseado de estos subproductos, consistentes en la sustitución

del TFA por ácido fórmico en la fase móvil combinado con el uso de la columna Waters SunFire™ PREP C18 OBD™ (**Figura 10.8B**).

Colectivamente, estos métodos nos permitieron obtener muestras en suficiente cantidad y pureza (la mayoría de los péptidos fueron obtenidos con una pureza de un 90% o superior) para nuestros estudios de interacción con YAP y Smurf1 en el caso de Smad1, y Pin1 y Nedd4L en el caso de Smad3. Los resultados estructurales y biofísicos de estos estudios fueron presentados en las tesis del Dr. Eric Aragón y de la Dra. Nina Görner respectivamente.

Finalmente, se exploró la posibilidad el uso de la ligación química nativa en un fosfopéptido, acompañada por la síntesis de un tioéster peptídico y la desulfuración de un fosfopéptido para convertir residuos de cisteína en alanina. La incorporación de estas técnicas a nuestro repertorio nos habilita para obtener secuencias fosforiladas mas largas para ser usadas en nuestros estudios.

Regulación de la actividad de la ubiquitín ligasa Nedd4L: la señalización por $\text{Ca}^{2+}/\text{IP}_3$ induce la transición de la conformación latente a la activa[‡]

Con el objetivo de caracterizar estructuralmente la conformación latente de Nedd4L así como el efecto del $\text{IP}_3/\text{Ca}^{2+}$ preparamos constructos recombinantes de los dominios C2 (17.9 kDa) y HECT (44.9 kDa) con distintos marcajes isotópicos. Estudiamos la interacción entre ambos dominios mediante ensayos de perturbación del desplazamiento químico basados en el experimento de RMN de correlación $^1\text{H}-^{15}\text{N}$ HSQC.

En primer lugar se preparó una muestra del dominio C2 marcada en $^2\text{H}-^{13}\text{C}-^{15}\text{N}$ en un tampón libre de Ca^{2+} (250 mM NaCl, 20 mM Tris, 1 mM EGTA). La muestra fue usada para la adquisición del par de experimentos tridimensionales de RMN HNcoCACB/HNCACB. Mediante el análisis de estos espectros se asignaron la mayoría resonancias del esqueleto peptídico del dominio.

Se preparó una muestra marcada en ^{15}N del dominio C2 y una no marcada del dominio HECT, ambas equilibradas en el mismo tampón libre de Ca^{2+} . Primero se adquirió un espectro HSQC de referencia del dominio C2. La adición de un equivalente del dominio HECT no marcado dio como resultado el ensanchamiento y la pérdida final de aproximadamente un 85% de las señales del esqueleto peptídico

[‡] Escobedo, A., Gomes, T., Aragón, E., Martín-Malpartida, P., Ruiz, L., Macias, M.J. (2014). *Structural basis of the activation and degradation mechanisms of the E3 ubiquitin ligase Nedd4L*. **Structure** 22, 1446-1457.

del dominio C2 (**Figura 10.9A**). Interpretamos dicho ensanchamiento como el resultado de la formación de un complejo HECT-C2 de elevado peso molecular. La formación de este complejo lleva a la reducción de los tiempos de relajación transversal de la mayoría de residuos del dominio C2, y solo los más flexibles permanecen en el espectro (el 15% restante corresponde al segmento N-terminal del dominio).

Para evitar dicho ensanchamiento debido a la relajación, repetimos el mismo experimento usando una muestra del dominio C2 marcada en ^2H - ^{15}N (**Figura 10.9B**). En esta muestra todos los átomos de hidrógeno fueron reemplazados por deuterio, a excepción de aquellos localizados en los grupos amida del esqueleto peptídico, que intercambian rápidamente con el tampón no marcado.

En añadir un equivalente del dominio HECT no marcado pudimos identificar perturbaciones en el desplazamiento químico de residuos localizados en los bucles $\beta 1$ - $\beta 2$ y $\beta 3$ - $\beta 4$ del dominio C2 (**Figura 10.9C**).

A continuación preparamos una muestra marcada en ^{15}N del dominio C2 equilibrada en un tampón con Ca^{2+} (250 mM NaCl, 20 mM Tris, 20 mM CaCl_2 , 1 mM EGTA). Diferencias sustanciales en el espectro HSQC respecto a la muestra libre de Ca^{2+} forzaron a la asignación de nuevo de las resonancias del esqueleto peptídico del dominio C2 en presencia de Ca^{2+} (**Figura 10.9G**). La comparativa de los dos espectros HSQC permitió mapear la interacción del Ca^{2+} con el dominio C2, localizada en los bucles $\beta 1$ - $\beta 2$, $\beta 5$ - $\beta 6$ y $\beta 7$ - $\beta 8$ (**Figura 10.9F**). Esta región se solapa parcialmente con el sitio de unión del dominio HECT. La unión del Ca^{2+} al dominio C2 fue también analizada por ITC, que reveló la presencia de dos sitios de unión con afinidades (K_D) de 26.5 ± 10.9 nM y 1.7 ± 0.2 μM respectivamente (**Figura 10.9E**).

Preparamos también una muestra del dominio HECT no marcado y equilibrado en el tampón con Ca^{2+} . Al añadirlo a la muestra en ^{15}N del C2 con Ca^{2+} , el 98% de las resonancias no se vieron afectadas (**Figura 10.9D**). Así, el Ca^{2+} dificulta enormemente la interacción entre estos dos dominios.

La interacción del IP_3 (análogo soluble de la porción citoplasmática del fosfolípido de membrana PIP_2) con el dominio C2 fue también estudiada mediante experimentos de titulación por RMN, basados en HSQC y también en espectros mono-dimensionales de ^{31}P . El IP_3 fue progresivamente añadido a una muestra marcada en ^{15}N del dominio C2 equilibrada en el tampón sin Ca^{2+} . Las resonancias ^{31}P del IP_3 sufrieron pequeñas perturbaciones, mientras que las resonancias ^1H - ^{15}N del dominio C2 apenas se vieron afectadas (**Figura 10.9H**).

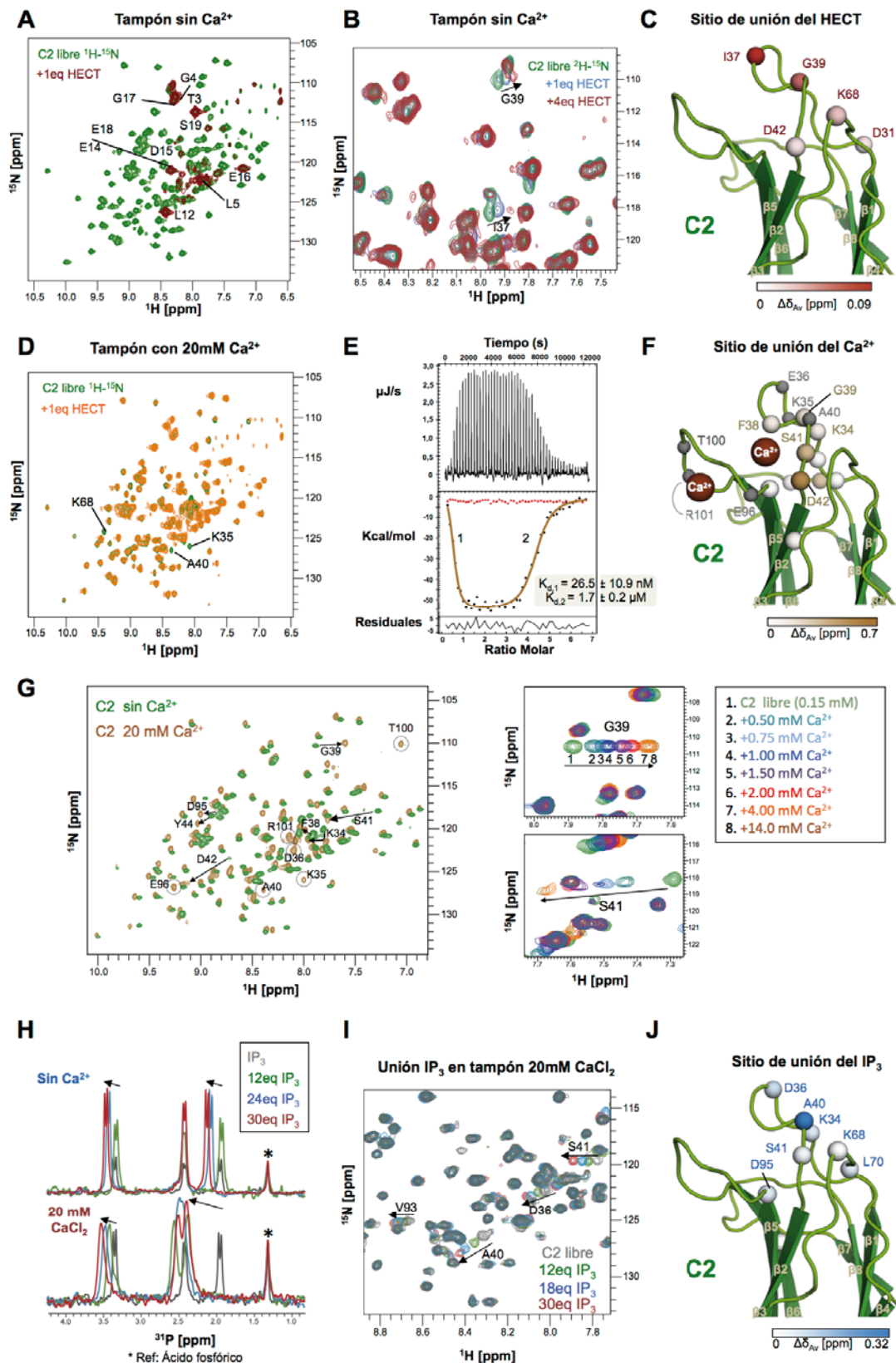


Figura 10.9: Bases estructurales y biofísicas para la activación de Nedd4L. (A) – (C) muestran los experimentos para el estudio de la interacción HECT-C2 en ausencia de Ca^{2+} . (D) – (G) muestran la unión del Ca^{2+} al dominio C2, incluyendo el experimento por ITC (E), así como el efecto en la interacción HECT-C2 (D). (H) – (J) muestran la unión del IP_3 al dominio C2 y el efecto del Ca^{2+} en esta interacción.

