



HAL
open science

Investigation of Pin1 interaction with the Cdc25C phosphatase and a novel series of domain-specific ligands

Stephanie Davidson

► **To cite this version:**

Stephanie Davidson. Investigation of Pin1 interaction with the Cdc25C phosphatase and a novel series of domain-specific ligands. Molecular biology. Sorbonne Université, 2023. English. NNT : 2023SORUS442 . tel-04762924

HAL Id: tel-04762924

<https://theses.hal.science/tel-04762924v1>

Submitted on 1 Nov 2024

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

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



THÈSE DE DOCTORAT

École doctorale Chimie-Physique et Analytique de Paris Centre – ED388

Laboratoire des Biomolécules UMR 7203 – Sorbonne Université

Investigation of Pin1 interaction with the Cdc25C phosphatase and a novel series of domain-specific ligands

Présentée par

Stephanie Davidson

Pour obtenir le grade DOCTEUR EN SCIENCES de Sorbonne Université

Directeur de thèse: Dr. Ludovic Carlier

Co-encadrante de thèse: Dr. Maud Larregola

Thèse soutenue le 31 Octobre 2023 devant le jury composé de :

Isabelle LANDRIEU	Rapporteuse
Laure GUILHAUDIS	Rapporteuse
Christina SIZUN	Examinatrice
François XAVIER-THEILLET	Examineur
Thierry BRIGAUD	Examineur
	Président de jury
Ludovic CARLIER	Directeur de thèse
Maud LARREGOLA	Invitée

Acknowledgements

First and foremost, I would like to express my gratitude to the jury members, namely **Dr. Isabelle Landrieu**, **Dr. Laure Guilhaudis**, **Dr. Christina Sizun**, **Dr. François Xavier-Theillet**, and **Pr. Thierry Brigaud**, for their willingness to assess the scientific relevance of my thesis. I am genuinely appreciative of their time and valuable contributions.

The research for this thesis was conducted at the Laboratoire des Biomolécules (LBM) of Sorbonne Université, under the direction of **Pr. Olivier Lequin**, to whom I extend my thanks for having welcomed me into this unit, and for always being friendly and helpful whenever I needed assistance.

I am deeply grateful to my PhD supervisor, **Dr. Ludovic Carlier**, for giving me the opportunity to undertake this challenging project and for his invaluable contributions of time, expertise, and support throughout my research journey. His mentorship and trust in me have been instrumental in my academic growth, and I am truly thankful for his guidance. I would also like to extend my gratitude to **Dr. Maud Larregola**, who, although not my official supervisor, took on a pivotal role as a mentor and guide. Her kindness, support, and the significant time she spent working alongside me, were invaluable to my academic growth.

I would also like to express my appreciation for the contributions of **Dr. Emeric Miclet**, for his generous support, sharing of knowledge, and the time he dedicated. In particular, his expertise in NMR EXSY experiments has been of great value, always making time to engage in discussions and answer my questions whenever I needed it.

I want to recognise the significant contributions of our collaborators of the BioCis laboratory of Cergy-Pontoise, namely, **Pr. Thierry Brigaud**, **Dr. Grégory Chaume** and **Dr. Chiara Zanato**, for their synthesis of all the ligands detailed in this thesis. I also want to extend a special thanks to **Guy Gouarin** and **Lorenzo Meneghelli**, with whom I had the privilege of collaborating and exchanging ideas at Jussieu. Your insights into your respective projects were incredibly helpful in shaping the content of this thesis.

I am sincerely grateful to **Dr. François Xavier-Theillet** for generously providing me with the ERK2 kinase, which played a pivotal role in advancing my research. Additionally, his

willingness to share his expertise in IDP phosphorylation was instrumental in the completion of this work.

I would like to extend my sincere appreciation to the **Infranalytis research federation** for granting me access to their 950 MHz NMR spectrometer at the ICSN, CNRS, Gif-sur-Yvette.

To my fellow members of office 514, and guest star Yadira, I thank you for providing a great working environment that inspired me to come into the lab with enthusiasm every day. Not forgetting my roots, I would like to thank **Rachel, Soha, Ed and Paula** for welcoming me to the lab with open arms (and in Ed's case with some very deep personal questions)! I was so lucky to have you in my first and second year, Rachel as the first person who welcomed me, Soha for teaching me the ways of Pin1, Paula for our gossip sessions and kinder breaks and Ed for being my consistent office buddy. To **Yadira**, who has been a huge support system throughout my entire time in this team, I am so grateful to always have you to come to for advice, gossip and Cuban coffee. I can't forget to thank my Italian queen, **Marta**, since you arrived, my days in the lab (and outside) only got happier. Thank you for bringing such a positive atmosphere, for the lasagne, for the coffee breaks, and for teaching me some very classy Italian phrases ;). Looking over and seeing you making hand gestures at your computer never fails to make me smile. And finally, to **Rapha**, I couldn't have asked for a better person to start my PhD with. You have been with me all the way, my PhD sister, not only supporting me in my work, but also in my personal life. You one of the kindest, most gracious and genuine people I have ever met, and I can't wait to see you get the PhD you so deserve.

I wish to acknowledge the support and encouragement of **Dr. Patrick Fuchs** for his kindness and willingness to engage in discussions, which provided a supportive atmosphere throughout my thesis. To all the other **members of the LBM** who have been incredibly kind and supportive, I, of course, cannot name you all individually, but I extend my heartfelt thanks to each and every one of you.

I would also like to thank **Max**, who I lived with for the majority of my time in Paris, thanks for introducing me to some great TV shows, cheesy gnocchi, and a reborn love for the children's toy section of supermarkets. Also, for showing an interest in my PhD to the point where we spent two full evenings of me describing my project to you. Respect for that, pop pop.

To **Lowie**, who has been the most incredible support system, especially in the final months of this PhD. I would like to thank you for the sacrifices you made, for cooking, cleaning, and welcoming me home after long days of writing. Thank you for sitting near me every evening while I worked, I can't imagine how much harder this would have been to do alone. Your constant love and support have been my anchor through the toughest times, and I am so lucky to have you in my life.

To my **family**, who believed in me and supported my dreams, even when that meant that I would be so far away from them, I express my greatest thanks. For my Uncle **Grahame**, who inspired me to a profession in science, always interested in discussing the complexities of my scientific projects. For my sister **Sarah** your relentless determination was an inspiration throughout this PhD journey, which felt like a marathon of its own! For my **Dad**, who has always been my biggest fan, your kindness, selflessness, and hard work have been a constant source of motivation. For my **Mum**, who came with me to every University open day, was the first person I would call every results day, who brought me cups of tea and snacks when I would spend all day in my room preparing for exams, who was always there to support me even when I was miles away, it all led to this. And finally, to my **Grandad**, your unwavering support, collecting every science award I received in school, still reminds me to this day of how proud you would be. Thank you for believing in me. I couldn't have asked for a more supportive, loving family, and for that I will always be extremely grateful.

I dedicate this thesis to my parents

Table of Contents

List of Figures.....	7
List of Tables.....	11
List of Abbreviations.....	12
English Abstract.....	15
Résumé Français	17
INTRODUCTION.....	19
Chapter 1 Introduction to the Cell Cycle.....	20
Chapter 2 The Cdc25C Phosphatase.....	26
2.1 The CDC25 family.....	26
2.2 The structure and function of Cdc25C.....	31
2.3 Differential phosphorylation status of Cdc25C throughout the cell cycle.....	36
2.4 Conclusion	39
Chapter 3 The Pin1 Isomerase.....	41
3.1 Peptidyl-prolyl isomerases (PPIases).....	41
3.2 The diverse roles of Pin1	43
3.3 Structure and catalytic mechanism	46
3.4 Substrate binding models.....	51
3.5 Interdomain allostery	54
3.6 The interaction between Pin1 and Cdc25C.....	63
Chapter 4 Pin1 as a Therapeutic Target.....	68
4.1 Dysregulation of Pin1 in disease.....	68
4.1.1. The involvement of Pin1 in Alzheimer’s disease	68
4.1.2. The overexpression of Pin1 in cancer	71
4.2 Pin1 inhibitors.....	73
4.2.1. Therapeutic application of Pin1 inhibitors.....	74
4.2.2. Classes of Pin1 inhibitors.....	74
4.3 The incorporation of fluorine into pharmaceutical compounds.....	80
Chapter 5 General Aims of the Thesis	83
5.1 Characterisation of the binding and activity of novel peptide ligands of Pin1	84
5.2 Structural characterisation of the interaction between Pin1 and the IDR of Cdc25C	88
RESULTS & DISCUSSION.....	91
Chapter 6 Production of Samples for NMR Analysis.....	92
6.1 Expression and purification of recombinant Pin1	92
6.1.1. Expression of unlabelled Pin1	92
6.1.2. Expression of ¹⁵ N-labelled Pin1	93
6.1.3. Purification of FL Pin1 Ni-NTA chromatography and gel filtration	93

6.1.4. Optimisation of the protocol	95
6.2 Expression and purification of the IDR of Cdc25C and its mutants.....	98
6.2.1. Expression of Cdc25C ⁴⁴⁻²²⁸ in LB media.....	99
6.2.2. Expression of Cdc25C ⁴⁴⁻²²⁸ in M9 media	100
6.2.3. Purification of WT Cdc25C ⁴⁴⁻²²⁸	101
6.2.4. Purification of phosphomimetic mutant of Cdc25C ⁴⁴⁻²²⁸	105
6.3 PreScission enzyme	106
6.3.1. Expression of PreScission enzyme in LB media.....	106
6.3.2. Purification of PreScission enzyme	107
Chapter 7 Characterising the binding and activity of novel peptide ligands of Pin1	108
7.1 The binding of Pin1 to the 2nd generation of peptide ligands.....	109
7.1.1. Quantifying ligand affinity: NMR-based K _D determination	109
7.1.2. Fluorinated series of Pin1 ligands	117
7.1.3. PROTAC series of Pin1 ligands.....	128
7.2 Optimising a Pin1 activity assay.....	141
7.2.1. Protease-free UV assay	142
7.2.2. NMR EXchange Spectroscopy (EXSY) assay.....	147
7.3 Discussion and perspectives	158
Chapter 8 Structural characterisation of the interaction between Pin1 and the IDR of Cdc25C.....	165
8.1 Overview of structural studies of IDPs using NMR spectroscopy	165
8.2 NMR investigation of unphosphorylated WT Cdc25C ⁴⁴⁻²²⁸	168
8.2.1. Backbone assignment of unphosphorylated WT Cdc25C ⁴⁴⁻²²⁸	168
8.2.2. Structure propensity and dynamics of unphosphorylated WT Cdc25C ⁴⁴⁻²²⁸	176
8.2.3. The interaction between Pin1 and unphosphorylated WT Cdc25C ⁴⁴⁻²²⁸	180
8.3 NMR investigation of the Cdc25C ⁴⁴⁻²²⁸ phosphomimetic mutant	183
8.3.1. The use of phosphomimetics to study phospho-proteins	183
8.3.2. NMR characterisation of pM-Cdc25C ⁴⁴⁻²²⁸	185
8.3.3. The interaction between Pin1 and pM-Cdc25C ⁴⁴⁻²²⁸	187
8.4 Phosphorylation of Cdc25C ⁴⁴⁻²²⁸	193
8.4.1. Challenges with multi-site-specific phosphorylation.....	193
8.4.2. Selection of the kinase for in vitro phosphorylation of Cdc25C ⁴⁴⁻²²⁸	196
8.4.3. Monitoring in vitro phosphorylation of Cdc25C ⁴⁴⁻²²⁸ by SDS-PAGE.....	197
8.4.4. Monitoring in vitro phosphorylation of Cdc25C ⁴⁴⁻²²⁸ by NMR.....	199
8.4.5. Interaction between Pin1 and phosphorylated Cdc25C ⁴⁴⁻²²⁸	207
8.5 Discussion and perspectives	214
MATERIALS & METHODS.....	217
Chapter 9 Materials and Methods.....	218
9.1 Reagents and Recipes	218
9.1.1. Bacterial Growth Media.....	218

9.1.2. Buffers and Solutions	219
9.2 General Techniques	223
9.2.1. UV/Vis Absorbance Spectrophotometry.....	223
9.2.2. Centrifugation	223
9.2.3. SDS-PAGE.....	224
9.3 Protein Expression	224
9.3.1. Plasmid Design and Preparation	225
9.3.2. Bacterial Transformation	225
9.3.3. Bacterial Protein Expression	225
9.4 Protein Purification	226
9.4.1. Isolating the CFE.....	226
9.4.2. Purification of Pin1	226
9.4.3. Purification of Cdc25C ⁴⁴⁻²²⁸	227
9.4.4. Purification of PreScission Enzyme.....	228
9.5 Mass Spectrometry.....	228
9.6 NMR Spectroscopy	229
9.6.1. Instruments Used for data Acquisition.....	229
9.6.2. 1D ¹ H Spectroscopy	229
9.6.3. 2D ¹ H- ¹ H Spectroscopy.....	230
9.6.4. 2D ¹ H- ¹³ C Spectroscopy.....	230
9.6.5. 2D ¹ H- ¹⁵ N Spectroscopy	231
9.6.6. NMR titrations	232
9.6.7. ¹⁵ N-(¹ H) Heteronuclear NOE Spectroscopy.....	237
9.6.8. 3D backbone assignment experiments.....	237
9.6.9. EXchange SpectroscopY (EXSY).....	238
9.7 Pin1 UV activity assay.....	239
9.8 Phosphorylation Reactions.....	240
Bibliography.....	244
Annex 1.....	255

List of Figures

1	A schematic representation of the two main stages of the cell cycle	20
2	The molecular mechanisms underlying the G ₂ to M checkpoint	23
3	Phylogenic tree illustrating the CDC25 family members from various species	27
4	Sequence alignment of Cdc25 family	28
5	A comparison of the X-ray crystal structures of the catalytic domains of Cdc25A and Cdc25B	29
6	The organisation of the catalytic site of Cdc25B (A) and general catalytic mechanism of the CDC25 phosphatases (B)	30
7	Sequence of Cdc25C and prediction of its disorder	32
8	Alpha fold model of Cdc25C (A) and its comparison to the catalytic domains of the other CDC25 phosphatases (B)	33
9	Spatio-temporal regulation of Cdc25C at the G ₂ to M transition	35
10	Summary of the key phosphorylation events occurring in the N-terminal regulatory region of Cdc25C	36
11	Sequence alignment of Cdc25C from <i>Homo sapiens</i> and <i>Xenopus laevis</i>	37
12	Schematic representation of the various phospho-isoforms of Cdc25C	39
13	Free energy diagram of prolyl isomerisation and its catalysis by PPIases	42
14	Comparison of the structural and active site organisation of the PPIase domains of FKBP, cyclophilins and parvulins	43
15	The varied mechanisms by which Pin1 regulates its target proteins.	44
16	X-ray crystal structure of Pin1.	46
17	The binding of a phospho-peptide to the WW domain of Pin1	47
18	The substrate binding pocket of the PPIase domain	49
19	The catalytic tetrad of Pin1.	50
20	The proposed non-covalent catalytic mechanism of Pin1 prolyl isomerisation	51
21	Pin1 substrate binding models	53
22	Schematic representation of the two states of Pin1 (A) and the residues involved in transient interdomain contacts (B)	56
23	The proposed allosteric hydrophobic conduit of Pin1	58
24	The structure of the compact and extended states of Pin1	60
25	Schematic representation of the effect of substrate sequence on inter- and intradomain allosteric signalling within Pin1.	62
26	Schematic representation of the proposed PP2A hypothesis of Pin1 regulation of Cdc25C	65
27	A model for Cdc25C regulation by Pin1 based on the activity of CDK1 in vivo in the presence and absence of Pin1	67
28	The protective role of Pin1 against the development of Alzheimer's disease	70
29	The involvement of Pin1 in promoting the 10 hallmarks of cancer	72
30	X-ray crystal structure of the Pin1 PPIase domain complexed with ATRA	76
31	X-ray crystal structure of the PPIase domain of Pin1 complexed with phosphate- (A) and carboxylate-containing (B) benzothiolene-based peptidomimetics.	78
32	Peptide bond isomerisation in pseudo-prolines	82
33	General structure of ligands containing the CF ₃ -pseudoproline group	87
34	Domain-specific ligands synthesised in the first-generation library of fluorinated peptide ligands of Pin1.	87
35	Chromatogram (A) and SDS-PAGE analysis (B) of the 1 st passage of FL Pin1 through the Ni-NTA column.	94

36	Chromatogram (A) and SDS-PAGE analysis (B) of the 2 nd passage of FL Pin1 through the Ni-NTA column.	94
37	Chromatogram (A) and SDS-PAGE analysis (B) of Pin1 purification by SEC	95
38	Characterisation of the first batch of Pin1 samples using SDS-PAGE analysis (A) and a 2D ¹ H- ¹⁵ N HSQC NMR Spectrum	97
39	Flow chart summarising the steps to produce NMR samples of FL Pin1.	98
40	Selection of the Cdc25C ⁴⁴⁻²²⁸ fragment for study (A) and expression vector design (B)	99
41	SDS-PAGE analysis of cultures after induction of GST-Cdc25C ⁴⁴⁻²²⁸ expression at 37°C or 20°C	100
42	Solubility tests of GST-Cdc25C ⁴⁴⁻²²⁸ after lysis and centrifugation of cells	101
43	Flow chart summarising the steps to produce high-purity NMR samples of Cdc25C ⁴⁴⁻²²⁸	102
44	SDS-PAGE analysis of affinity chromatography steps in the purification of Cdc25C ⁴⁴⁻²²⁸	103
45	Chromatogram (A) and SDS-PAGE analysis (B) of the Cdc25C ⁴⁴⁻²²⁸ purification by SEC	104
46	MALDI-TOF analysis of unlabelled Cdc25C ⁴⁴⁻²²⁸	104
47	Chromatogram (A) and SDS-PAGE analysis (B) of the passage of PreScission enzyme through the Ni-NTA column	107
48	Assigned ¹ H- ¹⁵ N-HSQC spectrum of apo Pin1 in Tris buffer (pH 7) recorded at 500 MHz, 298K	110
49	The dependence of 1D (upper) and 2D (lower) NMR spectra on the exchange rate	112
50	Overlay of nine ¹ H- ¹⁵ N HSQC spectra recorded on ¹⁵ N Pin1 in the presence of increasing concentrations of ligand at 298K on a 500 MHz spectrometer equipped with a cryoprobe.	113
51	Experimental CSP curves obtained for two residues of Pin1 by plotting CSP as a function of $[L]_t/[P]_t$ (number of equivalents)	116
52	CSP analysis of Pin1 in the presence of CZ282 or DMSO control (lower)	119
53	The structure of the phosphate group in comparison with the CF ₂ phosphonate group	120
54	Analysis of the interaction of Pin1 with GG156	121
55	Overlay of ¹ H- ¹⁵ N-HSQC spectra of apo Pin1 (black) and Pin1 + 1.5 molar equivalents of GG213 (orange).	122
56	CSPs in FL Pin1 induced by the binding of 1 molar equivalent of GG213.	123
57	Structure of the <i>cis</i> conformer of CZ115 highlighting its 3 chiral centres (red stars)	123
58	Analysis of the interaction of Pin1 with VB66.	125
59	Analysis of the interaction of Pin1 with GG168.	126
60	Analysis of the interaction of Pin1 with GG201	127
61	Schematic representation of targeting a POI for degradation using PROTAC technology	129
62	General structure and synthetic strategy of the Pin1 PROTAC molecules	131
63	Reference ligands from the first-generation series of Pin1 ligands	131
64	Analysis of the interaction of Pin1 with CZ360	133
65	Analysis of the interaction of Pin1 with LME27	135
66	Analysis of the interaction of Pin1 with LME51	136
67	Comparison of the linkage of the Fmoc via a carbamate (A) vs amide (B) bond	137
68	Analysis of the interaction of Pin1 with AGI24	138
69	Structure of complete PROTAC ligands comprised of the Pin1 LME27 ligand (blue), a (PEG) ₃ linker (red) and thalidomide scaffold (green)	139
70	Analysis of the interaction of Pin1 with LME66	140
71	Overlay of ¹ H- ¹⁵ N HSQC spectrum of Pin1 (black) and Pin1 in the presence of 18 μM chymotrypsin (red)	142
72	Sequence of substrates used in the protease-free Pin1 activity assay	143
73	Difference spectrum of the Glu-Substrate (75 μM) calculated from spectra recorded in 0.47 M LiCl/TFE and HEPES pH 7.8 recorded on a UV spectrometer at ambient temperature (A) and using a CD spectrometer at 10 °C (B)	144
74	Difference spectrum of the Ala-Substrate (48 μM) obtained in HEPES pH 7.8 calculated from spectra at 30s and 600s after a solvent jump from LiCl/TFE to HEPES	146

75	Time course of reversible prolyl isomerisation of the Ala-Substrate (48 μ M) was studied after a rapid transition from a solution containing 0.47 M LiCl/TFE to 35 mM Hepes at pH 7.8	146
76	Basic pulse sequence corresponding to the 2D EXSY experiments	148
77	Analysis of data obtained from typical EXSY experiments	148
78	The amide region of the 2D the $^1\text{H}^1\text{H}$ EXSY spectrum of pCdc25C ($t_m = 100$ ms) in the presence of Pin1	150
79	Assignment of the amide region of the 1D ^1H NMR spectrum of pCdc25C in Tris buffer, pH7 and in the presence of 60 μ M Pin1	150
80	Time dependence of the isomerisation activity of 60 μ M Pin1 on 2.5 mM pCdc25C	152
81	Concentration dependence of the isomerisation activity of Pin1	153
82	Dependence of the isomerisation activity of Pin1 on the concentration of CZ115	155
83	Dependence of the isomerisation activity of Pin1 on the concentration of CZ40	157
84	A potential two-step binding model to explain the comparable inhibitory effect of CZ115 and CZ40	163
85	^1H - ^{15}N -HSQC folded spectrum of apo Cdc25C ⁴⁴⁻²²⁸ in Bis-Tris buffer (pH 7) recorded at pH 6.5, 298K using a 500 MHz spectrometer	169
86	NMR assignment of ^{13}CO nuclei of residues T83-L87 of Cdc25C ⁴⁴⁻²²⁸	170
87	NMR assignment of $^{13}\text{C}\alpha$ nuclei of residues T83-L87 of Cdc25C ⁴⁴⁻²²⁸	171
88	NMR assignment of $^{13}\text{C}\beta$ nuclei of residues T83-L87 of Cdc25C ⁴⁴⁻²²⁸	172
89	NMR assignment of ^{15}N amide nuclei of residues T83-L87 of Cdc25C ⁴⁴⁻²²⁸	173
90	Assigned of ^1H - ^{15}N BEST-TROSY spectrum of Cdc25C ⁴⁴⁻²²⁸ at pH 6.5, 283 K recorded on a 950 MHz spectrometer	174
91	Overlay of ^1H - ^{15}N BEST-TROSY spectra of WT Cdc25C ⁴⁴⁻²²⁸ recorded throughout a temperature titration at pH 6.5, recorded using a 950 MHz spectrometer equipped with cryoprobe	175
92	Overlay of ^1H - ^{15}N SO-FAST spectra of WT Cdc25C ⁴⁴⁻²²⁸ recorded at titration at pH 6.5 (red) and pH 7 (blue) using a 500 MHz spectrometer equipped with cryoprobe	176
93	Secondary chemical shift ($\Delta\text{C}\alpha\text{-C}\beta$) predictions (upper panel), SSP scores (middle panel) and (^1H)- ^{15}N NOE values measured at 500 MHz for WT Cdc25C ⁴⁴⁻²²⁸	179
94	Overlay of ^1H - ^{15}N HSQC spectra of apo-Pin1 (blue) and Pin1 in the presence of 2 molar equivalents of WT Cdc25C ⁴⁴⁻²² (magenta)	180
95	Analysis of amide CSPs observed in Pin1 ^1H - ^{15}N HSQC spectra upon binding to 2 molar equivalents of Cdc25C ⁴⁴⁻²²⁸	181
96	Analysis of the relative signal intensity of amide cross-peaks in ^1H - ^{15}N HSQC spectra recorded on apo-Pin1 and in the presence of 1 (pale blue) or 2 (dark blue) molar equivalents of Cdc25C ⁴⁴⁻²²⁸	182
97	Comparison of the structures of phosphoserine, aspartate and glutamate	184
98	Overlay of ^1H - ^{15}N BEST-TROSY folded spectra of WT (black) and pM-Cdc25C ⁴⁴⁻²²⁸ (red)	185
99	Complete backbone amide assignment of Cdc25C ⁴⁴⁻²²⁸ -pM at pH 6.5, 283 K	186
100	Overlay of ^1H - ^{15}N HSQC spectra of apo-Pin1 (blue) and Pin1 in the presence of 2 molar equivalents of WT (magenta) or phosphomimetic (green) Cdc25C ⁴⁴⁻²²⁸	187
101	Analysis of amide CSPs observed in Pin1 ^1H - ^{15}N HSQC spectra upon binding to 2 molar equivalents of WT (dark green) or phosphomimetic (light green) Cdc25C ⁴⁴⁻²²⁸	188
102	Analysis of the relative signal intensity of amide cross-peaks in ^1H - ^{15}N HSQC spectra recorded on apo-Pin1 and in the presence of 1-2 molar equivalents of WT (upper) or phosphomimetic (lower) Cdc25C ⁴⁴⁻²²⁸	189
103	Overlay of ^1H - ^{15}N BEST-TROSY spectra of pM-Cdc25C ⁴⁴ throughout a titration with up to 4 molar equivalents of unlabelled Pin1	191
104	Binding-induced CSPs of amide cross-peaks in ^1H - ^{15}N BEST-TROSY spectra of the Cdc25C ⁴⁴⁻²²⁸ phosphomimetic upon binding to 4 molar equivalents of Pin1	191
105	SDS-PAGE analysis of phosphorylation reactions of 25 μ M Cdc25C ⁴⁴⁻²²⁸ using 250, 125 or 25 nM ERK2	198
106	SDS-PAGE analysis of phosphorylation reactions of 25 μ M Cdc25C ⁴⁴⁻²²⁸ using 250 nM ERK2	198
107	Overlay of 1D ^1H (A) and 2D ^1H - ^{15}N SOFAST-HMQC (B) spectra recorded at $t=0\text{h}$ and $t=6\text{h}$ after incubation of ^{15}N -Cdc25C ⁴⁴⁻²²⁸ with ERK2 kinase (at a ratio of 1:100 protein:kinase)	200
108	Assigned ^1H - ^{15}N -HSQC spectrum of tri-phosphorylated Cdc25C ⁴⁴⁻²²⁸ , recorded at pH 6.5, 298 K, using a 500MHz spectrometer equipped with cryoprobe	201
109	Minor phosphorylated population of Cdc25C ⁴⁴⁻²²⁸	202

110	Monitoring phosphorylation of Thr48, Thr130 and Ser168 of Cdc25C ⁴⁴⁻²²⁸ by ERK2 using ¹ H- ¹⁵ N SOFAST-HMQC experiments	204
111	Phosphorylation build-up curves and initial rates calculated for each phosphorylation event	205
112	Overlay of ¹ H- ¹⁵ N BEST-TROSY spectra of Pin1 recorded throughout a titration with up to 4 molar equivalents of triphospho-Cdc25C ⁴⁴⁻²²⁸ using a 950MHz spectrometer equipped with cryoprobe	208
113	Analysis of CSPs and signal loss (I/I _{max}) observed for Pin1 amide resonances throughout titration with increasing concentrations of triphospho-Cdc25C ⁴⁴⁻²²⁸ from 0.1 (A), 0.2 (B), 1.0 (C) and 4.0 (D) equivalents. .	209
114	Overlay of ¹ H- ¹⁵ N BEST-TROSY spectra of triphospho-Cdc25C ⁴⁴⁻²²⁸ recorded throughout a titration with up to 4 molar equivalents of Pin1.	211
115	Analysis of CSPs and signal loss (I/I _{max}) observed for triphospho-Cdc25C ⁴⁴⁻²²⁸ resonances throughout titration with increasing concentrations of Pin1 from 0.1 (A), 0.2 (B), 0.3 (C), 0.5 (D), 1.0 (E) and 4.0 equivalents (F)	213
116	Scheme of the process of performing a cross-titration	235

List of Tables

1	Table summarising the key CDK-Cyclin complexes formed during different stages of the cell cycle.	21
2	The structure, properties and application of selected Pin1 inhibitors.	79
3	Physiochemical properties of fluorine	81
4	The first-generation library of Pin1 peptide ligands.	86
5	Yields (mg of purified protein per litre of LB medium) of unlabelled Pin1 acquired throughout this thesis. . .	96
6	Yields (mg of purified protein per litre of M9 medium) of ¹⁵ N-labelled Pin1 acquired throughout this thesis. .	96
7	Comparison of protein yields (mg of purified protein per L of growth media) obtained after purification of Cdc25C ⁴⁴⁻²²⁸ following induction at 37°C or 20°C.	100
8	Yields of unlabelled, single- and double-labelled Cdc25C ⁴⁴⁻²²⁸ acquired throughout this thesis.	105
9	Yields of unlabelled, single- and double-labelled pM-Cdc25C ⁴⁴⁻²²⁸ acquired throughout this thesis.	106
10	The K _D values for the second-generation library of ligands calculated for each domain of Pin1	116
11	Comparison of K _D values of CZ115 and its chirally pure mimic VB66 for Pin1.	125
12	Comparison of K _D values of GG168 and GG201 for Pin1.	128
13	Comparison of K _D values of CZ55 and CZ360 for Pin1.	133
14	Comparison of K _D values of CZ40 and LME27 for Pin1.	135
15	Comparison of K _D values of CZ55, LME51 and LME27 for Pin1.	136
16	Comparison of K _D values of LME51 and AGI24 for Pin1.	138
17	Comparison of K _D values of CZ115 and CZ40 for Pin1	156
18	Dissociation constants calculated for the pCdc25C peptide for the WW domain (isolated or within FL Pin1) in different reports.	158
19	Analysis of protein loss throughout the phosphorylation process of several protein batches.	206
20	Acquisition parameters for 1D ¹ H NMR experiments	229
21	Acquisition parameters for 2D ¹ H NMR experiments	230
22	Acquisition parameters for 2D ¹ H- ¹³ C NMR experiments	231
23	Acquisition parameters for 2D ¹ H- ¹⁵ N NMR experiments.	232
24	Volumes of tube 1 and 2 that are mixed to reach each desired equivalent of Cdc25C.	234
25	Acquisition parameters for ¹ H-(¹⁵ N)-NOE NMR experiments	237
26	Acquisition parameters for 3D assignment NMR experiments	238
27	Acquisition parameters for EXSY NMR experiments	239

List of Abbreviations

- aa** = amino acid
A₂₈₀ = Absorption of UV light at a wavelength of 280 nm
AD = Alzheimer's disease
APP = Amyloid precursor protein
APS = Ammonium Persulphate
BC = Breast cancer
CDC25 = Cell division cycle 25
CDK = Cyclin-dependent kinase
Cha = Cyclohexane
CF₂PO₄ = Difluoro-phosphonate
CSP = Chemical shift perturbation
DDR = DNA damage response
DEER = Double Electron-Electron Resonance
DMSO = Dimethyl sulfoxide
DSP = Dual specificity phosphatase
eNOE = exact Nuclear Overhauser Effect
ER α = Estrogen receptor α
ERK = Extracellular signalling-related kinases
EXSY = EXchange Spectroscopy
FKBP = FK506-binding proteins
FL = Full-length
GSK- β 3 = Glycogen synthase kinase 3
ID = Interdomain
KO = Knock-out
M-phase = Mitotic phase
Nal = Naphtylalanine
NB = No binding
NMR = Nuclear magnetic resonance
NFT = Neurofibrillary tangle
pM = Phosphomimetic
POI = Protein of interest
PPIase = Peptidyl-prolyl isomerase
PRE = Paramagnetic relaxation enhancement
PROTAC = PROteolysis-TArgeting Chimeras

PTM = Post-translational modification
PTP = Protein tyrosine phosphatases
PTPA = Phosphatase 2A phosphatase activators
RDC = Residual Dipolar Coupling
SAC = Spindle assembly checkpoint
SFC = Skp1Cdc53/Cullin-F box protein
S/N = Signal-to-noise
TEMED = Tetramethylethylenediamine
TS = Transition state
TSA = Transition state analogue
WT = Wild-type

Nomenclature of natural amino acids

Amino acid	3 letter code	1 letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

English Abstract

Pin1 is a member of the PPIase family of enzymes which catalyse the *cis/trans* isomerisation of proline peptide bonds, inducing a conformational change in their target proteins. Pin1 is a regulator of key biological processes, including the cell cycle, and its overexpression has been linked to numerous cancers, highlighting its potential as a therapeutic target. Pin1 specifically recognises and isomerises pS/T-P motifs within substrates. The enzyme is small (18.2 kDa) and organised into two domains, a ligand-binding WW domain and a catalytic PPIase domain. Although both domains bind to the same phosphorylated motifs within substrates, the PPIase domain displays preferential binding to motifs in the *cis* conformation, whilst the WW domain almost exclusively binds to the *trans* isomer. The mechanism of Pin1 binding and the relationship between the two domains is not fully understood. In particular, it is unclear how Pin1 recognises and interacts with protein partners containing multiple pS/T-P motifs. Despite many attempts to develop Pin1 inhibitors as anticancer drugs, existing compounds have exhibited limited potency, specificity, and bioavailability.