El experimento fue repetido en presencia de Ca^{2+} . En esta situación, las perturbaciones en las resonancias ^{31}P del IP_3 así como en las resonancias ^1H - ^{15}N del dominio C2 fueron mucho mayores (**Figura 10.9H e I**). Dichas perturbaciones corresponden a residuos localizados en los bucles $\beta 1$ - $\beta 2$, $\beta 3$ - $\beta 4$ y en la lámina $\beta 5$ (**Figura 10.9J**), región que se solapa con el sitio de unión del dominio HECT y del Ca^{2+} . Así, el Ca^{2+} puentea la interacción entre el dominio C2 y el IP_3 mediante interacciones electrostáticas.

Moléculas funcionales de Nedd4L reconocen a sus homólogas parcialmente desestructuradas como medio de regulación de la vida media de la ligasa[‡]

Para comprobar si los dominios WW de Nedd4L reconocen el bucle que contiene el sitio PY, realizamos titulaciones mediante experimentos HSQC consistentes en la adición de un dominio WW no marcado a una muestra marcada en ^{15}N del subdominio HECT-C. Para ello fue necesaria la asignación de las resonancias del esqueleto peptídico correspondientes a dicho sub-dominio.

Usando el dominio WW2 o bien el WW3, los resultados fueron similares: solo se detectaron pequeñas perturbaciones en los desplazamientos químicos de las resonancias ^1H - ^{15}N del HECT-C, que además se restringen a unos pocos residuos expuestos (**Figura 10.10A y B**). Dichas perturbaciones podrían deberse a interacciones inespecíficas o a pequeñas perturbaciones estructurales.

Para evaluar la posible exposición del residuo de tirosina en el motivo HECT-PY (Y931), analizamos las propiedades dinámicas del subdominio HECT-C en la escala temporal de ps a ns mediante la medida de los parámetros T_1 , T_2 y NOE heteronuclear a 298 K (**Figura 10.10C**). Los residuos localizados en elementos de estructura secundaria y en bucles cortos muestran valores típicos de zonas compactas, incluyendo el bucle $\beta 4$ - $\alpha 4$ donde se localiza el motivo PY. El bucle $\alpha 3$ - $\beta 2$ muestra valores relacionados con una mayor flexibilidad, de acuerdo con su estructura desordenada.

A continuación realizamos una simulación de dinámica molecular para analizar la posibilidad de que la tirosina del motivo PY resulte expuesta cuando el dominio HECT se desestabiliza. Usamos una rampa de temperatura desde 310 hasta 600 K (estos valores no correlacionan directamente con valores experimentales) durante 40 ns. Después de 13.59 ns, y mientras el dominio todavía conserva un plegamiento

[‡] Escobedo, A., Gomes, T., Aragón, E., Martín-Malpartida, P., Ruiz, L., Macias, M.J. (2014). *Structural basis of the activation and degradation mechanisms of the E3 ubiquitin ligase Nedd4L*. **Structure** 22, 1446-1457.

pseudo-nativo, el residuo de tirosina resulta expuesto posibilitando así su reconocimiento por parte de los dominios WW (**Figura 10.10D**).

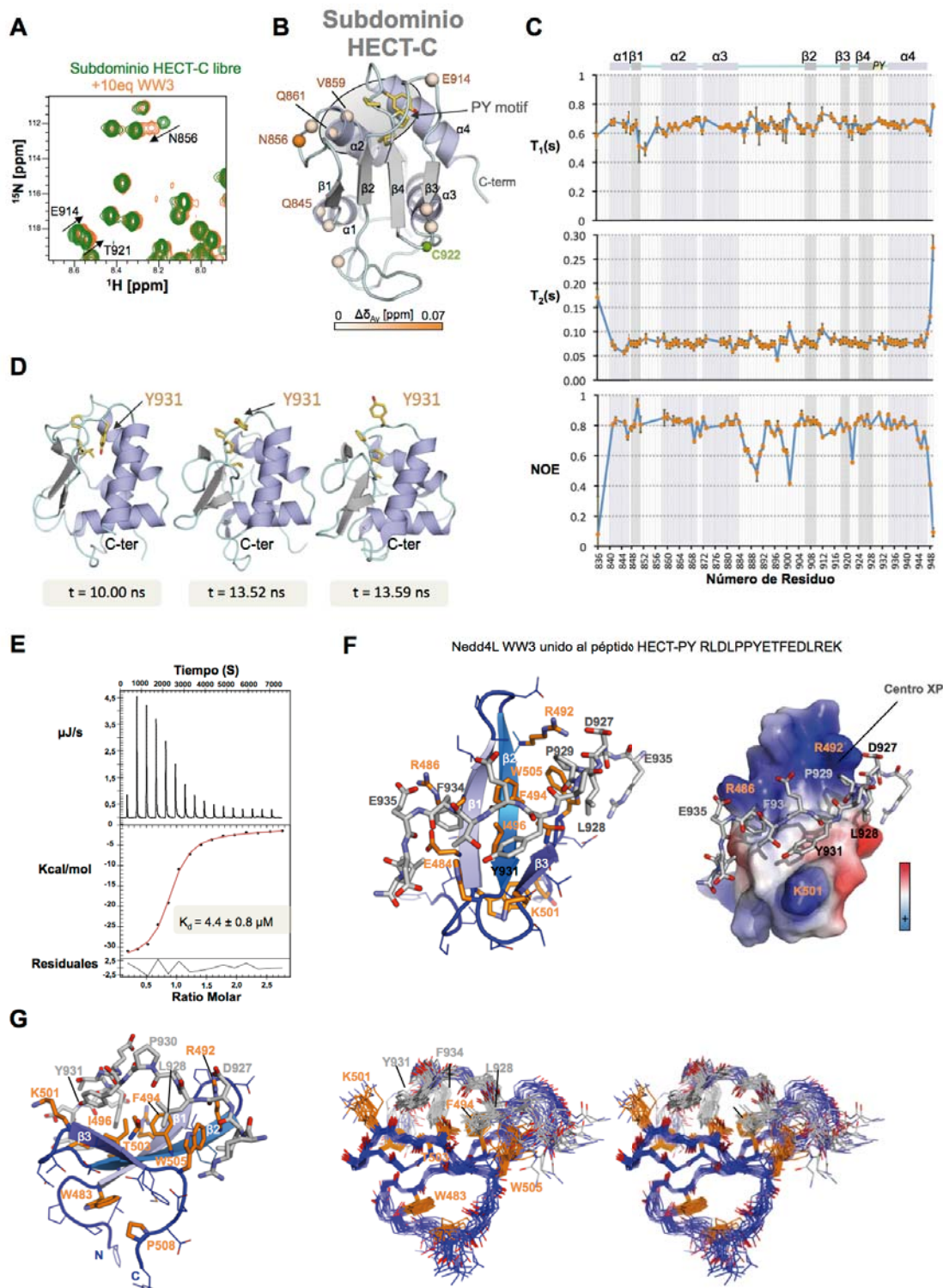


Figura 10.10: El dominio WW3 interacta con el motivo HECT-PY expuesto. (A) y (B) muestran la titulación del dominio WW3 sobre el subdominio HECT-C plegado. (C) Datos de relajación para el subdominio HECT-C. (D) Instantáneas de la dinámica molecular que muestran la exposición de Y931 mientras el subdominio HECT-C mantiene un plegamiento pseudo-nativo. (E) Experimento ITC para la medida de la afinidad de la interacción WW3-HECT-PY (péptido). (F) y (G) Estructura del complejo resultada por RMN.

Con el objetivo de estudiar en detalle la unión de los dominios WW al motivo HECT-PY, sintetizamos un péptido de 15 residuos incluyendo dicho motivo. Las interacciones fueron evaluadas por RMN e ITC. Ambas técnicas revelaron que el dominio WW3 une al motivo con la mayor afinidad ($K_D = 4.4 \pm 0.8 \mu\text{M}$) (**Figura 10.10E**).

A pesar de que el motivo HECT-PY difiere ligeramente de la secuencia canónica de motivos PPxY la fuerte afinidad con el dominio WW3 es comparable otras medidas para dominios WW que reconocen a sus sustratos (Aragón et al., 2012; Aragón et al., 2011). Decidimos investigar las características de esta interacción resolviendo la estructura tridimensional del complejo en solución mediante RMN (**Figura 10.10F y G**). Se prepararon muestras del dominio WW3 isotópicamente marcado en ^{13}C y/o ^{15}N en complejo con el péptido no marcado. Se adquirieron espectros 3D heteronucleares para la asignación de las resonancias del dominio, incluyendo el par CBCAcoNH/CBCANH y el par HSQC-TOCSY/HSQC-NOESY ^{15}N . El péptido y sus contactos con el dominio fueron asignados en los espectros 2D homonucleares de protón TOCSY y NOESY. La estructura muestra al dominio WW3 con un plegamiento canónico y al péptido unido en conformación extendida, con la prolina del motivo LPxY unida al centro XP y la tirosina coordinada por residuos del segundo bucle.