In order to shed more light on the binding mechanism of Pin1, two projects were undertaken. Firstly, we analysed the binding of a novel library of Pin1 peptide ligands, inspired by a first-generation of compounds produced previously by the lab, for potential application as anti-cancer drugs. Notably, the second-generation ligands tested strategies to enhance potency, facilitate synthesis and increase bioavailability. In addition, strategies for Pin1 abolition were tested: classic catalytic inhibition by competitive inhibitors, and the use of PROTAC technology. The PROTAC strategy involves the design of Pin1 ligand/ubiquitin ligase chimeras which target Pin1 for degradation in cancer cells. The binding of the second-generation library to Pin1 was assessed by observing CSPs in NMR titrations, facilitating the extraction of K_D values at the residue level. We have identified a lead compound of the PROTAC series which displays specific binding to the WW domain of Pin1 with a K_D within the micromolar range. In addition, we have optimised a Pin1 activity assay based on NMR EXSY experiments to compare the inhibitory activity of ligands of the first- and second-generation libraries. Significantly, we demonstrated that both WW- and PPIase-domain

specific ligands display comparable inhibitory activity. This finding suggests that occlusion of the WW domain substrate binding pocket is sufficient to inhibit isomerisation of a Pin1 substrate by the PPIase domain.

There is a lack of structural studies focusing on Pin1's interactions with full-length protein partners. Therefore, another major part of this work involved structurally characterising the interaction between Pin1 and the Cdc25C phosphatase, a critical regulator of the cell cycle that acts as a gatekeeper to mitosis. Cdc25C consists of a C-terminal catalytic domain and an N-terminal IDR, which is host to six potential Pin1 recognition motifs. Despite being a known Pin1 binding partner, the details of this interaction remain largely unexplored. Here, we used NMR to study the interaction between Pin1 and the IDR of Cdc25C. We present the first structural characterisation of the IDR of Cdc25C, experimentally confirming its intrinsically disordered nature and reporting a full NMR backbone assignment. We confirmed that Pin1 displayed weak interaction with both the unphosphorylated IDR and a phosphomimetic mutant of the Cdc25C IDR. Finally, the Cdc25C IDR was phosphorylated *in vitro* to produce a tri-phosphorylated form. Phosphorylation significantly increased affinity for Pin1, revealing that the WW domain acts as the principal binding module for the IDR of Cdc25C. In addition, it appears that Pin1 displays no strong binding preference between the three phosphorylated motifs.

Résumé Français

L'enzyme Pin1, membre de la famille des PPIases, joue un rôle clé en catalysant l'isomérisation *cis/trans* des liaisons peptidiques de la proline, modifiant la conformation des protéines cibles. Pin1 est un régulateur clé de divers processus biologiques, y compris le cycle cellulaire, et sa surexpression a été associée à de nombreux types de cancer, soulignant son potentiel en tant que cible thérapeutique. Pin1 diffère des autres PPIases en ce qu'elle reconnaît et isomérisse spécifiquement les motifs pS/T-P des substrats. Cette petite enzyme (18,2 kDa) se compose de deux domaines, un domaine WW de liaison au ligand et un domaine PPIase catalytique. Les deux domaines se lient aux mêmes motifs phosphorylés dans les substrats, mais le domaine PPIase se lie préférentiellement aux motifs dans la conformation *cis* et le domaine WW se lie presque exclusivement à l'isomère *trans*. Cependant, le mécanisme de liaison et la relation entre ces deux domaines demeurent incompris. Malgré de nombreuses tentatives pour développer des inhibiteurs de Pin1 contre le cancer, les composés existants ont montré des limites en termes de puissance, de spécificité et de biodisponibilité.

Nous avons entrepris deux projets pour comprendre le mécanisme de liaison de Pin1. Tout d'abord, nous avons analysé la liaison d'une nouvelle bibliothèque de ligands peptidiques de Pin1, inspirée d'une première génération de composés produits précédemment par le laboratoire, en vue d'une application future potentielle pour candidats anticancéreux. Les nouveaux ligands ont été conçus pour améliorer leur puissance, faciliter leur synthèse et augmenter leur biodisponibilité. En outre, des stratégies d'abolition de Pin1 ont été testées : l'inhibition catalytique classique par inhibiteurs compétitifs et l'utilisation de la technologie PROTAC. La stratégie PROTAC implique la conception de chimères ligand/ubiquitine ligase de Pin1, ciblant Pin1 pour la dégradation dans les cellules. Les nouveaux ligands ont été testés avec des titrages RMN pour extraire les valeurs K_D au niveau du résidu. Nous avons identifié un composé principal de la série PROTAC qui se lie au domaine WW de Pin1 avec une affinité de liaison de l'ordre du micromolaire. Nous avons aussi optimisé un test d'activité Pin1 basé sur des expériences NMR EXSY afin de comparer l'activité inhibitrice de nos ligands. Nous avons démontré que les ligands spécifiques des domaines WW et PPIase présentent une activité

inhibitrice comparable. Cela suggère que l'occlusion du domaine WW est suffisante pour inhiber l'isomérisation par le domaine PPIase.

Une partie essentielle de ce travail a été de caractériser structurellement l'interaction de Pin1 avec la phosphatase Cdc25C, un régulateur clé du cycle cellulaire, notamment de la mitose. Cdc25C comprend un domaine catalytique en C-terminal et une région N-terminale, intrinsèquement désordonnée (IDR) contenant six motifs potentiels de liaison à Pin1. Bien qu'il s'agisse d'un partenaire connu de Pin1, les détails, au niveau structurel, de cette interaction restent largement inexplorés. Ici, nous avons utilisé RMN pour étudier l'interaction entre Pin1 et l'IDR de Cdc25C. Nous présentons la première caractérisation structurale de l'IDR de Cdc25C, confirmant expérimentalement sa nature intrinsèquement désordonnée et rapportant une attribution complète du squelette par RMN. Nous avons confirmé que Pin1 interagissait faiblement avec l'IDR non phosphorylé et un mutant phosphomimétique de l'IDR de Cdc25C. Enfin, l'IDR de Cdc25C a été phosphorylé *in vitro* pour produire une forme triphosphorylée. La phosphorylation a augmenté l'interaction avec Pin1, révélant que le domaine WW est le principal module de liaison pour l'IDR phosphorylé de Cdc25C. De plus, nous avons révélé que Pin1 n'a pas de préférence de forte liaison entre les trois motifs phosphorylés.



INTRODUCTION

Chapter 1 Introduction to the Cell Cycle

At the heart of cell biology lies the cell cycle, a fundamental process that orchestrates the growth and division of cells to sustain life. Proper progression of the cell cycle relies on the intricate coordination of numerous proteins and pathways, and any disruption to this complex network can result in severe pathological consequences. Two key players in this network are Pin1 and Cdc25C, which perform critical regulatory roles in the cell cycle and will be the focus of this thesis. This chapter will provide a comprehensive overview of the basic process and stages of the cell cycle, offering a foundation for understanding the crucial roles of Pin1 and Cdc25C which will be explored in depth later.

The eukaryotic cell cycle is comprised of two main stages, interphase (including G_1 , S and G_2) during which DNA is replicated in preparation for division, and the mitotic phase (M) whereby the cell physically divides producing two identical daughter cells (see figure 1). Interphase can be further divided into three crucial phases, including the S phase where DNA replication occurs, and two gap phases (G_1 and G_2) which serve as critical control periods to guarantee the cell's readiness to proceed to the subsequent S or M phase (Norbury & Nurse, 1992). During the first stage of the cell cycle, G_1 , cells have the option of exiting the cell cycle before committing to chromosomal replication to enter a quiescent state known as G_0 . G_0 is not formally considered a stage of the cell cycle, but rather a phase where the cell has temporarily stopped dividing, characteristic of the majority of non-proliferating cells in the human body (Vermeulen *et al*, 2003). Alternatively, if cells *do* continue through interphase, genome replication occurs during the S phase resulting in the formation of two identical copies of each chromosome, known

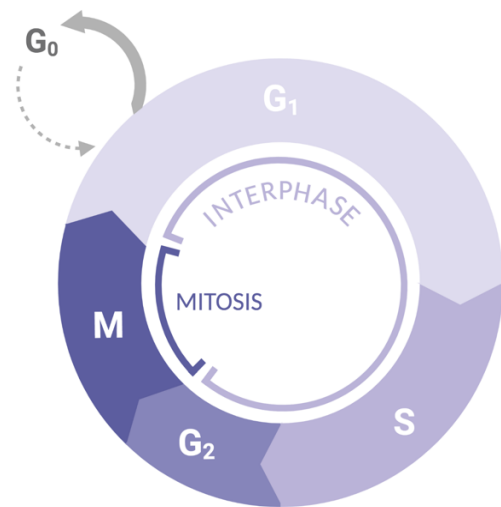


Figure 1: A schematic representation of the two main stages of the cell cycle (interphase and mitosis). Interphase is further subdivided into 3 stages, G_1 (gap stage 1), S (DNA synthesis) and G_2 (gap stage 2). Once interphase is complete, the cell progresses into stage M (mitosis), in which it divides into two identical daughter cells.

as sister chromatids. These sister chromatids will later separate and segregate to opposite poles of the cell during the first 4 stages of mitosis (prophase, metaphase, anaphase and telophase) before the cell eventually divides (cytokinesis) producing two genetically identical daughter progeny (McIntosh, 2016).

Proper progression through each phase of the cell cycle depends on an intricate network of proteins, with the cyclin-dependent kinase (CDK) family of threonine/serine kinases playing a dominant role (Malumbres & Barbacid, 2005). These key cell cycle regulators are activated at particular moments during the cycle and act as molecular timers to direct the cell through each phase via a series of phosphorylation events which provoke downstream pathways. Of the 20 characterised human CDKs (CDK1-20), 7 have been shown to perform direct functions in guiding the progression of the cell cycle, while the remaining family members play important roles in transcription (Malumbres *et al*, 2009). Although constitutively expressed throughout the duration of the cell cycle, their catalytic activity relies on the binding of additional regulatory proteins, known as cyclins (Malumbres, 2014). There are multiple different types of cyclins, each with distinct patterns of expression that rise and fall throughout the cycle (Pines, 1995). Notably, each CDK of the cell cycle displays promiscuous binding to several cyclins, hence, it is the specific combination of different cyclin-CDK complexes at particular moments of the cell cycle that allows for the temporal control of its progression (see table 1) (Malumbres, 2014). This ability to control the progression of the cell cycle is facilitated by the downstream effects that each of these cyclin-CDK complexes exerts on the cellular machinery (Ding *et al*, 2020). For example, the CDK2-cyclin E complex, formed at the G_1 to S transition, phosphorylates and activates several target proteins

Phase	CDK	Cyclin
G_1	CDK4	Cyclin D (1, 2, and 3)
	CDK6	Cyclin D (1, 2, and 3)
$G_1 \rightarrow S$	CDK2	Cyclin E
S	CDK2	Cyclin A
G_2	CDK2	Cyclin A
$G_2 \rightarrow M$	CDK1	Cyclin A
M	CDK1	Cyclin B
All	CDK7	Cyclin H

Table 1: Table summarising the key CDK-Cyclin complexes formed during different stages of the cell cycle. The detailed formation and downstream effects of these complexes has been reviewed in Ding *et al.*, 2020; Malumbres, 2014; Vermeulen *et al.*, 2003.

involved in replicating DNA, such as the replication licensing factor Cdc6. Notably, CDK activity is not exclusively reliant on cyclin binding, with a series of phosphorylation events and binding of inhibitory proteins allowing the fine tuning of CDK activity throughout the cell cycle (Malumbres, 2014). For example, the CDC25 phosphatases function to activate CDK-cyclin complexes through the removal of inhibitory phosphates in a spatio-temporal manner (Boutros *et al*, 2007).

Due to the importance of the cell cycle's seamless and ordered progression, it comes as no surprise that the regulation of this process is not solely dependent on CDK-cyclin complexes. To ensure the faultless execution of the cell cycle, several checkpoints have been established, providing a series of requirements which the cell must satisfy before progression to the next phase is granted (Hartwell & Weinert, 1989). Failure to meet these requirements can result in temporary or permanent cell cycle arrest for the purpose of maintaining genome integrity and proper ploidy in the dividing cells. Amongst the most significant checkpoints, the G₁ restriction point acts as a first checkpoint, marking the commitment of the cell to proliferation only if extracellular conditions are optimal (Pardee, 1974; Pennycook & Barr, 2020). Next, during the S phase and the G₂ to M transition, DNA damage checkpoints are present to ensure error-free replication (Panagopoulos & Altmeyer, 2021). During mitosis, the spindle assembly checkpoint (SAC) acts to ensure that the replicated chromosomes are equally distributed to each daughter cell. Each cell cycle checkpoint is meticulously orchestrated by a complex network of proteins to ensure its precise progression. Consider the G₂ to M transition, where the pivotal event for mitotic entry hinges upon the activation of the CDK1-CyclinB complex, which marks the irreversible commitment to mitosis (see figure 2) (Otto & Sicinski, 2017). This activation necessitates both the formation of a functional complex with cyclin B and the subsequent removal of inhibitory phosphorylations, accompanied by the addition of activating phosphorylations on CDK1. During the G₂ phase of the cell cycle, the MUVB complex forms an association with the FOXM1 transcription factor, thereby facilitating the transcription of genes crucial for the initiation and progression of mitosis, including those encoding cyclin B proteins. Further activation of the CDK1-CyclinB complex is facilitated by the phosphorylation of Thr161 on CDK1 by the CDK7-cyclin H complex, also known as the cyclin-activating kinase (CAK). Crucially, mitotic entry is contingent upon the removal of inhibitory phosphorylations (pThr14 and pTyr15) from CDK1, by the Cdc25 phosphatases, in particular Cdc25C. Notably, these inhibitory phosphorylations are maintained by the Myt1 and Wee1

kinases throughout interphase, but at the G₂ to M transition their activity is superseded by the Cdc25C phosphatase, resulting in their removal and marking the onset of mitosis.

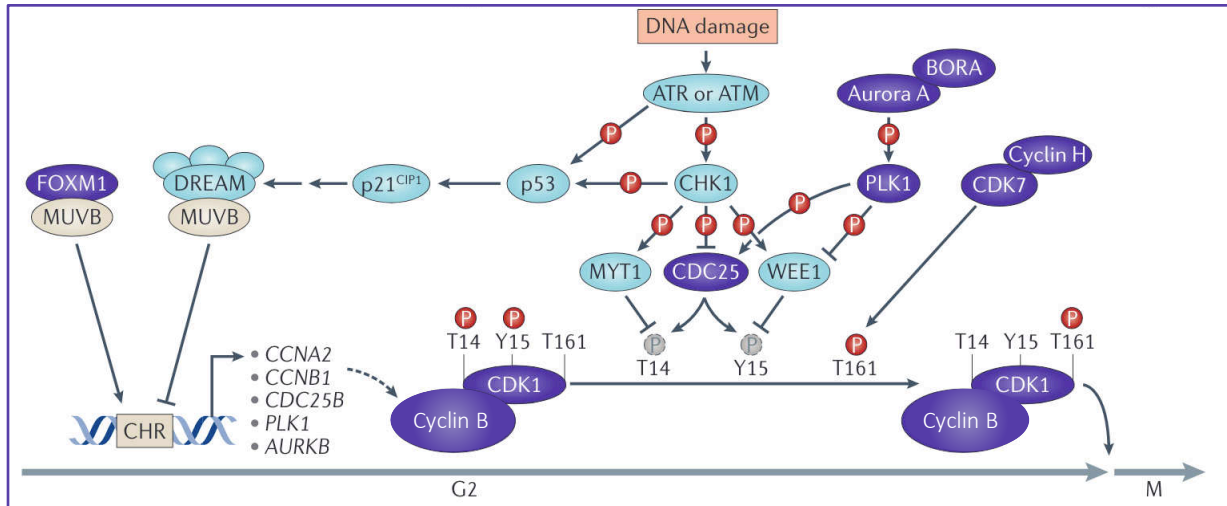


Figure 2: The molecular mechanisms underlying the G₂ to M checkpoint. Activation of CDK1-cyclin B complexes permits entry into mitosis. During G₂ phase, the MUVB complex, together with the FOXM1 transcription factor, activates genes necessary for entering and progressing through mitosis, including cyclin B. Activation of CDK1-cyclin B kinase requires phosphorylation of Thr161 by the CDK7-cyclin H complex (red circle “P”) and removal (dashed circle “P”) of inhibitory Thr14 and Tyr15 phosphorylations on CDK1 by Cdc25C. Thr14 and Tyr15 are phosphorylated by MYT1 and WEE1 kinases. CHK1-dependent G₂ DNA damage checkpoint prevents CDK1 activation. After DNA damage repair, PLK1 is essential for CDK1 reactivation. Purple ovals signify cell cycle progression promoters, while blue ovals present inhibitors. Diagram modified from Otto & Sicinski, 2017.

Notably, the checkpoint networks are heavily intertwined with the DNA damage response (DDR) pathway so that if any break or mutation in DNA is flagged by the DDR, cell cycle progression is halted until the DNA can be repaired. The DDR comprises an extensive network of signalling pathways, primarily being controlled by the ATM-Chk2 and ATR-Chk1 kinases (Petsalaki & Zachos, 2020). The ATM-Chk2 and ATR-Chk1 pathways respond to different types of DNA breaks, with ataxia telangiectasia mutated (ATM) activated at double stranded lesions and the ataxia telangiectasia and Rad3-related (ATR) kinase activated if DNA replication is blocked (for instance due to inhibition of DNA polymerases or UV damage). Catalytically active ATM and ATR in turn phosphorylate and activate their downstream kinases, Chk2 and Chk1 respectively. Once activated, these kinases phosphorylate various

substrates involved in the cell-cycle, often acting to prevent cell cycle progression while DNA damage persists. Some molecular consequences of the DDR can be seen in detail at the G₂ to M transition (see figure 2). Here, the Chk1-dependent G₂ DNA damage checkpoint can repress the transcription of genes crucial for mitotic entry and progression, primarily through the formation of the DREAM multiprotein complex, which acts as a transcriptional repressor (Otto & Sicinski, 2017). Additionally, Chk1's phosphorylation of Myt1 and Wee1 kinases promotes the maintenance of inhibitory phosphorylations, pThr14 and pThr15, on CDK1. Further reinforcing this control, Chk1 directly inhibits the Cdc25 phosphatases. Collectively, these mechanisms effectively prevent CDK1 activation and, consequently, entry into mitosis. The DDR plays additional crucial roles during mitosis, regulating spindle formation and anaphase onset as well as correcting mis-attached kinetochore–microtubules (reviewed in Petsalaki & Zachos, 2020). Overall, the complex cross-talk between the cell cycle and DDR pathways ultimately protects the cell from chromosomal instability, which is strongly associated with tumorigenesis and human disorders.

Of the large network of proteins that have been characterised as significant to ensuring the activation of and advancement through the cell cycle checkpoints, many have also been classified as proto-oncogenes. This is unsurprising considering one of the ten key hallmarks of cancer, uncontrolled cell proliferation, is triggered by cell cycle dysfunction and is provoked by deletions, mutations or differential expression patterns of its key governing proteins (Williams & Stoeber, 2012). The phenotype of sustained proliferative signalling is most often caused by the malfunctioning of proteins controlling the cell cycle checkpoints and the DDR, ultimately compromising the ability of cells to arrest in response to genetic errors (Matthews *et al*, 2022). While some cell cycle checkpoints remain functional in cancer, others, such as the DNA damage checkpoint, are often compromised. Mutations commonly found in cancer affect cell cycle control in response to DNA damage and growth signals, promoting S phase entry and inhibiting cell cycle exit. Despite this, certain checkpoints, like the DNA replication stress checkpoint and the spindle assembly checkpoint (SAC), remain vital in cancer cells and are rarely mutated. Commonly mutated proteins in cancer include ATM, Chk2 and p53 involved in the DNA damage response, with commonly overexpressed proteins including CDKs and CDC25 phosphatases which are key for the progression through DNA damage checkpoints (Boutros *et al*, 2007). The exact causes of CDC25 overexpression in cancer are not fully understood, and it appears to result from a complex interplay of various factors. There is no

evidence to suggest that gene amplification or specific genetic mutations are responsible for CDC25 overexpression in cancer. Instead, CDC25 deregulation can occur at multiple stages, including transcription, translation, and post-translational modifications. It is important to note that, like many singular protein modifications, CDC25 overexpression alone is inadequate to induce carcinogenesis. However, it can contribute to the disease by promoting abnormal cell cycle progression, genomic instability, and cooperation with other oncogenes.

In conclusion, the cell cycle represents a tightly regulated sequence of events that are fundamental for the proper growth and division of eukaryotic cells. It is composed of several stages that are monitored and driven by a complex network of proteins, including CDK-cyclin complexes which act as molecular timers driving the cell through each phase of the cycle. In addition to CDKs and cyclins, checkpoints play a crucial role in ensuring that the cell cycle progresses in an orderly and error-free manner and that any potential DNA damage is detected and repaired. The following sections will delve deeper into the structure and function of two key proteins known to be important in cell cycle regulation and CDK-cyclin activation. Cdc25C, a phosphatase involved in the G₂ to M transition, acts directly to activate CDK1-Cyclin B to control entry into mitosis. The result of this interaction is dependent on Cdc25C's regulation by another protein, Pin1, a unique isomerase regulating a range of cellular processes including the cell cycle. Pin1 acts on a range of key proteins involved at each stage of the cell cycle and is often coined a 'master regulator' of the cell cycle. Notably, due to its role in promoting cell cycle progression, Pin1 is also commonly overexpressed in human cancer and represents a well-characterised therapeutic target. However, the mode of action of Pin1 and its interaction with substrates, including Cdc25C is not fully understood. Further understanding of this interaction will deepen our knowledge of the critical stages of cell cycle regulation and provide more potential targets for anticancer therapies.

Chapter 2 The Cdc25C Phosphatase

As previously introduced, the CDC25 phosphatases play pivotal roles in overseeing the progression of the cell cycle through their regulation of CDK-cyclin activity. Human cells express three distinct isoforms of CDC25 - A, B and C – each acting at different stages of the cell cycle. Cdc25C specifically plays a role in the G₂ to M transition where it is involved in an activation loop with the CDK1-cyclin B complex which drives progression through mitosis. Its function as an important mitotic gateway protein makes Cdc25C subject to tight spatio-temporal control through a series of post-translational modifications and binding events which occur in its regulatory region. This chapter will explore the complexities of Cdc25C, including its structural features, its role in the cell cycle, and the interplay of modifications that influence its regulation.

2.1 The CDC25 family

As members of the larger class of protein tyrosine phosphatases (PTPs), the cell division cycle 25 (CDC25) family act to remove regulatory phosphate groups from their CDK protein targets in the cell cycle (Liu *et al*, 2020). The CDC25 family can be further classified as dual specificity phosphatases (DSPs) due to their ability to recognise and dephosphorylate both phospho-tyrosine and phospho-threonine residues. CDC25 phosphatases are highly conserved among eukaryotic cells, except for plants, with mammalian systems possessing three isoforms: CDC25A, CDC25B, and CDC25C (Boutros *et al*, 2007). Additionally, non-mammalian orthologues of these proteins have also been identified in frogs (*Xenopus laevis*, Cdc25A and Cdc25C) and chickens (*Gallus gallus*, Cdc25A and Cdc25B) further underscoring their evolutionary conservation and functional significance (see figure 3) (Boutros *et al*, 2007). The three human CDC25 homologs are of approximately similar chain lengths of 524, 580 and 473 amino acids (aa) for CDC25A, B and C respectively (Brenner *et al*, 2014). In addition, they all share a similar structural architecture, each being organised into a C-terminal catalytic and an N-terminal regulatory region which is predicted to be intrinsically disordered (Liu *et al*, 2020). Their respective catalytic domains display a high level of conservation, each containing the catalytic HCX₅R motif, characteristic of all PTPs (see figure 4).

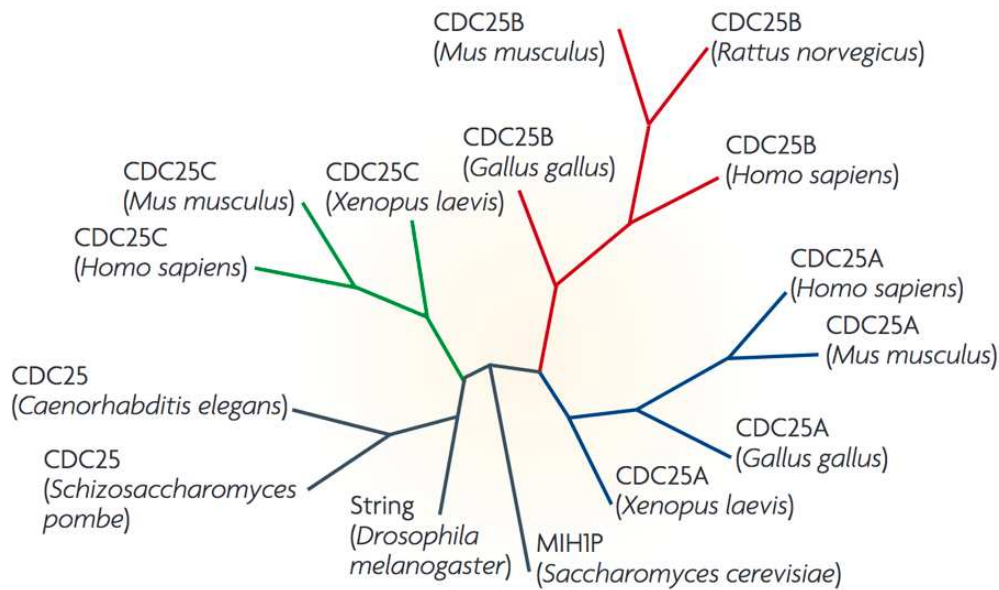


Figure 3: Phylogenetic tree illustrating the CDC25 family members from various species. The tree consists of four major branches representing CDC25 isoforms: CDC25A (depicted in blue), CDC25B (depicted in red), CDC25C (depicted in green), and other variants found in non-mammalian species (depicted in grey). Within the CDC25A-C branches, mammalian species, chickens (*Gallus gallus*), and frogs (*Xenopus laevis*) are represented. The tree provides an overview of the evolutionary relationships among CDC25 family members across different organisms. Diagram taken from Boutros et al., 2007.

The X-ray crystal structures of the catalytic domains of the CDC25A and B isoforms have been resolved and revealed slight differences in their active site organisation and substrate binding properties, though their overall folds are largely conserved (see figure 5) (Fauman *et al*, 1998; Reynolds, 1999). The characteristic fold of the catalytic domain consists of a central five-stranded parallel β -sheet that is flanked by three α -helices below and two α -helices above (Rudolph, 2007). The most significant distinguishing factor concerning the three isoforms lies in their divergent *N*-terminal regulatory regions which vary in both sequence and length. These differences are further accentuated due to the presence of splice variants. The *N*-termini contain multiple sites for post-translational modifications (PTMs), and particularly a range of phosphorylation and ubiquitination events can dynamically affect CDC25 phosphatase activity, binding events, subcellular localisation, as well as the overall abundance of the protein. Therefore, the composition of this regulatory domain and its post-translational regulation underlies the different functions of the CDC25 isomers in the cell cycle.

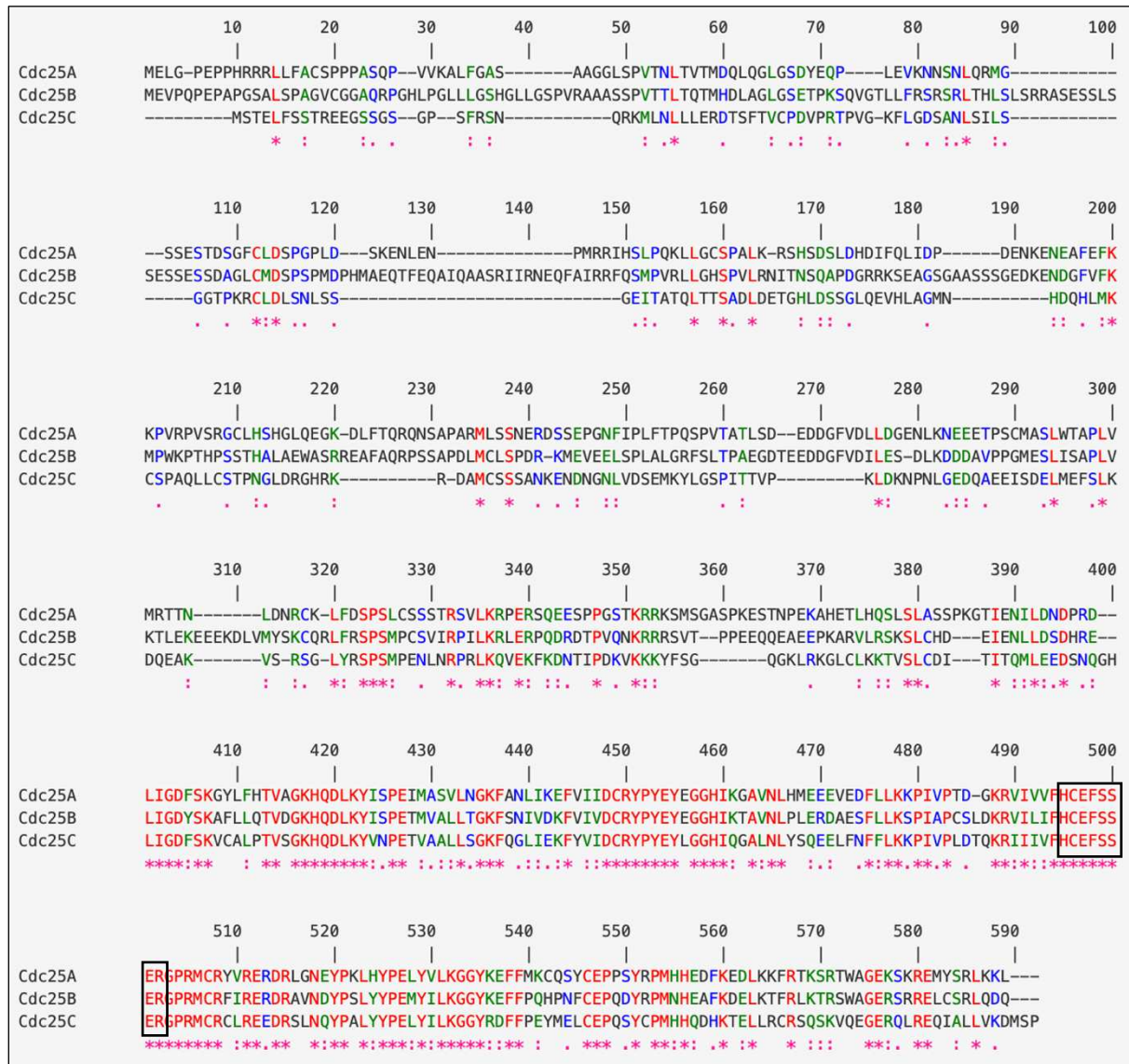


Figure 4: Sequence alignment of Cdc25 family. The *N*-terminal regions show little sequence homology however, the *C*-terminal catalytic domains show an overall high sequence identity between the 3 isomers. The conserved catalytic HCEFSR motif (highlighted by black box) is identical between all 3 isomers (HCEFSR). Residues highlighted in red and indicated by ‘*’ are identical between all 3 isomers, residues highlighted in green and marked with ‘:’ indicate conservation of highly similar properties, and residues displaying partial conservation of properties are highlighted in blue and indicated by ‘.’

A remarkable feature of CDC25 phosphatases is the absence of a well-defined substrate binding site, with the recognition of substrate phosphate groups predominantly governed by the residues of the catalytic site (Rudolph, 2007). The characteristic flat and shallow nature of the catalytic site is in fact well-suited for accommodating substrates containing both pThr and pTyr, aligning with the dual-specificity of Cdc25 enzymes. Instead, substrate recognition

seems to depend on specific hotspot residues (e.g., R488, R492, and Y497 in Cdc25B), which are located approximately 20–30 Å away from the active site.