Marcaje paramagnético del dominio WW3 de Nedd4L en complejo con el péptido HECT-PY

En primer lugar, se usó la mutagénesis dirigida para preparar dos mutantes del dominio WW3 de Nedd4L, introduciendo un residuo de cisteína en cada uno de ellos en posiciones distintas (**Figura 10.11A**). Los mutantes se marcaron en ^{15}N , y se prepararon dos muestras de cada uno de ellos, una derivatizada con una etiqueta paramagnética (MTSL) y la otra con su homólogo diamagnético (MTS). Se comprobó mediante experimentos HSQC que el plegamiento y la capacidad de unir al péptido HECT-PY seguían intactos en ambos mutantes (**Figura 10.11C**). Los efectos PRE se observaron comparando la intensidad de los picos HSQC correspondientes a los mutantes en presencia de las etiquetas paramagnética (I_{MTSL}) y diamagnética (I_{MTS}) (**Figura 10.11B**).

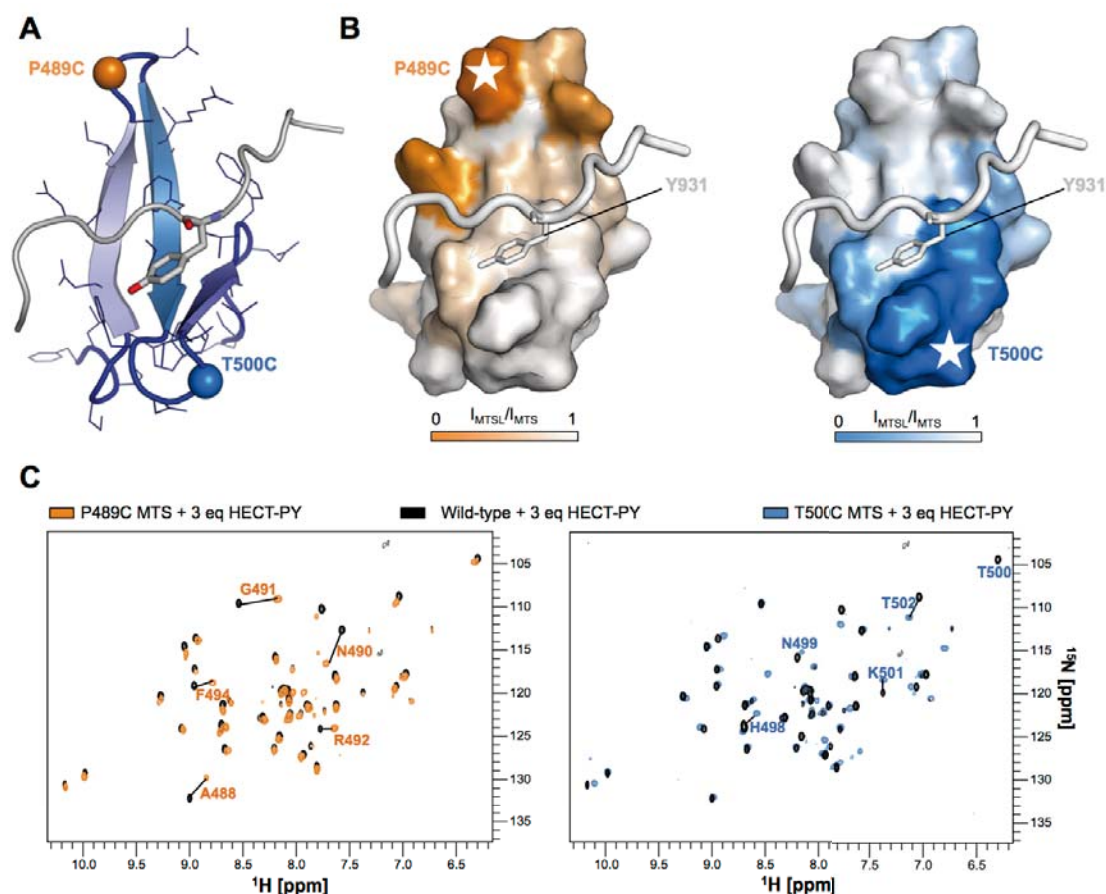


Figura 10.11: Incorporación de una etiqueta paramagnética al complejo Nedd4L WW3 – HECT-PY. (A) Posiciones mutadas a cisteína en el dominio WW3 de Nedd4L. (B) Ratios I_{MTSL}/I_{MTS} para cada residuo del dominio WW3 representados en gradiente de color. (C) Espectros HSQC de los mutantes del dominio WW3 derivatizados con MTS y con 3 equivalentes de HECT-PY comparados con la versión no mutante del experimento.

10.4. Discusión

La ubiquitinación es uno de los mecanismos que las células eucariotas usan para regular la vida media y la función de las proteínas, y su impacto biológico ha sido caracterizado extensamente. Afecta a proteínas implicadas en procesos biológicos fundamentales como el ciclo y la división celular, el desarrollo embrionario y la homeostasis en tejidos. Defectos en su regulación pueden dar lugar a enfermedades graves (Petroski, 2008). Aunque en un principio se pensaba que las ubiquitin ligasas eran constitutivamente activas y estaban reguladas solo a nivel de expresión génica, en los últimos años se han descrito diversos mecanismos de regulación post-translacional. Entre ellos encontramos la fosforilación de sustratos o de las ligasas mismas, que regula el reconocimiento de ligando, así como la transición de una conformación latente a una abierta y activa o la auto-ubiquitinación como medio de regulación de su propia vida media.

Nedd4L ubiquitina e induce la internalización de varios canales de iones en la membrana celular, entre ellos el canal de sodio epitelial ENaC (Staub et al., 1996). Esta ligasa ha sido vinculada con enfermedades como el síndrome de Liddle y la fibrosis quística en relación con defectos en esta regulación. Además, es una de las principales ligasas responsables de la ubiquitinación del receptor de membrana de TGF- β así como de R-Smad2/3 y de I-Smad7. La desregulación de esta vía de señalización ha sido vinculada a diversos tipos de cáncer y metástasis.

Los datos recogidos en la presente tesis doctoral contribuyen a una comprensión de los mecanismos de regulación de Nedd4L desde una perspectiva estructural y mecanística, usando para ello un enfoque multidisciplinar entre la química orgánica, la biología estructural y la biofísica.

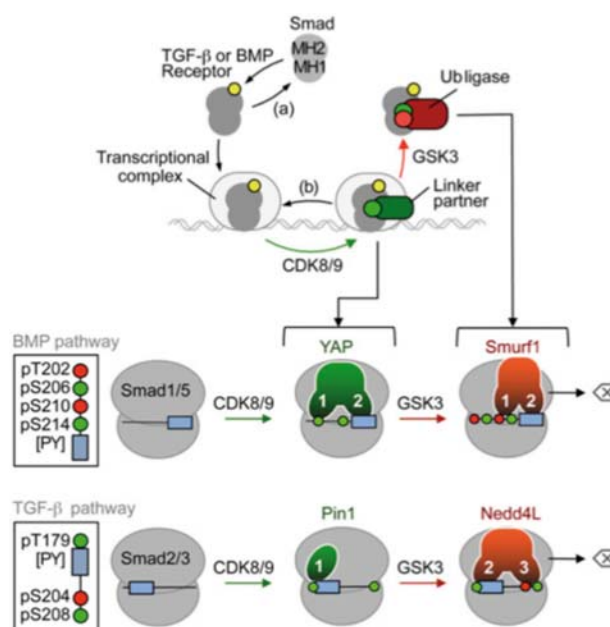


Figura 10.12: Esquema resumen de las bases estructurales detrás de la actividad y la degradación de las R-Smads. Derivado de los estudios estructurales realizados con los péptidos cuya síntesis se detalla en esta tesis. Figura preparada por el Dr. Joan Massagué.

En primer lugar, se ha utilizado la SPPS para sintetizar una familia de fosfopéptidos que representan diversos estados de fosforilación de la región desordenada inter-dominio de Smad1 y Smad3. En general, la incorporación de uno o más (hasta 4) aminoácidos fosforilados a secuencias relativamente largas (hasta 36 residuos) dificulta la obtención de los péptidos con el grado de pureza y en las cantidades requeridas. Así, fue necesaria la optimización de varios aspectos de estas síntesis. Los péptidos obtenidos fueron utilizados en nuestro laboratorio para un completo estudio descrito en las tesis de la Dra. Görner y el Dr. Aragón que reveló las bases moleculares detrás de un complejo mecanismo de regulación de la actividad y la degradación de los R-Smads (Aragón et al., 2011). Se encontró el fundamento

biofísico y estructural que permite a YAP y Pin1 asociarse preferencialmente y cooperar con Smad1 y Smad3 respectivamente después de las fosforilaciones por CDK8/9, así como el mecanismo por el que dichas preferencias se decantan a favor de Smurf1 y Nedd4L después de las fosforilaciones por GSK3- β (**Figura 10.12**).

Respecto a la conformación latente de Nedd4L, hemos encontrado que el dominio HECT y el Ca^{2+} se unen a zonas solapadas en el dominio C2 (**Figura 10.13A y B**). Hemos observado que en presencia de Ca^{2+} el complejo HECT-C2 no puede ser detectado por RMN. La elevada afinidad de la interacción del Ca^{2+} con el dominio C2 contrasta con el carácter transitorio de los contactos auto-inhibitorios. Así, hemos corroborado que la presencia de Ca^{2+} obstaculiza los contactos entre los dominios HECT y C2, proporcionando el fundamento estructural y biofísico para la activación de la ligasa. Además el Ca^{2+} favorece la interacción del IP_3 también en la misma región del dominio C2 (**Figura 10.13B**). La liberación de IP_3 al citoplasma, resultado de la hidrólisis del fosfolípido de membrana PIP_2 , es un evento central que provoca incrementos intracelulares de la concentración de Ca^{2+} . Proponemos que la señalización por $\text{IP}_3/\text{Ca}^{2+}$ actúa modulando las poblaciones activa e inactiva de la ligasa así como su localización. Mediante contactos con el dominio C2 y dependiendo de la ratio PIP_2/IP_3 , la ligasa puede localizarse y ejercer su función ya sea en el citoplasma (mediante interacciones con el IP_3) o en la membrana plasmática (unida a PIP_2) (**Figura 10.13B y C**) (Escobedo et al., 2014).

También hemos investigado un posible mecanismo de auto-ubiquitinación de la ligasa, basado en el reconocimiento por parte de los dominios WW del motivo PY en el dominio HECT cuando dicho dominio está desplegado (Bruce et al., 2008). Mediante experimentos de relajación, hemos comprobado que la tirosina del motivo se mantiene casi inaccesible en solución en el dominio plegado. Hemos usado simulaciones de dinámica molecular que muestran que cuando se induce su desestabilización, el residuo puede resultar expuesto. Dicha desestabilización puede ocurrir cuando la proteína envejece en la célula y requiere ser degradada. Mediante ITC, hemos comprobado que el dominio WW3 une con la mayor afinidad al motivo HECT-PY expuesto, y hemos resuelto la estructura tridimensional de este complejo mediante RMN. De acuerdo con este escenario, proponemos un mecanismo cruzado de auto-ubiquitinación, en qué una molécula funcional de Nedd4L reconoce a otra parcialmente desestabilizada y la ubiquitina para su degradación (**Figura 10.13C**).

Además, hemos aprovechado esta extensa caracterización del complejo WW3-HECT PY para realizar un estudio preliminar con marcadores paramagnéticos que nos permitirá usar esta técnica en futuros experimentos en nuestro laboratorio.

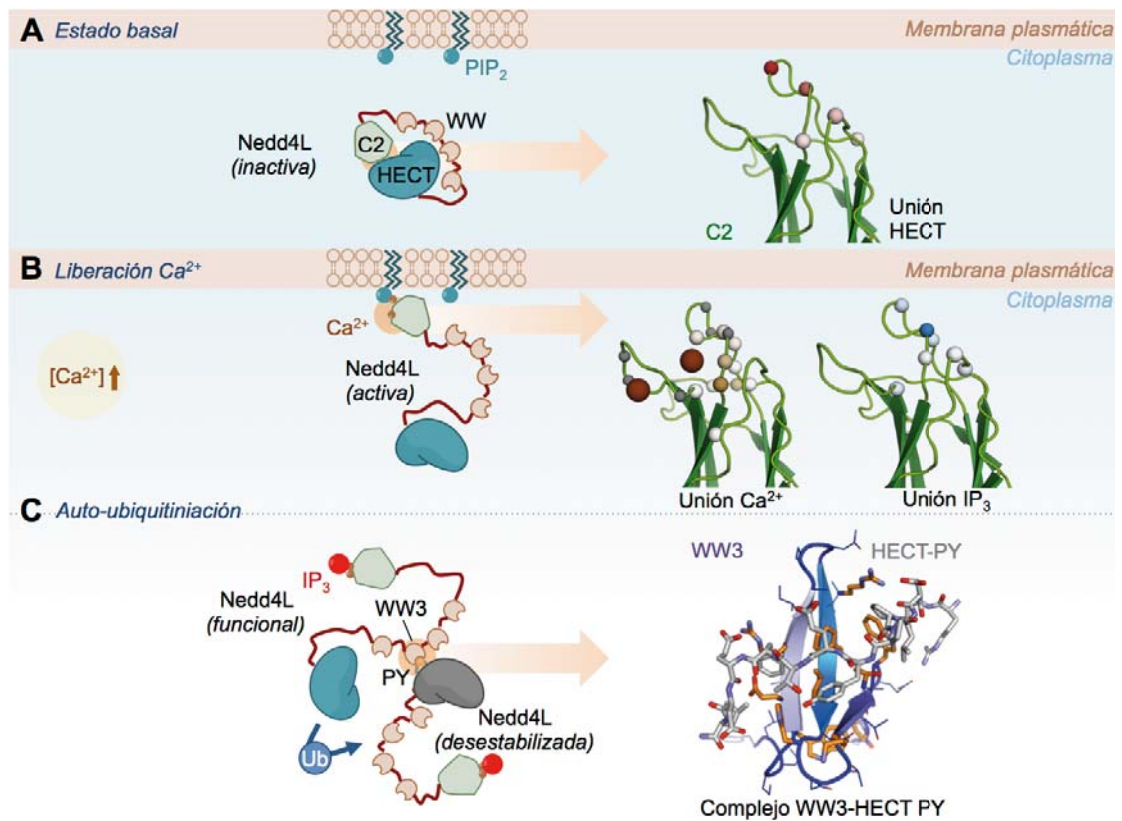


Figura 10.13: Resumen de los modelos propuestos para la activación y la degradación de Nedd4L. Junto a cada modelo se muestra el fundamento estructural en el que se basa.

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- β Pathways. *Cell* *139*, 757-769.
- Albesa, M., Grilo, L.S., Gavillet, B., and Abriel, H. (2011). Nedd4-2-dependent ubiquitylation and regulation of the cardiac potassium channel hERG1. *Journal of Molecular and Cellular Cardiology* *51*, 90-98.
- An, H., Krist, D.T., and Statsyuk, A.V. (2014). Crosstalk between kinases and Nedd4 family ubiquitin ligases. *Molecular BioSystems* *10*, 1643-1657.
- Aragón, E., Goerner, N., Xi, Q., Gomes, T., Gao, S., Massagué, J., and Macias, M.J. (2012). Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF-beta Pathways. *Structure* *20*, 1726-1736.
- 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 & development* *25*, 1275-1288.
- Attard, T.J., O'Brien-Simpson, N.M., and Reynolds, E.C. (2009). Identification and suppression of β -elimination byproducts arising from the use of Fmoc-Ser(PO₃Bzl,H)-OH in peptide synthesis. *International Journal of Peptide Research and Therapeutics* *15*, 69-79.
- Baca, M., Muir, T.W., Schnölzer, M., and Kent, S.B.H. (1995). Chemical ligation of cysteine-containing peptides: Synthesis of a 22 kDa tethered dimer of HIV-1 protease. *Journal of the American Chemical Society* *117*, 1881-1887.
- Backes, B.J., and Ellman, J.A. (1999). An alkanesulfonamide 'safety-catch' linker for solid-phase synthesis. *Journal of Organic Chemistry* *64*, 2322-2330.
- Bartels, C., Xia, T.H., Billeter, M., Guntert, P., and Wuthrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *Journal of biomolecular NMR* *6*, 1-10.
- Baumeister, W., Walz, J., Zühl, F., and Seemüller, E. (1998). The proteasome: Paradigm of a self-compartmentalizing protease. *Cell* *92*, 367-380.
- Bernassola, F., Karin, M., Ciechanover, A., and Melino, G. (2008). The HECT Family of E3 Ubiquitin Ligases: Multiple Players in Cancer Development. *Cancer Cell* *14*, 10-21.
- Berridge, M.J. (1993). Inositol trisphosphate and calcium signalling. *Nature* *361*, 315-325.
- Berridge, M.J. (2009). Inositol trisphosphate and calcium signalling mechanisms. *Biochimica et Biophysica Acta - Molecular Cell Research* *1793*, 933-940.
- Berridge, M.J., Lipp, P., and Bootman, M.D. (2000). The versatility and universality of calcium signalling. *Nature Reviews Molecular Cell Biology* *1*, 11-21.
- Bhattacharya, S., Zoladek, T., and Haines, D.S. (2008). WW domains 2 and 3 of Rsp5p play overlapping roles in binding to the LPKY motif of Spt23p and Mga2p. *International Journal of Biochemistry and Cell Biology* *40*, 147-157.
- Bieri, M., Kwan, A.H., Mobli, M., King, G.F., MacKay, J.P., and Gooley, P.R. (2011). Macromolecular NMR spectroscopy for the non-spectroscopist: Beyond