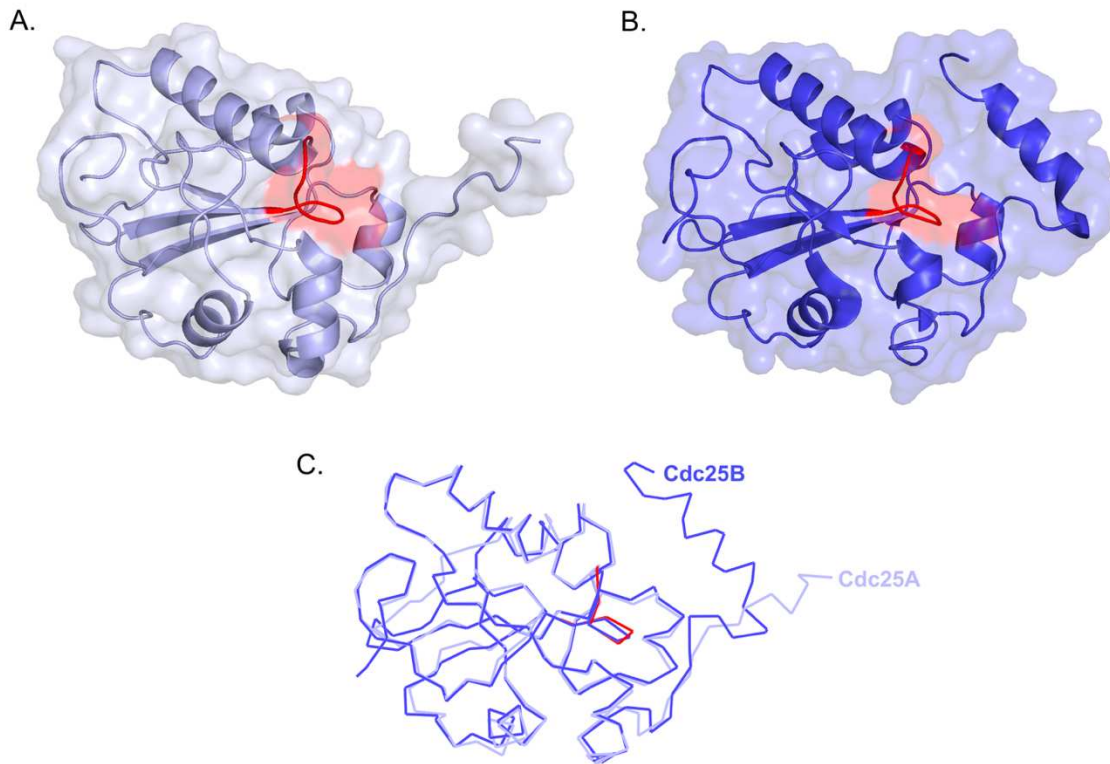


Figure 5: A comparison of the X-ray crystal structures of the catalytic domains of Cdc25A and Cdc25B. The catalytic domains of **A.** Cdc25A (residues 336–523) and **B.** Cdc25B (residues 394–580) are folded into a small α/β domain with a parallel β sheet core. Cdc25B contains an additional α -helix proximal to the catalytic site. **C.** The C α trace of the catalytic domain of Cdc25CA and B is overlaid, demonstrating their near identical fold. The conserved HCX₅R catalytic loop is highlighted in red. Figure created using PyMol (PDB IDs 1C25: Cdc25CA and 1QB0: Cdc25CB)

All three human CDC25 enzymes catalyse the removal of inhibitory phosphate groups at residues Thr14 and Tyr15 of the ATP binding loop in their CDK targets to generate the active kinase (Malumbres & Barbacid, 2005). CDC25 phosphatases act on CDKs through a stepwise mechanism, where the enzyme dissociates after dephosphorylating pThr and must reassociate for the subsequent dephosphorylation of pTyr (Rudolph, 2007). The active site primarily comprises the conserved HCX₅R motif. The dipole moment of the α -helix adjacent to this motif promotes the deprotonation of the catalytic cysteine, priming it for catalysis. The remaining residues within this motif participate in coordinating the substrate's phosphate group (see figure 6a). The backbone amide groups of the five X residues, together with the arginine sidechain

amide group, collectively form hydrogen bonds with the oxygen atoms of the phosphate group. The catalytic mechanism, common among PTPs, relies on an unprotonated thiolate (provided by the catalytic cysteine residue) and a protonated catalytic acid (see figure 6b). The thiolate anion is positioned directly beneath the phosphate group, enabling an in-line attack on the P-O bond in the initial chemical step of the reaction. This leads to the formation of the hydroxyl group on the Thr or Tyr residue of the substrate, as well as a covalently-linked phospho-cysteine intermediate, facilitated by the protonation of the oxyanion by the catalytic acid. The identity of the catalytic acid is not definitively established for the three CDC25 phosphatases.

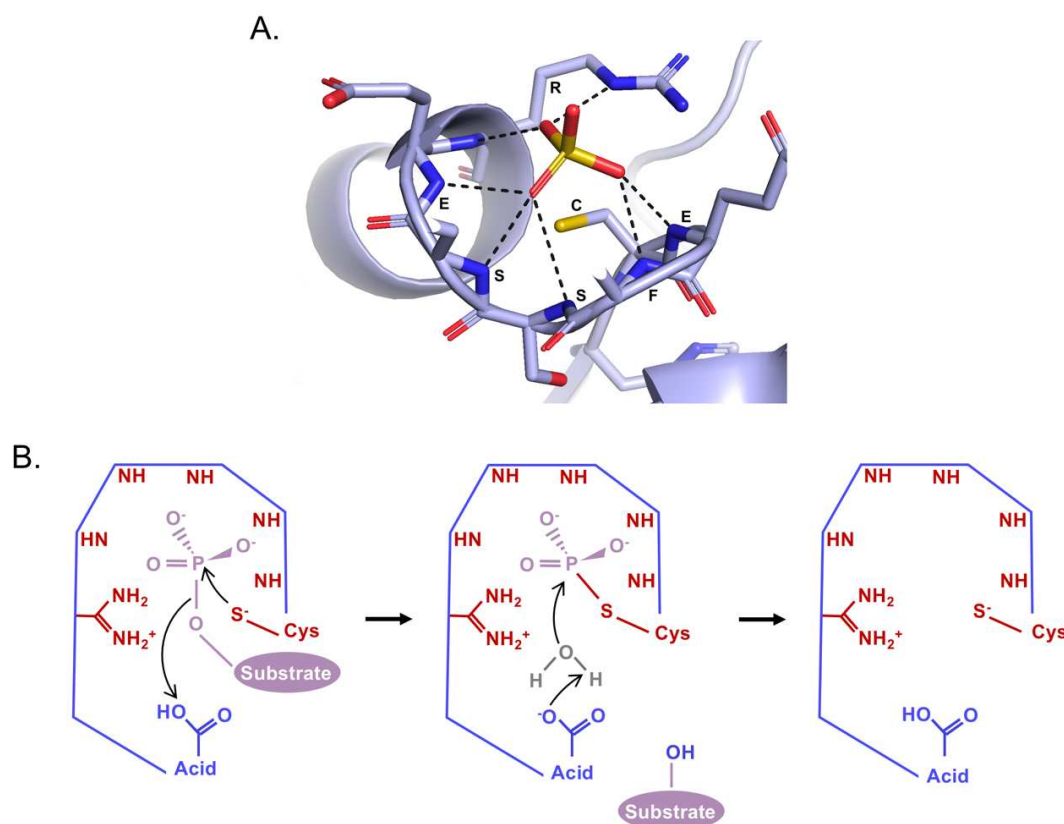


Figure 6: The organisation of the catalytic site of Cdc25B (A) and general catalytic mechanism of the CDC25 phosphatases (B). A. Cdc25B was co-crystallised with SO_4 bound in the active site, indicative of the interaction site of the substrate phosphate group. The backbone amide groups of residues E, F, S, S, E and R, as well as the sidechain amide of the R residue (of the catalytic HCX₅R motif) form hydrogen bonds (black dashed lines) with the oxygen atoms of sulphate/phosphate groups. The catalytic Cys residue lies below the sulphate/phosphate group. Figure made in PyMol (PDB ID: 1QB0) B. Residues of the HCX₅R motif shown in red. The catalytic acid is shown in blue. The substrate groups are represented in purple. The catalytic mechanism involves nucleophilic attack of the P-O bond of the substrate phosphate group by the thiolate anion of the catalytic cysteine. The product is released and passage through a covalently-linked phospho-cysteine intermediate occurs before regeneration of the active site by a water molecule.

While the three isomers seem to display partial functional redundancy, their differential temporal and spatial regulation *in vivo* limits their activity to different points during the cell cycle, where they play distinct roles. Cdc25A is a nuclear protein acting during the progression of both the G₁ to S and G₂ to M transitions of the cell cycle (Hoffmann *et al*, 1994). During late G₁, Cdc25A is involved in an activation loop with CDK2-cyclin E, driving entry into the S phase (Brenner *et al*, 2014). During the S phase, Cdc25A, in parallel with Cdc25B, dephosphorylate and activate CDK2-cyclin A, the CDK-cyclin complex which promotes DNA replication during the S phase. All three isomers appear to play a functional role in the G₂ to M transition of the cell cycle, where the dephosphorylation of the CDK1-cyclin B complex is the rate-limiting step to mitotic entry (Sur & Agrawal, 2016). During late G₂, Cdc25B shuttles from the nucleus to the cytoplasm and is proposed to play a role as the “initiator” of CDK1-Cyclin B activity (Boutros *et al*, 2007). After initial activation by Cdc25A and B, Cdc25C appears to play the dominant role in mitotic initiation by promoting complete CDK1-Cyclin B activity and maintaining it throughout mitosis (de Gooijer *et al*, 2017).

2.2 The structure and function of Cdc25C

Building upon its brief introduction as a cell cycle regulator and key mitotic gatekeeper, the structure and specific function of the smallest CDC25 family member, Cdc25C, will now be explored in more detail. Similar to the other members of the family, the human Cdc25C protein is organised into a folded C-terminal catalytic domain (residues 260-273) and an N-terminal regulatory region (residues 1-259) which is predicted to be disordered by the IUPred3 program (see Figure 7) (Liu *et al*, 2020). While the structure of the catalytic domain of Cdc25C has remained elusive, recent advancements in structural biology, such as AlphaFold, have revolutionised our ability to predict protein structures with remarkable accuracy. AlphaFold utilises a deep learning algorithm which has learned from the PDB database to generate highly reliable structural predictions. In the case of Cdc25C, despite the absence of a resolved crystal structure, its substantial sequence homology (over 60%) with family members Cdc25A and Cdc25B strongly implies the presence of a conserved structural fold. The AlphaFold prediction for Cdc25C's C-terminal catalytic domain is particularly compelling, displaying a high confidence level (see Figure 8a) (Jumper *et al*, 2021; Varadi *et al*, 2022). This structural model closely resembles the crystal structures of Cdc25A and Cdc25B catalytic domains, with one notable difference—an extended α -helix spanning residues Gln438-Met471 (see Figure 8b). This elongated helical structure in Cdc25C, absent in Cdc25A

and shorter in Cdc25B, could potentially serve as a lid-like feature, possibly involved in substrate binding and regulating access to the nearby active site. However, the *N*-terminal region of Cdc25C poses a greater challenge for AlphaFold due to its intrinsically disordered nature. Consequently, the predictions for this region are less confident. Most of the *N*-terminal region is anticipated to lack a defined structure. Yet, AlphaFold does suggest a short α -helix spanning residues Phe21 to Arg35 within this otherwise disordered region. However, it is crucial to highlight that this prediction is associated with a low confidence level. Additionally, AlphaFold predicts another alpha-helix formation between residues Ile269 and Gly280, likely marking the commencement of the catalytic domain. This prediction, while offering insight into the connectivity between the intrinsically disordered region and the structured domain, also carries a relatively low confidence level.

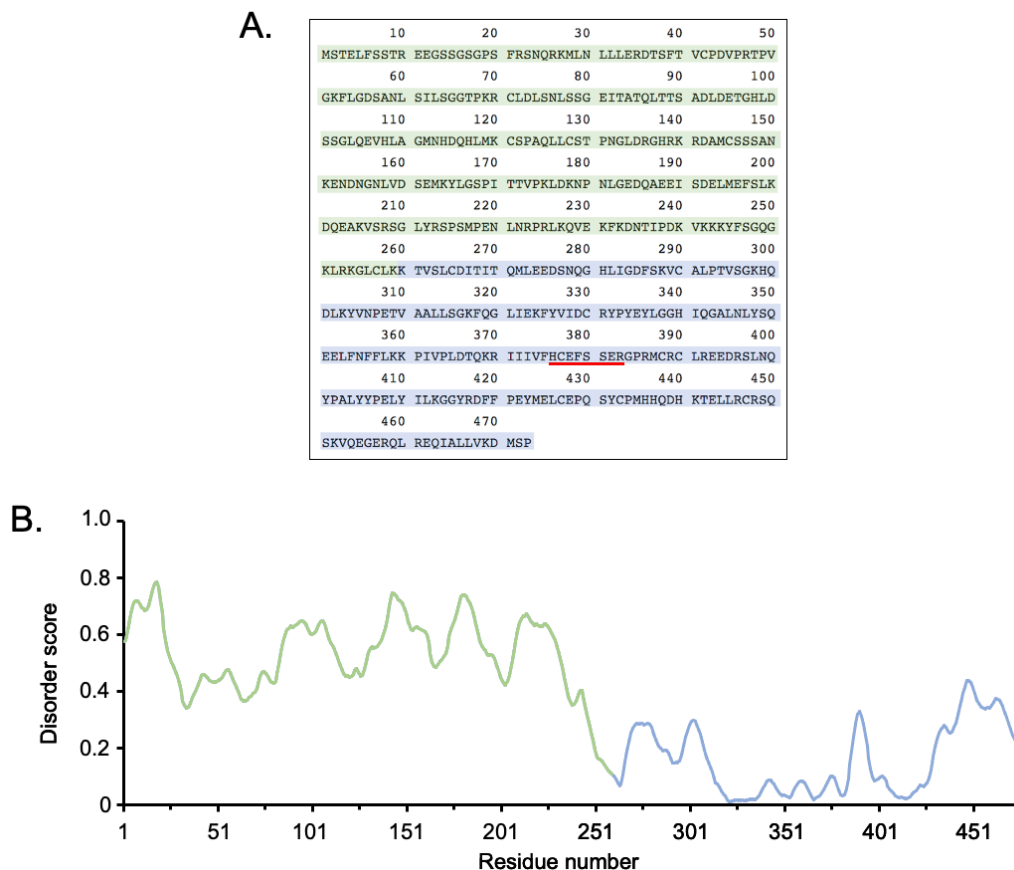


Figure 7: Sequence of Cdc25C and prediction of its disorder. The *N*-terminal regulatory region is highlighted in green and the *C*-terminal catalytic domain in blue. **A.** The characteristic catalytic HCX₅R motif is underlined in red. **B.** The IUPred graph displays the disorder scores for each residue along the Cdc25C sequence. Disorder scores range from 0 to 1, where higher values indicate a higher likelihood of intrinsic disorder, while lower values suggest a more ordered structure.

Overall, the catalytic domain is predicted to display the same topology as the other CDC25 phosphatases, comprising a 5-stranded parallel beta-sheet core sandwiched by two alpha-helices on top and three alpha-helices below. The *N*-terminal region of Cdc25C, however, is unique among its family members, sharing less than 40% sequence identity, and is further distinguished by the presence of at least five splice variants (Cdc25C1-5) (Bonnet *et al*, 2008; Sur & Agrawal, 2016). This region is rich in Ser and Thr residues which can undergo phosphorylation or ubiquitination events to exert regulatory effects. The plethora of regulatory modifications occurring in this region control the activity and subcellular location of Cdc25C during the different phases of the cell cycle (summarised in figure 9).

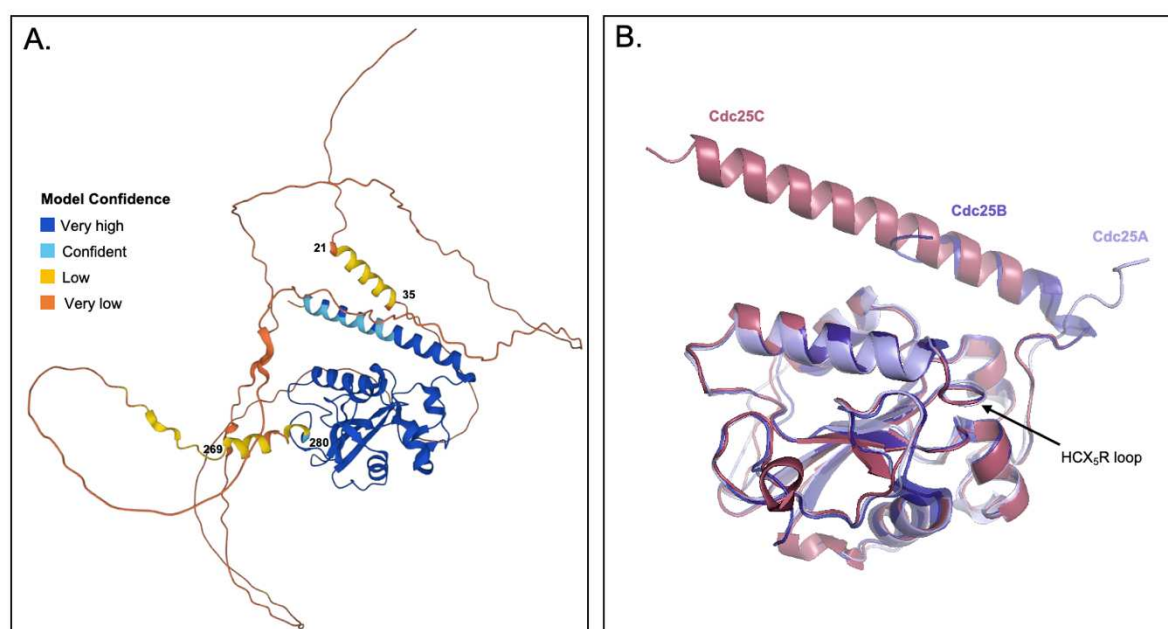


Figure 8: Alpha fold model of Cdc25C (A) and its comparison to the catalytic domains of the other CDC25 phosphatases (B). A. The alpha-fold model of Cdc25C predicts, with very high confidence (dark blue), the fold of the catalytic domain (residues 280-473). The *N*-terminal region is predicted to be intrinsically disordered with two short α -helices predicted with a low confidence level (yellow). B. The overlay of crystal structures of the catalytic domains of Cdc25A (residues 336–523) and Cdc25B (residues 394–580) with the alpha-fold prediction of the catalytic domain of Cdc25C (residues 277–473) displays a conserved fold between the three isomers. The Cdc25C prediction displays a notably longer α -helix which is absent in Cdc25A and significantly shorter in Cdc25B. Cdc25C model downloaded from AlphaFold PDB. Diagram made in PyMol using PDB IDs 1C25: Cdc25CA and 1QB0: Cdc25CB.