- macromolecular solution structure determination. *FEBS Journal* 278, 704-715.
- Blaustein, M.P., and Lederer, W.J. (1999). Sodium/calcium exchange: Its physiological implications. *Physiological Reviews* 79, 763-854.
- Bruce, M.C., Kanelis, V., Fouladkou, F., Debonneville, A., Staub, O., and Rotin, D. (2008). Regulation of Nedd4-2 self-ubiquitination and stability by a PY motif located within its HECT-domain. *The Biochemical journal* 415, 155-163.
- Brunger, 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 crystallographica Section D, Biological crystallography* 54, 905-921.
- Budd, S.L., and Nicholls, D.G. (1996). A reevaluation of the role of mitochondria in neuronal Ca²⁺ homeostasis. *Journal of Neurochemistry* 66, 403-411.
- Burgess, G.M., Godfrey, P.P., McKinney, J.S., Berridge, M.J., Irvine, R.F., and Putney Jr, J.W. (1984). The second messenger linking receptor activation to internal Ca release in liver. *Nature* 309, 63-66.
- Cao, X.R., Lill, N.L., Boase, N., Shi, P.P., Croucher, D.R., Shan, H., Qu, J., Sweezer, E.M., Place, T., Kirby, P.A., *et al.* (2008). Nedd4 controls animal growth by regulating IGF-1 signaling. *Science Signaling* 1.
- Caohuy, H., Jozwik, C., and Pollard, H.B. (2009). Rescue of Δ F508-CFTR by the SGK1/Nedd4-2 signaling pathway. *Journal of Biological Chemistry* 284, 25241-25253.
- Chen, C., and Matesic, L.E. (2007). The Nedd4-like family of E3 ubiquitin ligases and cancer. *Cancer and Metastasis Reviews* 26, 587-604.
- Cheng, H., Lederer, W.J., and Cannell, M.B. (1993). Calcium sparks: elementary events underlying excitation-contraction coupling in heart muscle. *Science* 262, 740-744.
- Cheng, P.L., Lu, H., Shelly, M., Gao, H., and Poo, M.M. (2011). Phosphorylation of E3 ligase smurf1 switches its substrate preference in support of axon development. *Neuron* 69, 231-243.
- Cho, W., and Stahelin, R.V. (2006). Membrane binding and subcellular targeting of C2 domains. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1761, 838-849.
- Christensen, T. (1979). A Qualitative Test for Monitoring Coupling Completeness in Solid Phase Peptide Synthesis Using Chloranil. *Acta Chem Scand* 33B, 763.
- Ciechanover, A., Heller, H., Elias, S., Haas, A.L., and Hershko, A. (1980). ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proceedings of the National Academy of Sciences of the United States of America* 77, 1365-1368.
- Clapham, D.E. (2007). Calcium Signaling. *Cell* 131, 1047-1058.
- Clapper, D.L., Walseth, T.F., Dargie, P.J., and Hon Cheung, L. (1987). Pyridine nucleotide metabolites stimulate calcium release from sea urchin egg microsomes desensitized to inositol trisphosphate. *Journal of Biological Chemistry* 262, 9561-9568.

- Corbalan-Garcia, S., Garcia-Garcia, J., Rodriguez-Alfaro, J.A., and Gomez-Fernandez, J.C. (2003). A new phosphatidylinositol 4,5-bisphosphate-binding site located in the C2 domain of protein kinase Calpha. *The Journal of biological chemistry* *278*, 4972-4980.
- Dalgarno, D.C., Levine, B.A., and Williams, R.J.P. (1983). Structural information from NMR secondary chemical shifts of peptide α C-H protons in proteins. *Bioscience Reports* *3*, 443-452.
- Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S.B.H. (1994). Synthesis of proteins by native chemical ligation. *Science* *266*, 776-779.
- De Bie, P., and Ciechanover, A. (2011). Ubiquitination of E3 ligases: Self-regulation of the ubiquitin system via proteolytic and non-proteolytic mechanisms. *Cell Death and Differentiation* *18*, 1393-1402.
- De Koninck, P., and Schulman, H. (1998). Sensitivity of CaM kinase II to the frequency of Ca²⁺ oscillations. *Science* *279*, 227-230.
- 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. *Journal of biomolecular NMR* *6*, 277-293.
- DeLano, W.L. (2002). The PyMOL Molecular Graphics System. <http://www.pymol.org/>.
- Ding, Y., Zhang, Y., Xu, C., Tao, Q.H., and Chen, Y.G. (2013). HECT domain-containing E3 ubiquitin ligase NEDD4L negatively regulates Wnt signaling by targeting dishevelled for proteasomal degradation. *Journal of Biological Chemistry* *288*, 8289-8298.
- Doreleijers, J.F., Vranken, W.F., Schulte, C., Markley, J.L., Ulrich, E.L., Vriend, G., and Vuister, G.W. (2012). NRG-CING: integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB. *Nucleic acids research* *40*, D519-524.
- Doucleff, M., Hatcher-Skeers, M., and Crane, N.J. (2011). Pocket guide to biomolecular NMR.
- Duchen, M.R. (1999). Contributions of mitochondria to animal physiology: From homeostatic sensor to calcium signalling and cell death. *Journal of Physiology* *516*, 1-17.
- Dunn, R., Klos, D.A., Adler, A.S., and Hicke, L. (2004). The C2 domain of the Rsp5 ubiquitin ligase binds membrane phosphoinositides and directs ubiquitination of endosomal cargo. *The Journal of cell biology* *165*, 135-144.
- Ebisawa, T., Fukuchi, M., Murakami, G., Chiba, T., Tanaka, K., Imamura, T., and Miyazono, K. (2001). Smurf1 Interacts with Transforming Growth Factor- β Type I Receptor through Smad7 and Induces Receptor Degradation. *Journal of Biological Chemistry* *276*, 12477-12480.
- Escobedo, A., Gomes, T., Aragon, E., Martin-Malpartida, P., Ruiz, L., and Macias, M.J. (2014). Structural Basis of the Activation and Degradation Mechanisms of the E3 Ubiquitin Ligase Nedd4L. *Structure* *22*, 1446-1457.
- Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* *33*, 5984-6003.

- Finch, E.A., Turner, T.J., and Goldin, S.M. (1991). Calcium as a coagonist of inositol 1,4,5-trisphosphate-induced calcium release. *Science* *252*, 443-446.
- Flasza, M., Gorman, P., Roylance, R., Canfield, A.E., and Baron, M. (2002). Alternative splicing determines the domain structure of WWP1, a Nedd4 family protein. *Biochemical and Biophysical Research Communications* *290*, 431-437.
- Fouladkou, F., Landry, T., Kawabe, H., Neeb, A., Lu, C., Brose, N., Stambolic, V., and Rotin, D. (2008). The ubiquitin ligase Nedd4-1 is dispensable for the regulation of PTEN stability and localization. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 8585-8590.
- 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.
- Galan, J.M., and Peter, M. (1999). Ubiquitin-dependent degradation of multiple F-box proteins by an autocatalytic mechanism. *Proceedings of the National Academy of Sciences of the United States of America* *96*, 9124-9129.
- Gallagher, E., Gao, M., Liu, Y.C., and Karin, M. (2006). Activation of the E3 ubiquitin ligase Itch through a phosphorylation-induced conformational change. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 1717-1722.
- 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- β Signaling. *Molecular Cell* *36*, 457-468.
- García-Martín, F., Quintanar-Audelo, M., García-Ramos, Y., Cruz, L.J., Gravel, C., Furic, R., Côté, S., Tulla-Puche, J., and Albericio, F. (2006). ChemMatrix, a poly(ethylene glycol)-based support for the solid-phase synthesis of complex peptides. *Journal of Combinatorial Chemistry* *8*, 213-220.
- Garrone, N.F., Blazer-Yost, B.L., Weiss, R.B., Lalouel, J.M., and Rohrwasser, A. (2009). A human polymorphism affects NEDD4L subcellular targeting by leading to two isoforms that contain or lack a C2 domain. *BMC cell biology* *10*, 26.
- Guerrero-Valero, M., Ferrer-Orta, C., Querol-Audí, J., Marin-Vicente, C., Fita, I., Gómez-Fernández, J.C., Verdager, N., and Corbalán-García, S. (2009). Structural and mechanistic insights into the association of PKC α -C2 domain to PtdIns(4,5)P₂. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 6603-6607.
- Haas, A.L., Warms, J.V., Hershko, A., and Rose, I.A. (1982). Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. *Journal of Biological Chemistry* *257*, 2543-2548.
- Harris, P.W.R., Williams, G.M., Shepherd, P., and Brimble, M.A. (2008). The synthesis of phosphopeptides using microwave-assisted solid phase peptide synthesis. *International Journal of Peptide Research and Therapeutics* *14*, 387-392.
- Harvey, K.F., Dinudom, A., Cook, D.I., and Kumar, S. (2001). The Nedd4-like Protein KIAA0439 is a Potential Regulator of the Epithelial Sodium Channel. *Journal of Biological Chemistry* *276*, 8597-8601.

- 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 and Development* 12, 186-197.
- Hein, C., Springael, J.Y., Volland, C., Haguenaer-Tsapis, R., and Andre, B. (1995). NPI1, an essential yeast gene involved in induced degradation of Gap1 and Fur4 permeases, encodes the Rsp5 ubiquitin-protein ligase. *Molecular Microbiology* 18, 77-87.
- Hershko, A., Ciechanover, A., Heller, H., Haas, A.L., and Rose, I.A. (1980). Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proceedings of the National Academy of Sciences of the United States of America* 77, 1783-1786.
- Hershko, A., Ciechanover, A., and Rose, I.A. (1979). Resolution of the ATP-dependent proteolytic system from reticulocytes: a component that interacts with ATP. *Proceedings of the National Academy of Sciences of the United States of America* 76, 3107-3110.
- Hershko, A., Heller, H., Elias, S., and Ciechanover, A. (1983). Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *Journal of Biological Chemistry* 258, 8206-8214.
- Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* 4, 435-447.
- Hesselberth, J.R., Miller, J.P., Golob, A., Stajich, J.E., Michaud, G.A., and Fields, S. (2006). Comparative analysis of *Saccharomyces cerevisiae* WW domains and their interacting proteins. *Genome Biology* 7.
- Hoeller, D., Hecker, C.M., and Dikic, I. (2006). Ubiquitin and ubiquitin-like proteins in cancer pathogenesis. *Nature Reviews Cancer* 6, 776-788.
- Hu, G., and Fearon, E.R. (1999). Siah-1 N-terminal RING domain is required for proteolysis function, and C-terminal sequences regulate oligomerization and binding to target proteins. *Molecular and Cellular Biology* 19, 724-732.
- Huibregtse, J.M., Scheffner, M., Beaudenon, S., and Howley, P.M. (1995). A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proceedings of the National Academy of Sciences of the United States of America* 92, 5249.
- Huminiacki, L., Goldovsky, L., Freilich, S., Moustakas, A., Ouzounis, C., and Heldin, C.H. (2009). Emergence, development and diversification of the TGF-signalling pathway within the animal kingdom. *BMC Evolutionary Biology* 9.
- Huse, M., Holford, M.N., Kuriyan, J., and Muir, T.W. (2000). Semisynthesis of hyperphosphorylated type I TGF β receptor: Addressing the mechanism of kinase activation [20]. *Journal of the American Chemical Society* 122, 8337-8338.
- Hyvonen, M., Macias, M.J., Nilges, M., Oschkinat, H., Saraste, M., and Wilmanns, M. (1995). Structure of the binding site for inositol phosphates in a PH domain. *EMBO Journal* 14, 4676-4685.
- Ichimura, T., Yamamura, H., Sasamoto, K., Tominaga, Y., Taoka, M., Kakiuchi, K., Shinkawa, T., Takahashi, N., Shimada, S., and Isobe, T. (2005). 14-3-3 proteins modulate the expression of epithelial Na⁺ channels by