Throughout interphase, the activity of Cdc25C remains low and is spatially regulated via its sequestration in the cytoplasm by binding to 14-3-3 proteins (Dalal *et al*, 1999; Peng *et al*,

1997). At the G₂ to M transition, its dissociation from 14-3-3 proteins along with the hyperphosphorylation of its N-terminal region characterises the active, mitotic form of Cdc25C. Notably, hyperphosphorylation of Cdc25C is initiated by its substrate, CDK1-Cyclin B (Strausfeld *et al*, 1994). Upon nuclear translocation of the CDK1-cyclin B complex, its full activation is reliant on dephosphorylation of inhibitory Thr14 and Tyr15 phosphorylations by Cdc25C. Reciprocally, the activation of Cdc25C by CDK1-cyclin B is accomplished through the phosphorylation of five key sites in its N-terminus. This creates a positive feedback loop, as the activity of CDK1-cyclin B is reinforced by its activation of Cdc25C, resulting in a gradual and complete activation of the CDK1-cyclin B complex, ultimately leading to the initiation of cell division.

In addition to CDK1-cyclin B and the 14-3-3 proteins, several other proteins play a crucial role in the regulation of Cdc25C throughout the cell cycle. *Xenopus* Plx1 (human homolog Plk1) was the first kinase other than CDK1 found to phosphorylate and activate Cdc25C at mitosis (Kumagai & Dunphy, 1996). Members of the polo-like kinase family, Plk1 and Plk3, phosphorylate several residues in the nuclear export sequence (NES) of Cdc25C, promoting nuclear import and retention, hence supporting its function at mitosis (Liu *et al*, 2020). Moreover, protein phosphatase 1 (PP1) directly removes inhibitory phosphorylations from Cdc25C, further enhancing its activity at mitosis. Additional studies have suggested that the complete activation of the mitotic form of Cdc25C is dependent upon its interaction with the Pin1 isomerase, however, the exact details and consequences of this regulation remain unclear, and will be dissected in more detail later. Conversely, there are also multiple proteins involved in the repression of Cdc25C, particularly limiting its activity in the presence of stressors (Liu *et al*, 2020). Significantly, Cdc25C is targeted as part of the DDR and DNA replication checkpoint pathways, crucial to the detection of DNA replication errors and preventing their propagation to daughter cells. Human Chk1 phosphorylates Cdc25C in response to activation by the DDR, provoking an inhibitory effect by promoting its nuclear export (Sur & Agrawal, 2016). Furthermore, Cdc25C is negatively regulated by p53, coined the ‘guardian of the genome’ due to its crucial role in provoking cell cycle arrest in response to various stress signals (Lane, 1992; Taylor & Stark, 2001). The exact mechanisms by which p53 exerts its inhibitory effect on Cdc25C are not fully understood, however p53 appears to act through Cdc25C transcriptional repression and direct binding (Clair & Manfredi, 2006; Ruppenthal *et al*, 2007).

Antagonistically, Cdc25C is also believed to have an additional role monitoring p53 expression and suppressing p53-induced cell cycle arrest (Sur & Agrawal, 2016).

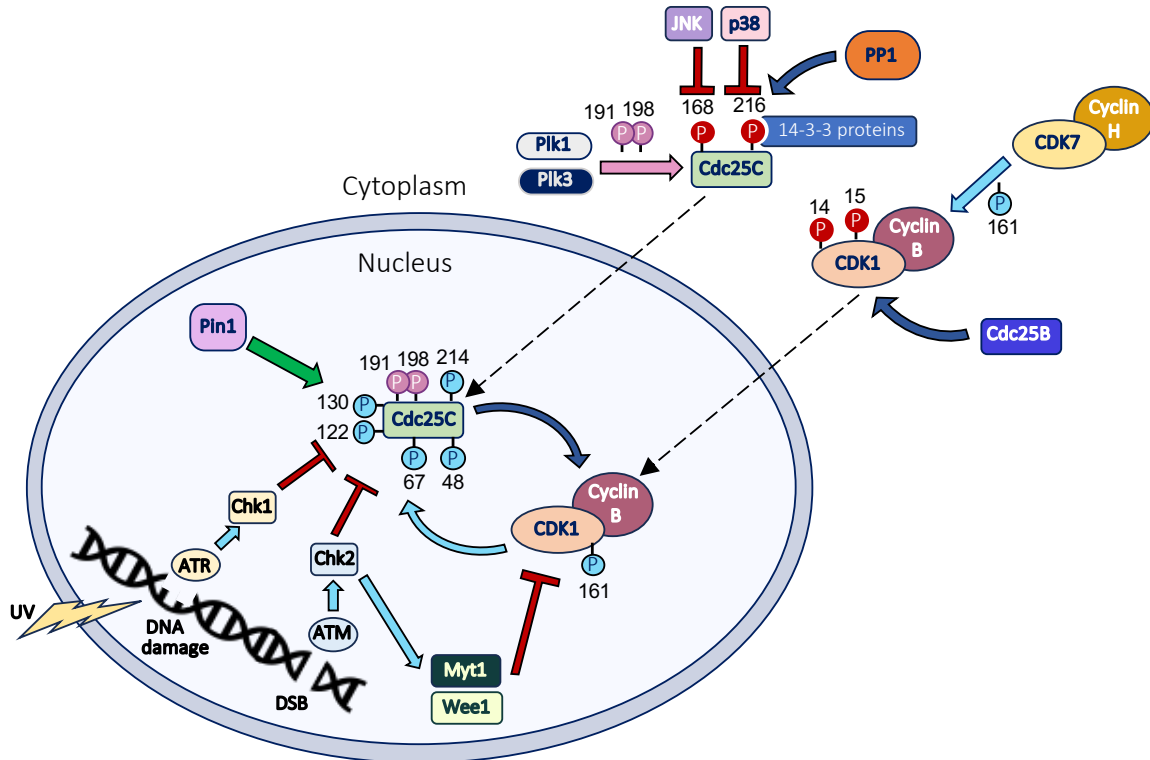


Figure 9: Spatio-temporal regulation of Cdc25C at the G₂ to M transition. Phosphorylation events are indicated by coloured 'P' symbols: red 'P' for inhibitory phosphorylation, blue 'P' for activating phosphorylation, and pink 'P' for phosphorylation promoting nuclear import. Activation and inhibition are depicted using arrows; standard arrows represent activation, while arrows with a vertical bar indicate inhibition. The colour of the arrows reflects the type of catalytic activity: light blue for activating phosphorylation, dark blue for activating dephosphorylation, pink for phosphorylation promoting nuclear import, green for unknown activation, and red for inhibitory phosphorylation. During interphase, Cdc25C is sequestered in the cytoplasm by 14-3-3 proteins bound to pS216. Phosphorylation of S168 further inactivates Cdc25C. Similarly, inactivating phosphorylations (pT14/pT15) on CDK1 prevent its activation during interphase. At the G₂ to M transition, the PP1 phosphatase removes inhibitory phosphorylations from Cdc25C, promoting dissociation from 14-3-3 proteins, which, accompanied by phosphorylation of the NES (p191/p198) by Plk kinases, promotes its nuclear import. Similarly, inhibitory phosphorylations are removed from CDK1 by Cdc25B and an activating phosphorylation (pThr161) is catalysed by CDK7 kinase. In the nucleus, an activation loop between Cdc25C and CDK1 promotes their full activation. Additionally, Pin1 contributes to complete Cdc25C activation by an unknown mechanism. Both Cdc25C and CDK1 are targeted as a part of the DDR and are inactivated in response to DNA damage or DSBs.

2.3 Differential phosphorylation status of Cdc25C throughout the cell cycle

The critical function of Cdc25C as a mitotic trigger necessitates its tight control by a complex network of aforementioned protein partners, ensuring its activation at the correct time and only in the presence of optimal cellular conditions and successful DNA replication. Its N-terminal regulatory region hosts a diverse array of serine and threonine residues that serve as crucial targets for a range of phosphorylation events, thereby providing a powerful means of controlling its activity. The pivotal regulatory phosphorylation events that can occur in the IDR of Cdc25C are summarised in figure 10 and will be examined in greater detail below.

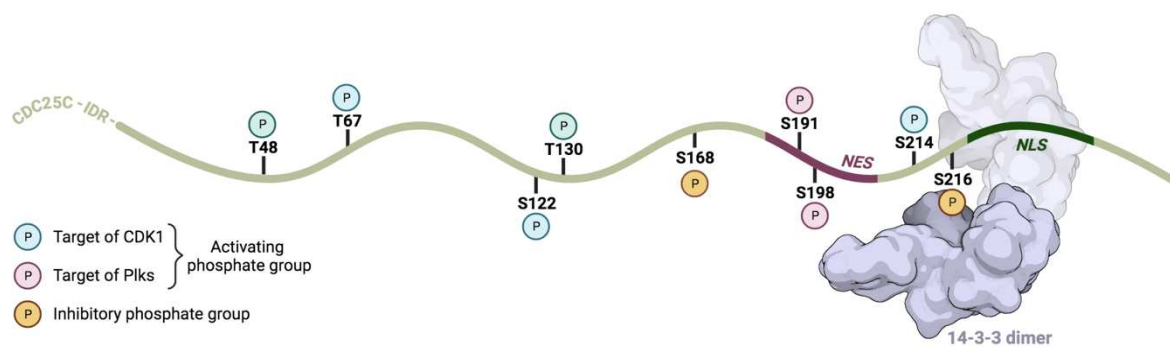


Figure 10: Summary of the key phosphorylation events occurring in the N-terminal regulatory region of Cdc25C. The five phosphorylation events (blue) which characterise the active form of Cdc25C are catalysed by CDK1-Cylin B. Additional activating phosphorylation events (pink) by the Plk kinases occur in the nuclear export sequence (NES) of Cdc25C ensure the nuclear localisation of Cdc25C during mitosis. Inhibitory phosphorylation events (yellow) help to maintain Cdc25C in the inactive state throughout interphase.

The repression of Cdc25C activity during interphase is heavily reliant on one particular inhibitory phosphorylation at Ser216 (Hutchins & Clarke, 2004). Phosphorylation of Ser216 provides a consensus sequence for members of the highly conserved 14-3-3 protein family, enabling their binding to Cdc25C (Peng *et al*, 1997; Yang *et al*, 2006). The binding of 14-3-3 dimers at this site has no apparent effect on catalysis but blocks the nuclear localisation signal (NLS) hence sequestering Cdc25C in the cytoplasm (Hutchins & Clarke, 2004).

Phosphorylation of this residue can be achieved by several stress-activated kinases, including Chks, p38, and other MAPKs. Hence modification of a single residue seemingly underlies the inhibition of Cdc25C in response to DNA damage or other stress signals. During mitosis, the active form of Cdc25C undergoes hyperphosphorylation, a phenomenon initially observed in *Xenopus laevis* Cdc25C, which displayed phosphorylation at five characteristic residues (Thr48, Thr67, Thr138, S205 and Ser285) (Sur & Agrawal, 2016). Notably, these phosphorylation sites exhibit a high level of conservation in humans, with the corresponding residues (Thr48, Thr67, Thr130, Ser168, and Thr214) also found in human Cdc25C (see figure 11).

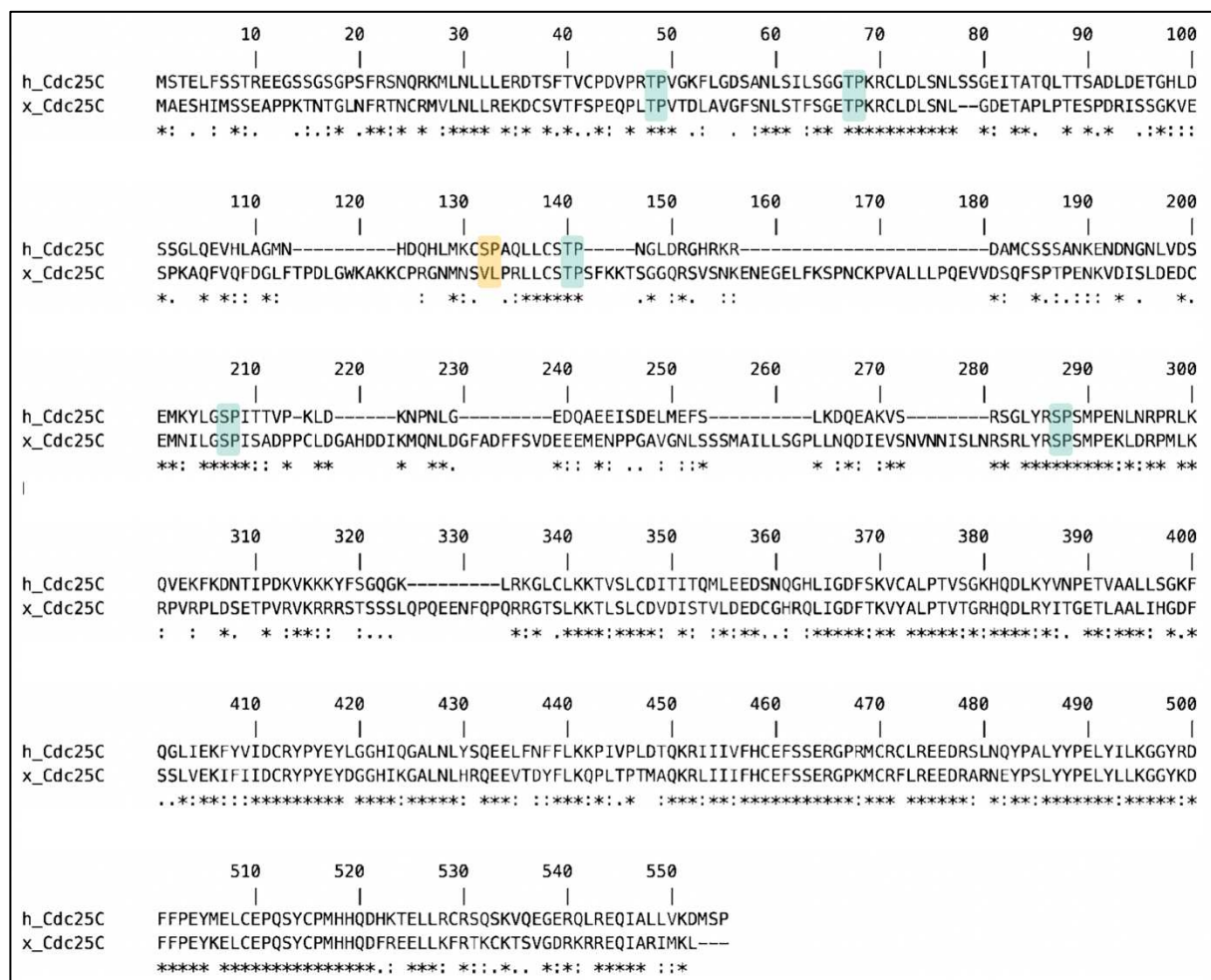


Figure 11: Sequence alignment of Cdc25C from *Homo sapiens* (h_Cdc25C) and *Xenopus laevis* (x_Cdc25C), highlighting the S/T-P sites in their regulatory N-terminal regions. Conserved sites between both proteins are marked in blue, while non-conserved sites are marked in yellow. Among the six sites in human Cdc25C, only one (Ser122) is not conserved in *Xenopus* Cdc25C. The remaining five sites, Thr48 (*Xenopus*: Thr48), Thr67 (*Xenopus*: Thr67), Thr130 (*Xenopus*: Thr138), Ser168 (*Xenopus*: Ser205), and Ser214 (*Xenopus*: Ser285), show conservation between the two proteins.

In humans, the hyperphosphorylated, active form of Cdc25C has since been characterised, identifiable by its phosphorylation at five key sites (Thr48, Thr67, Ser122, Thr130 and Ser214) in accordance with the characteristic CDK recognition sequence (Ser/Thr-Pro modules) (Sur & Agrawal, 2016). The five residues are all phosphorylated by the CDK1-cyclin B kinase *in vitro* and *in vivo*. Interestingly, phospho-Ser168 in Cdc25C also conforms to the recognition site of CDKs and is enriched at the G₂ to M transition (Liu *et al*, 2020). However, studies have suggested that its phosphorylation is achieved by the stress-activated kinase, JNK, during the G₂ phase, which elicits an inhibitory effect on Cdc25C activity. Thus, it is not confirmed whether CDK1 can phosphorylate Ser168 *in vivo* (Gutierrez *et al*, 2010). Adding further complexity to the effect of phosphorylation at this residue, its equivalent in *Xenopus* Cdc25C (Ser205) is characterised as one of the five activating phosphorylation events that occur during the transition to mitosis (Perry & Kornbluth, 2007). This disparity suggests that phosphorylation of this residue (Ser168 in humans and Ser205 in *Xenopus*) may have antagonistic effects, possibly due to differences in cellular context, adding complexity to the role of phosphorylation of Cdc25C. Of the five confirmed activating phosphorylation events in human Cdc25C, some have clear functions. For example, the presence of phospho-Thr130 appears to be required for the dissociation of Cdc25C from 14-3-3 proteins (Margolis, 2003; Margolis *et al*, 2006). Furthermore, phosphorylation Ser214 blocks the inhibitory phosphorylation of Ser216, supporting maintained Cdc25C activity and nuclear localisation throughout mitosis (Bulavin *et al*, 2003). The inhibitory effect of phospho-Ser214 on the phosphorylation of Ser216 persists even under stress conditions, highlighting the irreversibility of the mitotic transition once the final G₂ to M checkpoint has been cleared. During early mitosis, additional activating phosphorylation events occur in the nuclear export sequence (NES) of Cdc25C, on residues Ser191 and Ser198 by Plk3 and Plk1 respectively (Bahassi *et al*, 2004; Toyoshima-Morimoto *et al*, 2002). The phosphorylation of these residues disrupts nuclear export hence working to maintain Cdc25C in the nucleus for the duration of mitosis (Myer *et al*, 2005).