- phosphorylation-dependent interaction with Nedd4-2 ubiquitin ligase. *The Journal of biological chemistry* 280, 13187-13194.
- Ingham, R.J., Gish, G., and Pawson, T. (2004). The Nedd4 family of E3 ubiquitin ligases: Functional diversity within a common modular architecture. *Oncogene* 23, 1972-1984.
- Inman, G.J., Nicolás, F.J., and Hill, C.S. (2002). Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF- β receptor activity. *Molecular Cell* 10, 283-294.
- Itoh, S., Landström, M., Hermansson, A., Itoh, F., Heldin, C.H., Heldin, N.E., and Ten Dijke, P. (1998). Transforming growth factor β 1 induces nuclear export of inhibitory Smad7. *Journal of Biological Chemistry* 273, 29195-29201.
- Jennings, M.D., Blankley, R.T., Baron, M., Golovanov, A.P., and Avis, J.M. (2007). Specificity and autoregulation of notch binding by tandem wwdomains in suppressor of deltex. *Journal of Biological Chemistry* 282, 29032-29042.
- Johnson, E.C.B., and Kent, S.B.H. (2006). Insights into the mechanism and catalysis of the native chemical ligation reaction. *Journal of the American Chemical Society* 128, 6640-6646.
- Johnson, L.N., and Lewis, R.J. (2001). Structural basis for control by phosphorylation. *Chemical Reviews* 101, 2209-2242.
- Kaiser, E., Colescott, R.L., Bossinger, C.D., and Cook, P.I. (1970). Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Analytical Biochemistry* 34, 595-598.
- Kamadurai, H.B., Souphron, J., Scott, D.C., Duda, D.M., Miller, D.J., Stringer, D., Piper, R.C., and Schulman, B.A. (2009). Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. *Mol Cell* 36, 1095-1102.
- Kamynina, E., Dedonneville, C., Dens, M., Vandewalle, A., and Staub, O. (2001). A novel mouse Nedd4 protein suppresses the activity of the epithelial Na⁺ channel. *FASEB Journal* 15, 204-214.
- 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.
- Kato, Y., Ito, M., Kawai, K., Nagata, K., and Tanokura, M. (2002). Determinants of ligand specificity in groups I and IV WW domains as studied by surface plasmon resonance and model building. *Journal of Biological Chemistry* 277, 10173-10177.
- Kato, Y., Nagata, K., Takahashi, M., Lian, L., Herrero, J.J., Sudol, M., and Tanokura, M. (2004). Common mechanism of ligand recognition by group II/III WW domains: Redefining their functional classification. *Journal of Biological Chemistry* 279, 31833-31841.
- 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 β receptor for degradation. *Molecular Cell* 6, 1365-1375.
- Keeler, J. (2010). *Understanding NMR Spectroscopy* (Wiley).
- Kent, S., Sohma, Y., Liu, S., Bang, D., Pentelute, B., and Mandal, K. (2012). Through the looking glass - a new world of proteins enabled by chemical synthesis. *Journal of Peptide Science* 18, 428-436.

- Komander, D., Clague, M.J., and Urbe, S. (2009). Breaking the chains: structure and function of the deubiquitinases. *Nature reviews Molecular cell biology* 10, 550-563.
- Komander, D., and Rape, M. (2012). The ubiquitin code. In *Annual Review of Biochemistry*, pp. 203-229.
- Kuratomi, G., Komuro, A., Goto, K., Shinozaki, M., Miyazawa, K., Miyazono, K., and Imamura, T. (2005). NEDD4-2 (neural precursor cell expressed, developmentally down-regulated 4-2) negatively regulates TGF- β (transforming growth factor- β) signalling by inducing ubiquitin-mediated degradation of Smad2 and TGF- β type I receptor. *Biochemical Journal* 386, 461-470.
- Kwak, Y.D., Wang, B., Li, J.J., Wang, W., Deng, Q., Diao, S., Chen, Y., Xu, R., Masliah, E., Xu, H., *et al.* (2012). Upregulation of the E3 ligase NEDD4-1 by oxidative stress degrades IGF-1 receptor protein in neurodegeneration. *Journal of Neuroscience* 32, 10971-10981.
- Kwan, A.H., Mobli, M., Gooley, P.R., King, G.F., and MacKay, J.P. (2011). Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS Journal* 278, 687-703.
- Ladbury, J.E., and Chowdhry, B.Z. (1996). Sensing the heat: The application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chemistry and Biology* 3, 791-801.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
- 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. *Journal of biomolecular NMR* 8, 477-486.
- Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P.D., Smith, C.A., Sheffler, W., *et al.* (2011). ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology* 487, 545-574.
- Lemmon, M.A. (2008). Membrane recognition by phospholipid-binding domains. *Nature Reviews Molecular Cell Biology* 9, 99-111.
- Levitt, M.H. (2008). *Spin Dynamics: Basics of Nuclear Magnetic Resonance* (Wiley).
- Li, Y., Ozaki, T., Kikuchi, H., Yamamoto, H., Ohira, M., and Nakagawara, A. (2008). A novel HECT-type E3 ubiquitin protein ligase NEDL1 enhances the p53-mediated apoptotic cell death in its catalytic activity-independent manner. *Oncogene* 27, 3700-3709.
- Lian, L.Y., and Roberts, G. (2011). *Protein NMR Spectroscopy: Practical Techniques and Applications*.
- 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 TGF β Signaling. *Cell* 125, 915-928.
- Lin, X., Liang, M., and Feng, X.H. (2000). Smurf2 is a ubiquitin E3 ligase mediating proteasome-dependent degradation of Smad2 in transforming growth factor- β signaling. *Journal of Biological Chemistry* 275, 36818-36822.

- Lorick, K.L., Jensen, J.P., Fang, S., Ong, A.M., Hatakeyama, S., and Weissman, A.M. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proceedings of the National Academy of Sciences of the United States of America* *96*, 11364-11369.
- 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. *Nature Cell Biology* *10*, 994-1002.
- Macias, M.J., Gervais, V., Civera, C., and Oschkinat, H. (2000). Structural analysis of WW domains and design of a WW prototype. *Nature Structural Biology* *7*, 375-379.
- 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.
- Macias, M.J., Wiesner, S., and Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Letters* *513*, 30-37.
- Maddika, S., Kavela, S., Rani, N., Palicharla, V.R., Pokorny, J.L., Sarkaria, J.N., and Chen, J. (2011). WWP2 is an E3 ubiquitin ligase for PTEN. *Nature Cell Biology* *13*, 728-733.
- Mandell, D.J., Coutsiias, E.A., and Kortemme, T. (2009). Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature methods* *6*, 551-552.
- Mann, M., and Jensen, O.N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology* *21*, 255-261.
- Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *Journal of biomolecular NMR* *20*, 71-75.
- Massagué, J. (1998). TGF- β signal transduction. In *Annual Review of Biochemistry*, pp. 753-791.
- Massagué, J. (2000). How cells read TGF- β signals. *Nature Reviews Molecular Cell Biology* *1*, 169-178.
- Massagué, J. (2003). Integration of Smad and MAPK pathways: A link and a linker revisited. *Genes and Development* *17*, 2993-2997.
- Massagué, J. (2012). TGF β signalling in context. *Nature Reviews Molecular Cell Biology* *13*, 616-630.
- 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- β -induced migration and invasion. *Journal of Biological Chemistry* *285*, 1754-1764.
- Matsuura, I., Denissova, N.G., Wang, G., He, D., Long, J., and Liu, F. (2004). Cyclin-dependent kinases regulate the antiproliferative function of Smads. *Nature* *430*, 226-231.
- Matsuzaki, K., Kitano, C., Murata, M., Sekimoto, G., Yoshida, K., Uemura, Y., Seki, T., Taketani, S., Fujisawa, J.I., and Okazaki, K. (2009). Smad2 and Smad3 phosphorylated at both linker and COOH-terminal regions transmit malignant TGF- β signal in later stages of human colorectal cancer. *Cancer Research* *69*, 5321-5330.
- Merrifield, R.B. (1963). Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *Journal of the American Chemical Society* *85*, 2149-2154.