Although early studies often refer to two states of Cdc25C, characterised as the inactive hypo- and active hyper-phosphorylated forms, accumulating evidence suggests the existence of functional intermediary phospho-isomers (Bonnet *et al*, 2008). A complex network of Cdc25C phospho-isoforms has been identified, including those of full-length Cdc25C and its splice variants, each containing different combinations of the five characteristic mitotic

phosphorylation events (see figure 12). Furthermore, these different phosphorylation intermediates display unique spatio-temporal patterns and binding properties, suggesting that they may perform distinct functions in the cell (Franckhauser *et al*, 2010). Although the precise functions of these intermediary forms are not yet fully understood, their existence suggests that the conventional model of a CDK1-cyclin B/Cdc25C positive amplification loop may oversimplify the complexity of mitotic Cdc25C activation (Sur & Agrawal, 2016). Moreover, the presence of distinct Cdc25C splice variants and phosphorylation isoforms adds another layer of intricacy to the process of mitotic initiation, warranting further investigation.

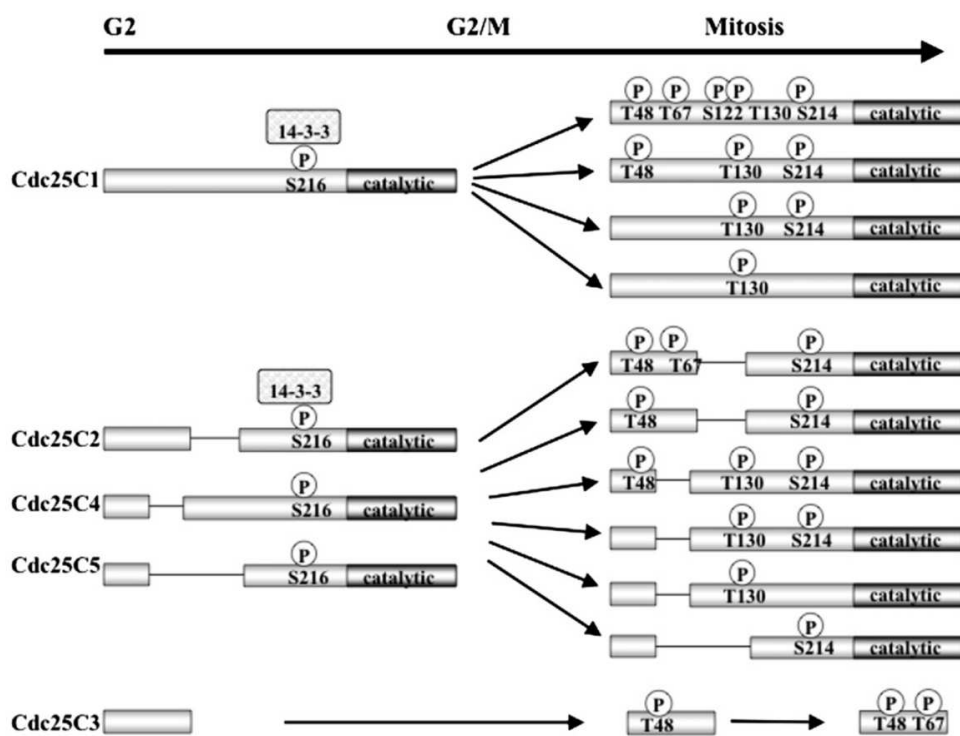


Figure 12: Schematic representation of the various phospho-isoforms of Cdc25C. Both full-length Cdc25C (Cdc25C1) and its isoforms (Cdc25C2-5) can undergo different numbers of activating phosphorylations at mitosis. Diagram taken from Bonnet *et al.*, 2008.

2.4 Conclusion

In summary, the CDC25 phosphatases play crucial roles in the regulation of CDK-cyclin complexes during the cell cycle, with Cdc25C serving as a key gatekeeper in the

initiation of mitosis. The activity of Cdc25C is under tight regulation by a complex network of proteins and pathways, and is achieved via a diverse range of PTMs occurring in its regulatory region, with Ser/Thr phosphorylation playing a dominant role. While the CDK1-cyclin B/Cdc25C positive amplification loop is a well-established mechanism of Cdc25C activation, the presence of intermediate phospho-forms indicates that the activation of Cdc25C is more intricate than initially thought. Further research on the specific phosphorylation events and their spatio-temporal regulation, as well as their interplay with binding partners, will be crucial for gaining a comprehensive understanding of Cdc25C function. In the following chapter, we will shift our focus to another significant regulator of the cell cycle, the peptidyl-prolyl isomerase Pin1, and its role in modulating the activity of Cdc25C and other key cell cycle regulators.

Chapter 3 The Pin1 Isomerase

The Pin1 isomerase has emerged as a master regulator of the cell cycle with its ability to modulate the activity of various proteins critical to cell division. As a prolyl isomerase, Pin1 catalyses the *cis-trans* isomerisation of phosphorylated serine/threonine-proline motifs, thereby regulating the conformation and activity of its target proteins. Pin1 is unique among prolyl isomerases in its specificity for these motifs and its ability to regulate a diverse range of substrates involved in a multitude of significant cellular processes including the cell cycle, transcription, apoptosis and DNA damage. This chapter will investigate the molecular mechanisms underlying Pin1 function, including its structure, catalytic mechanism and substrate binding properties, focussing on how these mechanisms link to its role in cell cycle regulation.

3.1 Peptidyl-prolyl isomerases (PPIases)

The functional properties of proteins are intricately linked to their structural organisation, which can be modulated by a range of PTMs and conformational rearrangements. Peptide bonds, which exhibit a partial double bond character, can adopt two distinct conformations: *cis* or *trans* (Lu *et al*, 2007). While most peptide bonds exist in the more energetically favourable *trans* conformation (with an energy barrier, $\Delta G^\circ \approx 10-17$ kJ.mol⁻¹ between the *cis* and *trans* ground states), proline residues, with their unique cyclic structure, display an unusually conformationally restrained peptide bond which facilitates conversion between the *cis* and *trans* states, which are closer in free energy ($\Delta G^\circ \approx 3-4$ kJ.mol⁻¹) (Drakenberg *et al*; Matena *et al*, 2018). The conformational flexibility offered by prolyl isomerisation can have significant implications for protein structure and function, as it can act as a molecular switch to regulate the activity and binding properties of a protein through conformational changes. However, in accordance with Linus Pauling's model of the strain energy associated with peptide bond rotation, the process of *cis-trans* prolyl isomerisation involves passing through a high energy transition state (TS), which is approximately 30 kcal mol⁻¹ higher in free energy than the *trans* state (see figure 13). Due to this significant energy barrier, prolyl isomerisation is an inherently slow process, making it unsuitable for use in fast

signal transduction pathways. To overcome this challenge, specialised enzymes called peptidyl-prolyl isomerases (PPIases) have evolved to accelerate this isomerisation process into physiologically relevant timescales (Eberhardt *et al*, 1992; Matena *et al*, 2018).

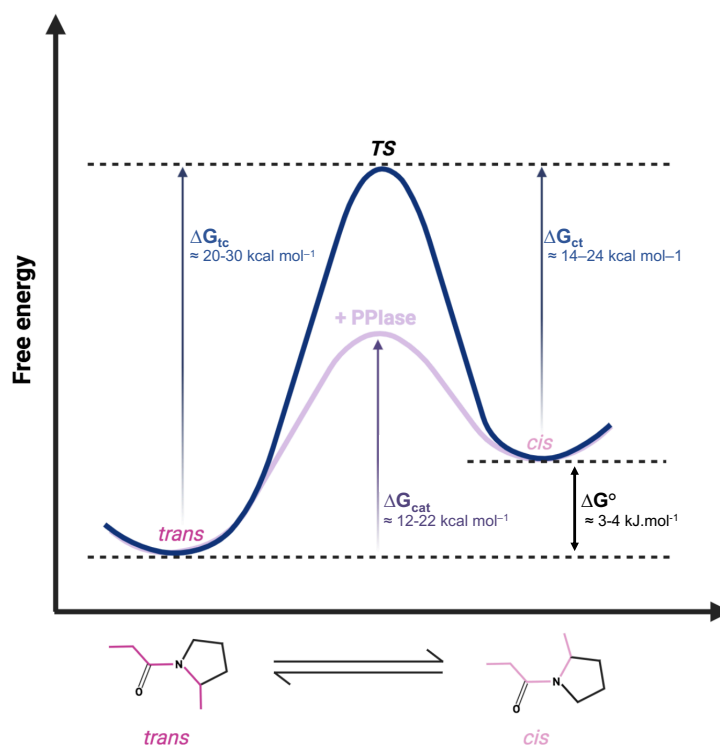


Figure 13: Free energy diagram of prolyl isomerisation and its catalysis by PPIases. PPIases decrease the free energy of activation by about 8 kcal.mol⁻¹.

PPIases can be broadly categorised into four distinct families based on their structural and functional characteristics: FK506-binding proteins (FKBPs), cyclophilins, parvulins, and protein phosphatase 2A phosphatase activators (PTPAs). While all PPIases share a common catalytic activity, the different families have distinct sequential, structural, and substrate binding properties. Although their overall fold and sizes differ, cyclophilins, FKBP and parvulins display similarities in the structure of their PPIase domains and the organisation of their catalytic sites (see figure 14) (Lu *et al*, 2007). Parvulins are the smallest subclass of the PPIase family, and so far, three members have been characterised in humans, the Par14/Par17 isoforms and Pin1. Pin1 was the first PPIase to be discovered that is essential for cell viability (Lu *et al*, 1996). Pin1 is known for its unique substrate specificity in exclusively recognising phosphorylated serine or threonine residues preceding the prolyl peptide bond. Since its discovery, Pin1 has been heavily studied due to its promiscuous binding to a large range of

proteins involved in essential cellular processes. Despite significant progress in understanding its mode of action, the precise mechanisms by which Pin1 recognises and exerts a regulatory effect upon its target proteins remain incompletely understood and are the subject of ongoing investigation.

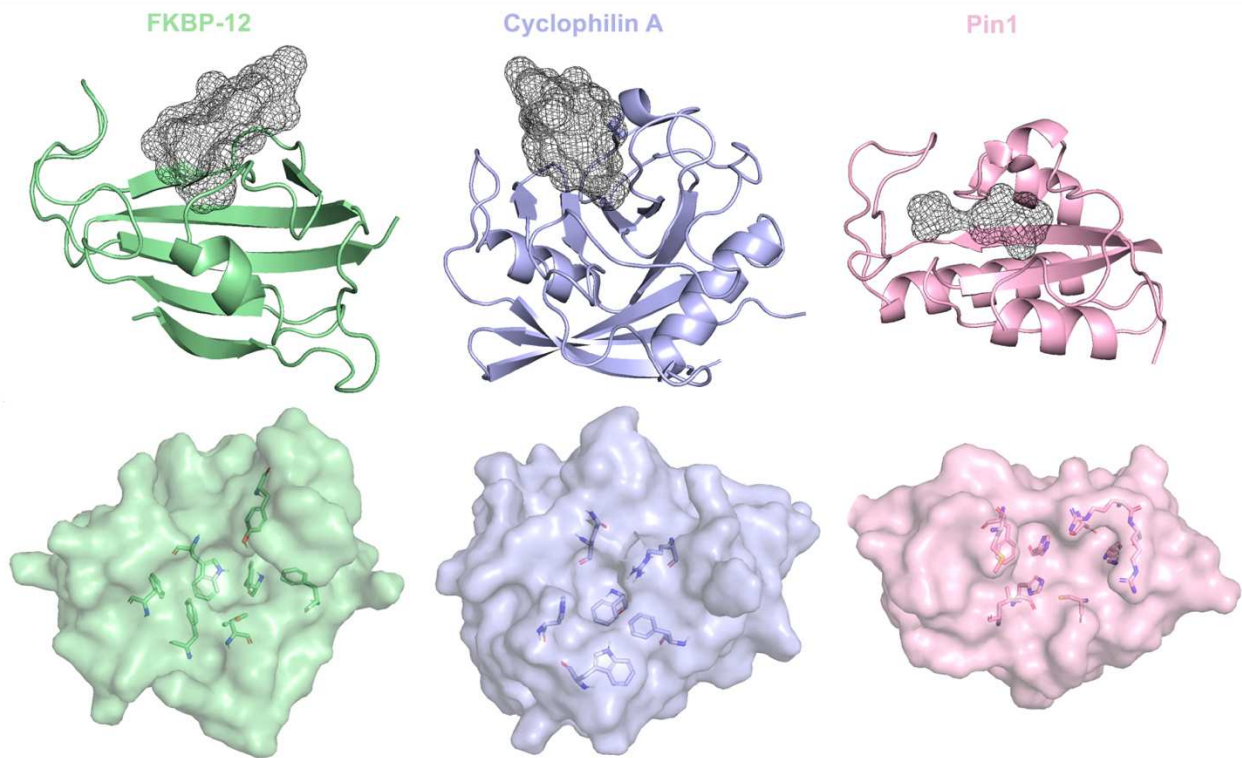


Figure 14: Comparison of the structural and active site organisation of the PPIase domains of FKBP, cyclophilins and parvulins. The overall structure of the PPIase domains of human FKBP-12, cyclophilin A and Pin1 are compared (top). Their respective high-affinity ligand binding sites are represented in mesh format. Their respective PPIase domains show a similar fold, organised into antiparallel β -sheets which position a short α -helix to form a solvent-exposed active site. The shallow, broad active site (bottom) is characteristic of PPIase domains and key residues involved in catalysis and ligand binding are shown for each residue. Diagram created in PyMol PDB ID: 1FKB, 1BCK and 1PIN).

3.2 The diverse roles of Pin1

One of the key aspects underlying Pin1's role in cell vitality is its extensive interactome, which encompasses numerous proteins that play fundamental roles in essential cellular processes such as the cell cycle, DNA damage response, cell development and transcriptional

regulation (Liou *et al*, 2011). Each of these processes is tightly regulated to ensure their accurate execution, as previously demonstrated for the cell cycle. A significant regulatory mechanism underlying many complex signalling cascades in the cell is the reversible phosphorylation of serine and threonine residues, which modulates their activity and interaction with other proteins (Pawson & Scott, 2005). Pin1's ability to act as a specific regulator of key serine/threonine phosphoproteins acting in these processes has firmly established it as an attractive subject for research. Considering the significant interactome of Pin1, the question arises as to how this enzyme can differentially regulate such a wide range of targets. Prolyl isomerisation by Pin1 can exert regulatory effects on its substrate through several biological pathways, dependent on the specific target protein (see figure 15) (Zhou & Lu, 2016).