- Metzger, M.B., Hristova, V.A., and Weissman, A.M. (2012). HECT and RING finger families of E3 ubiquitin ligases at a glance. *Journal of Cell Science* 125, 531-537.
- Miranda, M., and Sorkin, A. (2007). Regulation of receptors and transporters by ubiquitination: New insights into surprisingly similar mechanisms. *Molecular Interventions* 7, 157-167.
- Miyazaki, K., Fujita, T., Ozaki, T., Kato, C., Kurose, Y., Sakamoto, M., Kato, S., Goto, T., Itoyama, Y., Aoki, M., *et al.* (2004). NEDL1, a Novel Ubiquitin-protein Isopeptide Ligase for Dishevelled-1, Targets Mutant Superoxide Dismutase-1. *Journal of Biological Chemistry* 279, 11327-11335.
- Miyazaki, K., Ozaki, T., Kato, C., Hanamoto, T., Fujita, T., Irino, S., Watanabe, K.I., Nakagawa, T., and Nakagawara, A. (2003). A novel HECT-type E3 ubiquitin ligase, NEDL2, stabilizes p73 and enhances its transcriptional activity. *Biochemical and Biophysical Research Communications* 308, 106-113.
- 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.
- Mullen, A.C., Orlando, D.A., Newman, J.J., Lovén, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., Dekoter, R.P., and Young, R.A. (2011). Master transcription factors determine cell-type-specific responses to TGF- β signaling. *Cell* 147, 565-576.
- Mund, T., and Pelham, H.R. (2009). Control of the activity of WW-HECT domain E3 ubiquitin ligases by NDFIP proteins. *EMBO reports* 10, 501-507.
- Murray, D., and Honig, B. (2002). Electrostatic control of the membrane targeting of C2 domains. *Molecular Cell* 9, 145-154.
- 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.
- Nuber, U., Schwarz, S.E., and Scheffner, M. (1998). The ubiquitin-protein ligase E6-associated protein (E6-AP) serves as its own substrate. *European Journal of Biochemistry* 254, 643-649.
- Oancea, E., and Meyer, T. (1998). Protein kinase C as a molecular machine for decoding calcium and diacylglycerol signals. *Cell* 95, 307-318.
- Okamoto, K., Taya, Y., and Nakagama, H. (2009). Mdmx enhances p53 ubiquitination by altering the substrate preference of the Mdm2 ubiquitin ligase. *FEBS Letters* 583, 2710-2714.
- Payne, R.J., Ficht, S., Greenberg, W.A., and Wong, C.H. (2008). Cysteine-free peptide and glycopeptide ligation by direct aminolysis. *Angewandte Chemie - International Edition* 47, 4411-4415.
- Perry, W.L., Hustad, C.M., Swing, D.A., Norene O'Sullivan, T., Jenkins, N.A., and Copeland, N.G. (1998). The itchy locus encodes a novel ubiquitin protein ligase that is disrupted in a(18H) mice. *Nature Genetics* 18, 143-146.
- Persaud, A., Alberts, P., Amsen, E.M., Xiong, X., Wasmuth, J., Saadon, Z., Fladd, C., Parkinson, J., and Rotin, D. (2009). Comparison of substrate specificity of

- the ubiquitin ligases Nedd4 and Nedd4-2 using proteome arrays. *Molecular Systems Biology* 5.
- Petroski, M.D. (2008). The ubiquitin system, disease, and drug discovery. *BMC Biochemistry* 9.
- Petroski, M.D., and Deshaies, R.J. (2005). Function and regulation of cullin-RING ubiquitin ligases. *Nature Reviews Molecular Cell Biology* 6, 9-20.
- 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 GTPPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. *Journal of Molecular Biology* 314, 1147-1156.
- Plant, P.J., Yeager, H., Staub, O., Howard, P., and Rotin, D. (1997). The C2 domain of the ubiquitin protein ligase Nedd4 mediates Ca²⁺-dependent plasma membrane localization. *Journal of Biological Chemistry* 272, 32329-32336.
- Pozzan, T., Rizzuto, R., Volpe, P., and Meldolesi, J. (1994). Molecular and cellular physiology of intracellular calcium stores. *Physiological Reviews* 74, 595-636.
- Rahighi, S., Ikeda, F., Kawasaki, M., Akutsu, M., Suzuki, N., Kato, R., Kensche, T., Uejima, T., Bloor, S., Komander, D., *et al.* (2009). Specific Recognition of Linear Ubiquitin Chains by NEMO Is Important for NF- κ B Activation. *Cell* 136, 1098-1109.
- Rizo, J., and Sudhof, T.C. (1998). C2-domains, structure and function of a universal Ca²⁺-binding domain. *The Journal of biological chemistry* 273, 15879-15882.
- Ryan, P.E., Davies, G.C., Nau, M.M., and Lipkowitz, S. (2006). Regulating the regulator: Negative regulation of Cbl ubiquitin ligases. *Trends in Biochemical Sciences* 31, 79-88.
- Sapkota, G., Alarcón, C., Spagnoli, F.M., Brivanlou, A.H., and Massagué, J. (2007). Balancing BMP Signaling through Integrated Inputs into the Smad1 Linker. *Molecular Cell* 25, 441-454.
- Sato, Y., Yoshikawa, A., Mimura, H., Yamashita, M., Yamagata, A., and Fukai, S. (2009). Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *EMBO Journal* 28, 2461-2468.
- Scheffner, M., and Staub, O. (2007). HECT E3s and human disease. *BMC Biochemistry* 8.
- Scialpi, F., Malatesta, M., Peschiaroli, A., Rossi, M., Melino, G., and Bernassola, F. (2008). Itch self-polyubiquitylation occurs through lysine-63 linkages. *Biochemical Pharmacology* 76, 1515-1521.
- Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR* 44, 213-223.
- Shi, Y., and Massagué, J. (2003). Mechanisms of TGF- β signaling from cell membrane to the nucleus. *Cell* 113, 685-700.
- Sims, J.J., and Cohen, R.E. (2009). Linkage-Specific Avidity Defines the Lysine 63-Linked Polyubiquitin-Binding Preference of Rap80. *Molecular Cell* 33, 775-783.

- Soond, S.M., and Chantry, A. (2011). Selective targeting of activating and inhibitory Smads by distinct WWP2 ubiquitin ligase isoforms differentially modulates TGF B signalling and EMT. *Oncogene* *30*, 2451-2462.
- Staub, O., Dho, S., Henry, P.C., Correa, J., Ishikawa, T., McGlade, J., and Rotin, D. (1996). WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na⁺ channel deleted in Liddle's syndrome. *EMBO Journal* *15*, 2371-2380.
- Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* *403*, 41-45.
- Streb, H., Irvine, R.F., Berridge, M.J., and Schulz, I. (1983). Release of Ca²⁺ from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate. *Nature* *306*, 67-69.
- Su, X.C., and Otting, G. (2010). Paramagnetic labelling of proteins and oligonucleotides for NMR. *Journal of Biomolecular NMR* *46*, 101-112.
- Taichi, M., Kimura, T., and Nishiuchi, Y. (2009). Suppression of side reactions during final deprotection employing a strong acid in Boc chemistry: Regeneration of methionyl residues from their sulfonium salts. *International Journal of Peptide Research and Therapeutics* *15*, 247-253.
- Tanaka, K., Waxman, L., and Goldberg, A.L. (1983). ATP serves two distinct roles in protein degradation in reticulocytes, one requiring and one independent of ubiquitin. *Journal of Cell Biology* *96*, 1580-1585.
- Theobald, D.L., and Steindel, P.A. (2012). Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics* *28*, 1972-1979.
- Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO Journal* *19*, 94-102.
- Trompouki, E., Bowman, T.V., Lawton, L.N., Fan, Z.P., Wu, D.C., Dibiase, A., Martin, C.S., Cech, J.N., Sessa, A.K., Leblanc, J.L., *et al.* (2011). Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* *147*, 577-589.
- Verdaguer, N., Corbalan-Garcia, S., Ochoa, W.F., Fita, I., and Gómez-Fernández, J.C. (1999). Ca²⁺ bridges the C2 membrane-binding domain of protein kinase C α directly to phosphatidylserine. *EMBO Journal* *18*, 6329-6338.
- Verdecia, M.A., Joazeiro, C.A., Wells, N.J., Ferrer, J.L., Bowman, M.E., Hunter, T., and Noel, J.P. (2003). Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. *Mol Cell* *11*, 249-259.
- Vojtkovsky, T. (1995). Detection of secondary amines on solid phase. *Peptide research* *8*, 236-237.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* *59*, 687-696.
- 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.
- Wang, J., Peng, Q., Lin, Q., Childress, C., Carey, D., and Yang, W. (2010). Calcium activates Nedd4 E3 ubiquitin ligases by releasing the C2 domain-mediated auto-inhibition. *The Journal of biological chemistry* *285*, 12279-12288.

- Wang, X., Trotman, L.C., Koppie, T., Alimonti, A., Chen, Z., Gao, Z., Wang, J., Erdjument-Bromage, H., Tempst, P., Cordon-Cardo, C., *et al.* (2007). NEDD4-1 Is a Proto-Oncogenic Ubiquitin Ligase for PTEN. *Cell* *128*, 129-139.
- Wiemuth, D., Lott, J.S., Ly, K., Ke, Y., Teesdale-Spittle, P., Snyder, P.M., and McDonald, F.J. (2010). Interaction of serum- and Glucocorticoid regulated Kinase 1 (SGK1) with the WW-domains of Nedd4-2 is required for epithelial sodium channel regulation. *PLoS ONE* *5*.
- 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.
- Wiseman, T., Williston, S., Brandts, J.F., and Lin, L.N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical Biochemistry* *179*, 131-137.
- Xu, L., Kang, Y., Çöl, S., and Massagué, J. (2002). Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGF β signaling complexes in the cytoplasm and nucleus. *Molecular Cell* *10*, 271-282.
- Yan, L.Z., and Dawson, P.E. (2001). Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *Journal of the American Chemical Society* *123*, 526-533.
- Yang, C., Zhou, W., Jeon, M.S., Demydenko, D., Harada, Y., Zhou, H., and Liu, Y.C. (2006). Negative regulation of the E3 ubiquitin ligase Itch via Fyn-mediated tyrosine phosphorylation. *Molecular Cell* *21*, 135-141.
- Yao, Y., Choi, J., and Parker, I. (1995). Quantal puffs of intracellular Ca²⁺ evoked by inositol trisphosphate in *Xenopus* oocytes. *Journal of Physiology* *482*, 533-553.
- Yeung, B., Ho, K.C., and Yang, X. (2013). WWP1 E3 Ligase Targets LATS1 for Ubiquitin-Mediated Degradation in Breast Cancer Cells. *PLoS ONE* *8*.
- Zhu, H., Kavsak, P., Abdollah, S., Wrana, J.L., and Thomsen, G.H. (1999). A SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation. *Nature* *400*, 687-693.

Abbreviations and Units

Å	Angstroms (10^{-10} m)
AA	Amino acid
Amp	Ampicillin
ACH	α -cyano-4-hydroxycinnamic acid
ACN	Acetonitrile
AcOH	Acetic acid
Aria	Ambiguous restraints for interactive assignment
Au	Absorbance units
BMP	Bone morphogenic protein
BMRB	Biological magnetic resonance bank
Boc	<i>t</i> -Butoxycarbonyl
Bzl	Benzyl
°C	Celsius
CDK8/9	G1 cyclin-dependent kinases 8 and 9
cDNA	Complimentary DNA
CNS	Crystallography & NMR system
Co-IP	Co-immunoprecipitation
Co-Smad	Common partner-Smad
CSP	Chemical shift perturbation
δ	Chemical shift (ppm)
Da	Dalton
DAG	Diacylglycerol
DBU	1,8-Diazabicycloundec-7-ene
DCM	Dichloromethane
DHB	2,5-Dihydroxybenzoic acid
DIC	<i>N,N'</i> -Diisopropylcarbodiimide
DIEA	<i>N,N</i> -Diisopropylethylamine
DMF	<i>N,N</i> -Dimethylformamide
DNA	Deoxyribonucleic acid
DODT	3,6-dioxa-1,8-octanedithiol
DTT	Dithiothreitol
DUB	Deubiquitinating enzyme
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethyleneglycoltetraacetic acid

Eq	Equivalents
ENaC	Epithelial sodium channel
ER	Endoplasmic reticulum
ESI	Electrospray ionization
ExPASy	Expert protein analysis system
FA	Formic acid
FID	Free induction decay
Fmoc	9 <i>H</i> -Fluoren-9-ylmethoxy carbonyl
FPLC	Fast protein liquid chromatography
Gdn-HCl	Guanidine hydrochloride
GPCR	G protein-coupled receptor
GSK3-β	Glycogen synthase kinase 3-beta
GST	Glutathion S-transferase
h	Hours
HATU	<i>N,N</i> -Methylmethan-aminium hexafluorophosphate
HECT	Homologous to E6-AP C-terminus
HOAt	1-Hydroxy-7-azabenzotriazole
HOBt	1-Hydroxybenzotriazole
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single-quantum correlation
I-Smad	Inhibitor-Smad
IP₃	Inositol 1,4,5-trisphosphate
IP₃R	IP ₃ receptor
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Itch	E3 ubiquitin-protein ligase Itchy homolog
K	Kelvin
KAN	Kanamycin
K_D	Dissociation constant
LB	Lysogeny broth
LC	Liquid chromatography
M	Molar (g/mol)
MALDI	Matrix-assisted laser desorption ionization
MAPK	Mitogen-activated protein kinases
MBP	Maltose-binding protein
MD	Molecular dynamics
MH domain	Mad-homology domain
min	Minutes
MPAA	4-mercaptophenylacetic acid

MS	Mass spectrometry
MTS	(1-Acetyl-2,2,5,5-tetramethylpyrroline-3-methyl)methanethiosulfonate
MTSL	(1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)methanethiosulfonate
Mw	Molecular weight
MW	Microwave
N	Stoichiometry parameter (a-dimensional)
NCL	Native chemical ligation
Nedd4	Neural precursor cell expressed developmentally down-regulated protein 4
Nedd4L	Neural precursor cell expressed developmentally down-regulated protein 4 like
NEDL1/2	Nedd4-like E3 ubiquitin-protein ligase 1/2
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NUS	Non-uniform sampling
NusA	N utilization substance protein A
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCS	Pseudo-contact shifts
PDB	Protein data bank
PEG	Polyethylene glycol
PH domain	Pleckstrin homology domain
Pin1	Protein interacting with NIMA-1 (never in mitosis A-1)
PIP₂	Phosphatidylinositol 4,5-bisphosphate
PKCα	Protein kinase C alpha
PMSF	Phenylmethanesulphonyl fluoride
ppm	Parts-per-million (10^{-6})
PRE	Paramagnetic relaxation enhancement
PTM	Post-translational modification
PyBOP	Benzotriazol-1-yl-oxytrypyrrolidinophosphonium
2-PySH	2-mercaptopyridine
RDC	Residual dipolar coupling
RF	Radio frequency
RING	Really interesting new gene
RNA	Ribonucleic acid

RP	Reversed phase
rpm	Rounds per minute
R-Smad	Receptor regulated-Smad
Rsp5	Reverses SPT-phenotype protein 5
RT	Room temperature
RTK	Receptor tyrosine kinase
RYR	Ryanodine receptor
s	Seconds
SA	Sinapinic acid
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sgk1	Serum- and glucocorticoid-regulated kinase 1
Smad	Small mothers against decapentaplegic
Smurf1/2	Smad ubiquitin regulatory factor 1/2
SOC	Super optimal broth with catabolite repression
SPPS	Solid phase peptide synthesis
tBu	<i>t</i> -Butyl
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TGF-β	Transforming growth factor beta
TIS	Triisopropylsilane
TOCSY	Total correlation spectroscopy
TOF	Time of flight
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
TROSY	Transverse relaxation optimized spectroscopy
Trt	Trityl
TGFβRI	TGF- β receptor type I
TGFβRII	TGF- β receptor type II
Ub	Ubiquitin
UPP	Ubiquitin-proteasome pathway
WT	Wild type
WWP1/2	WW domain containing protein 1/2
YAP	Yes-associated protein

Curriculum Vitae

Personal Information

Birth Date: 19 Oct 1987
Birth place: Lleida, Spain
Nationality: Spain

Contact

Carrer Congost 18 2n 4a
08024 Barcelona
Cell phone: +34 669920525

albert.escobedo@irbbarcelona.org

Education

09/09 – Current	Universitat de Barcelona PhD in Biomedicine	Barcelona, Spain
09/08 – 09/09	Cranfield University MSc in Molecular Medicine ★ <i>Course Director's Prize</i>	Cranfield, UK
09/05 – 09/09	Universitat de Lleida BSc in Biotechnology ★ <i>Spain's Ministry of Education and Science Special Mention</i> ★ <i>Universitat de Lleida End of Degree Extraordinary Award</i>	Lleida, Spain

Work Experience

09/09 – Current	Institute for Research in Biomedicine of Barcelona (IRB Barcelona) Pre-doctoral Researcher – Ministry of Education and Science fellow (FPU) <i>Major Workshops & Conferences:</i> <ul style="list-style-type: none">• 55th ENC (Experimental NMR Conference), 03/14, Boston USA – poster• Multidimensional NMR in Structural Biology, EMBO workshop, 08/12• Summer School on Medicines 3, PCB-UB, Biocat, Université Paul Sabatier Toulouse III, Oncopôle Toulouse, Université Montréal, workshop, 07/11• Advanced Course on NMR of Jaca, GERMN workshop, 06/11	Barcelona, Spain
02/13 – 06/13	Forschungsinstitut für Molekulare Pharmakologie (FMP-Berlin) PhD Internship – DAAD Fellow (German Government). <i>Solid-state NMR Spectroscopy of Proteins at Prof Oschkinat's laboratory.</i>	Berlin, Germany
04/09 – 08/09	Astbury Centre for Molecular and Structural Biology, University of Leeds MSc Thesis – <i>Atomic Force Microscopy (AFM) for protein biophysics at Dr Radford's laboratory.</i>	Leeds, UK
07/08 – 09/08	Centro Nacional de Biotecnología (CNB) CSIC Undergraduate fellow – JAE Fellowships for research initiation. <i>Virus and cancer group, Dr Rivas' laboratory.</i>	Madrid, Spain
11/07 – 06/08	Universitat de Lleida, Chemistry Department Undergrad. student – <i>Process optimization for biodiesel production at the organic chemistry laboratory.</i>	Lleida, Spain

Publications

- Escobedo, A., Gomes, T., Aragón, E., Martín-Malpartida, P., Ruiz, L., Macias, M.J. (2014) *Structural basis of the activation and degradation mechanisms of the E3 ubiquitin ligase Nedd4L*. **Structure** 22, 1446-1457.
- Aragón, E., Goerner, N., Zaromytidou, A.I., Xi, Q., Escobedo, A., Massagué, J., Macias, M.J. (2011). A Smad action turnover switch operated by WW domain readers of a phosphoserine code. **Genes & Development** 25, 1275-1288

Languages: Catalan (native), Spanish (native), English (fluent), French (fluent)