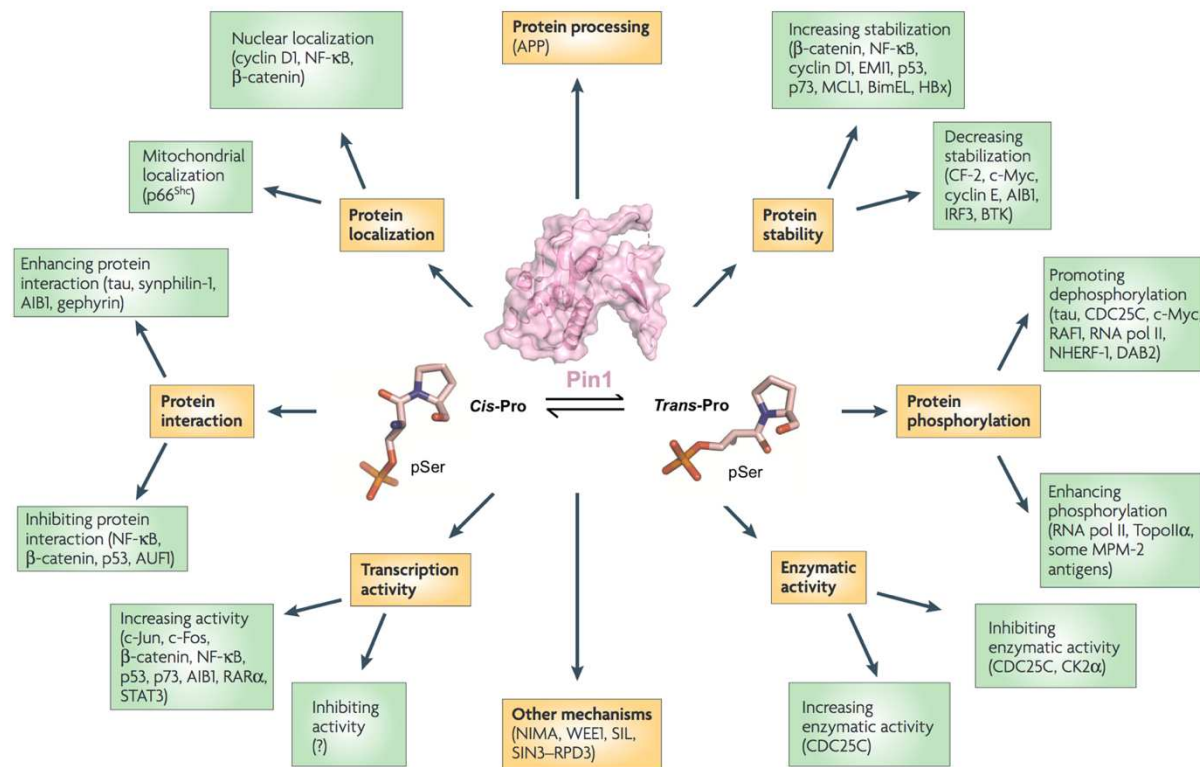


Figure 15: The varied mechanisms by which Pin1 regulates its target proteins. Prolyl isomerisation catalysed by Pin1 can lead to large conformational changes in its target protein. The resulting regulatory effect of this conformational change can take various forms, underlying how Pin1 can differentially regulate such a vast interactome. For example, Pin1-catalysed isomerisation directly influences the proteolytic processing pathway of amyloid precursor protein (APP), directing it down the non-amyloidogenic route. Other regulatory mechanisms of Pin1-catalysed prolyl isomerisation, such as those influencing phosphorylation and degradation, rely on the specific recognition of trans or cis motifs by downstream phosphatases, kinases, and ubiquitin ligases. Diagram modified from Zhou *et al.*, 2016.

One of the most well-established pathways by which Pin1 exerts control is by regulating of the stability of its substrates (Liou *et al*, 2011). In the cell, the selective degradation of proteins is a key regulatory mechanism to ensure the population of each component is spatially and temporally controlled, significant for the orchestration of complex signalling pathways. The most common cellular degradation process occurs via the ubiquitin proteasome pathway (UPS), which relies on the timely ubiquitination of proteins by E3 enzymes (ubiquitin ligases) which target them to the 26S proteasome where they subsequently undergo proteolysis (Schwartz & Ciechanover, 2009; Rock *et al*, 1994). The targeting of certain proteins for ubiquitination, and hence degradation, can be phosphorylation-dependent, as some E3 ligases exclusively recognise phospho-proteins (Liou *et al*, 2011). This is the case for the multi-subunit ubiquitin ligase, Skp1Cdc53/Cullin-F box protein (SCF) complex, as some F-box proteins (the variable substrate-targeting modules of these complexes) selectively recruit substrates containing phosphorylated Ser/Thr residues (Orlicky *et al*, 2003). Notably, this recognition can be isomer dependent, as is the case for Cdc4 f-box subunit of the SCF^{Cdc4} complex which has displayed a binding preference to pSer/Thr-Pro modules in the *trans* conformation. Therefore, Pin1 prolyl isomerisation can play a role here by either promoting or protecting against ubiquitination, depending on the direction of the isomerisation reaction it catalyses (Liou *et al*, 2011). For instance, Pin1 is seen to play a protective role in the stability of cyclin D1, a key regulator of the G1 to S transition of the cell cycle (Liou *et al*, 2002). Pin1 directly binds to and isomerises the Thr286-Pro motif of Cyclin-D1 which prevents its nuclear export and ubiquitin-mediated degradation (Diehl *et al*, 1997; Liou *et al*, 2002).

Another method by which Pin1 is seen to regulate some of its targets is via controlling their phosphorylation status. The reversible phosphorylation of serine, threonine and tyrosine residues acts as a crucial signal in cellular pathways, enabling the recruitment of these proteins, provoking a conformational change to regulate their activity or altering their ability to bind partners (Lu & Zhou, 2007). A significant phospho-motif, Ser/Thr-Pro, is specifically targeted by the large superfamily of Pro-directed protein kinases, encompassing the CDKs, JNKs, PLKs, p38 kinases, extracellular signalling-related kinases (ERKs), and glycogen synthase kinase 3 (GSK3). The post-phosphorylation isomerisation of S/T-P motifs by Pin1 is particularly significant in such signalling pathways as many pro-directed kinases and phosphatases of the cell are isomer specific, hence shifting the balance towards the *cis* or *trans* population can greatly impact their ability to recognise their substrate. For instance, in the case

of tau protein, which is involved in neuronal regeneration, Pin1-catalysed prolyl isomerisation of the Thr231-Pro motif triggers its dephosphorylation by the *trans*-specific phosphatase PP2A (Zhou *et al*, 2000). This regulatory mechanism plays a notable protective role, as it prevents the accumulation of pathogenic hyperphosphorylated tau, a hallmark feature observed in Alzheimer's disease (Lu *et al*, 1999a). Pin1's ability to selectively recognise and regulate such a diverse array of substrates raises the question of how it maintains such a remarkable level of substrate specificity. To gain a deeper insight into this phenomenon, a thorough investigation of Pin1's structure and catalytic mechanism is warranted.

3.3 Structure and catalytic mechanism

Pin1 is a small protein of 163 amino acids organised into two distinct domains, the *N*-terminal WW domain (residues 1-39) and the *C*-terminal PPIase domain (residues 50-163), which are separated by a short flexible linker (see figure 16) (Ranganathan *et al*, 1997).

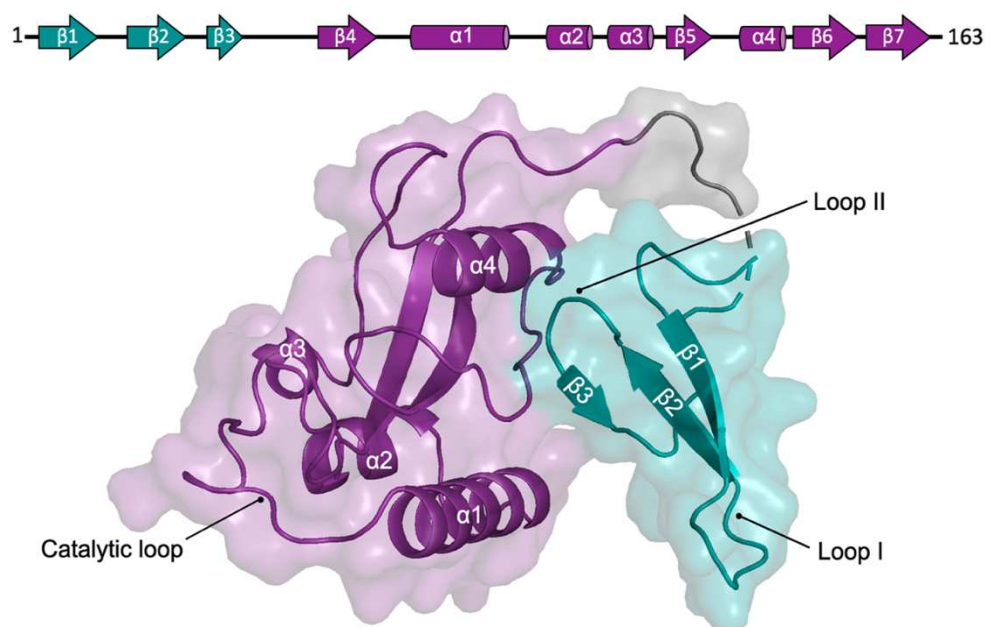


Figure 16: X-ray crystal structure of Pin1. The cartoon depiction of the structure of Pin1 shows its organisation into 2 domains, the WW domain (cyan) and the PPIase domain (magenta). The WW domain is composed of 3 antiparallel β -sheet (β 1-3). The PPIase domain is composed of a core 4-stranded antiparallel β -sheet (β 4-7) and 3 α helices (α 1-3). Diagram created using PyMol (PDB ID: 1PIN)

The WW domain of Pin1, which gets its name from a pair of conserved tryptophan residues, W11 and W34, shares similarities to WW domains of other proteins (Lee *et al.*, 2022). WW domains are short protein modules with a binding preference for proline-rich motifs, and are often organised into a 3-stranded antiparallel β -sheet (Smerdon & Yaffe, 2010). A subclass of WW domains, the group IV WW domains, show specific binding to sequences containing phosphorylated Ser/Thr-Pro (Lee *et al.*, 2022; Lu *et al.*, 1999b). The specificity of the Pin1 WW domain for these motifs is largely determined by the residues present in its loop I region (the β 1- β 2 loop) (Smerdon & Yaffe, 2010). The structure of the WW domain of Pin1 bound to a short peptide sequence of the RNA polymerase II ($YpSPTpSPS$, where p indicates phosphorylation of the following residue) highlighted key residues involved in substrate binding (see figure 17) (Verdecia *et al.*, 2000). The crucial residues involved in the binding of the phosphate group are Ser16, Arg17 of loop I and Tyr23 of the β 2-strand, which each form hydrogen bonds with the oxygens of the phosphate to stabilise it. Additionally, the aromatic groups of Tyr23 and Trp34 act to sandwich the substrate proline residue. Markedly, the WW domain has been shown to display a binding selectivity for pSer/Thr-Pro motifs in the *trans* conformation. Studies using *cis* and *trans*-locked phosphopeptides found a K_D of 53 μ M for the *trans*-locked inhibitor to this domain, while the *cis*-locked inhibitor showed no specific binding to the WW domain (Namanja *et al.*, 2011).

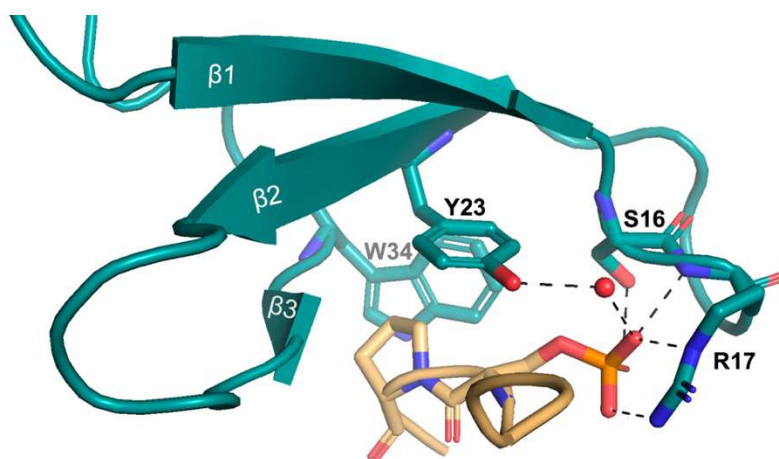


Figure 17: The binding of a phospho-peptide to the WW domain of Pin1. The WW domain of Pin1 is represented in cyan, and the $YpSPTpSPS$ peptide in beige. The sidechains of S16 and R17, as well as the mainchain nitrogen of R17, form hydrogen bonds (dashed lines) to the oxygens of the phosphate group of the second pSer residue (beige). The sidechain of Y23 also forms a hydrogen bond to the oxygen atom of the phosphate group, mediated by a water molecule (red sphere). The aromatic rings of Y23 and W34 coordinate the Pro. Diagram created using PyMol PDB ID:1F8A).

Like the WW domain, the PPIase domain of Pin1 also has the ability to bind sequences containing a pSer/Thr-Pro motif. Notably, while both domains can recognise the same substrate motifs, the PPIase domain is solely responsible for catalysing the isomerisation of prolyl peptide bonds, with the WW domain being devoid of catalytic activity. The structural architecture of the PPIase domain of Pin1 displays a characteristic parvulin topology whereby the typical secondary element ($\beta_1\alpha_1\alpha_2h\beta_2\alpha_3\beta_3\beta_4$, h =short helix) is folded into a core of 4 antiparallel β -strands and 4 external α -helices. The binding and catalysis of pSer/Thr-Pro motifs is facilitated by three critical regions of the PPIase domain, the substrate binding pocket, the basic triad and the catalytic tetrad. The first X-ray crystal structure of Pin1 in complex with an Ala-Pro dipeptide, revealed a hydrophobic substrate binding pocket present in the PPIase domain (see figure 18a) (Ranganathan *et al*, 1997). This structure allowed the identification of three conserved residues, Leu122, Met130 and Phe134 which form hydrophobic interactions with the proline ring of the bound peptide (see figure 18b). The initial crystal structure of Pin1 additionally showed the binding of a sulphate ion *N*-terminal to the Ala-Pro peptide, giving the first indication of the residues interacting with the phosphate group in native phospho-substrates. The binding of such multivalent anions is coordinated by three basic residues: Lys63, Arg68 and Arg69 (see figure 18b) (Matena *et al*, 2018). This basic triad underlies the unique specificity of Pin1 for phosphorylated substrates. Subsequent structures of Pin1 complexed to *cis*- or *trans*-locked phospho-peptides have confirmed that the two key binding zones of the PPIase domain can equally accommodate both isomers (Zhang *et al*, 2012b). However, the PPIase domain displays a higher affinity binding to phospho-peptides in the *cis* conformation ($K_D = 7 \mu\text{M}$) than those of the *trans* conformation ($K_D = 66 \mu\text{M}$) (Namanja *et al*, 2011). Notably, it has been highlighted that Arg68 of the basic triad is highly flexible, facilitating the rotation of the phosphate group during catalysis and enabling the binding of both isomers.

The catalytic mechanism underlying prolyl isomerisation by Pin1 has been subject to some debate; nonetheless, it is widely acknowledged that this process involves four essential residues, Cys113, His59, His157, and Ser154, collectively known as the catalytic tetrad (see figure 19) (Matena *et al*, 2018; Ranganathan *et al*, 1997). Based on analysis of the molecular structure of Pin1, it was initially suggested that the catalytic tetrad was arranged in a manner that permitted a covalent catalytic mechanism (Ranganathan *et al*, 1997). According to this mechanism, His59 was proposed to act as the catalytic base by deprotonating Cys113,

facilitating a subsequent nucleophilic attack of the peptide bond's carbonyl carbon by the thiolate group of Cys113. The process would involve the passage through a high-energy covalent tetrahedral intermediate, bound to the catalytic cysteine, which would ultimately relax back to the *cis* or *trans* ground state. Here, protonated Ser154 and His157 residues would act to stabilise the intermediate and prime His59 for catalysis.

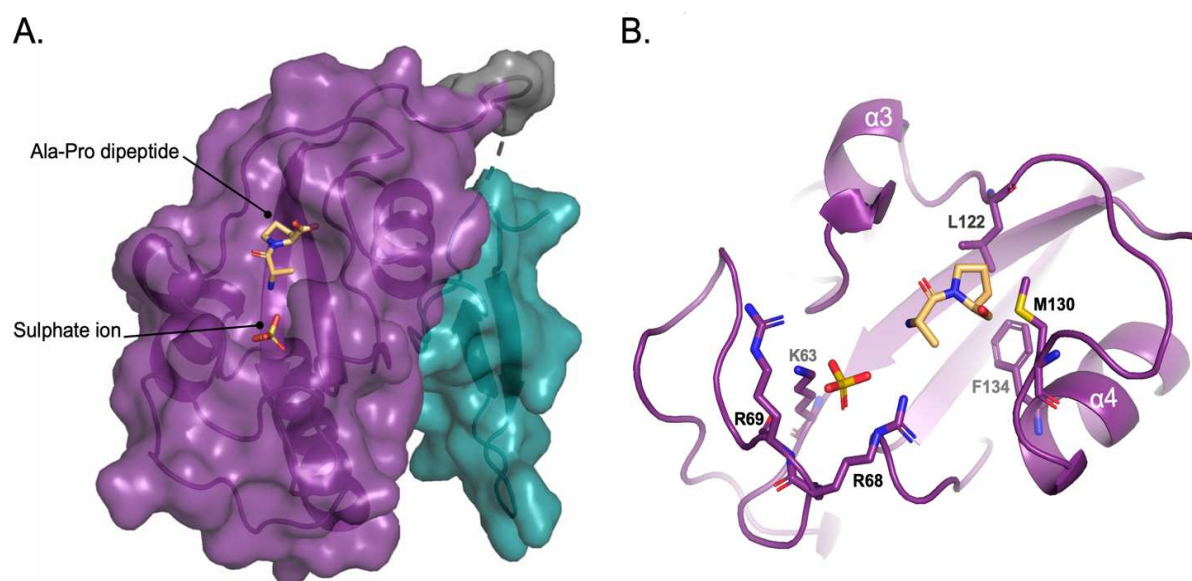


Figure 18: The substrate binding pocket of the PPIase domain. **A.** The binding pocket of the Ala-Pro dipeptide and sulphate ion to the back side of the PPIase domain (purple). **B.** The key hydrophobic residues involved in coordinating of the proline residue of the Ala-Pro dipeptide (beige) are indicated (L122, M130, and F134). These residues form hydrophobic interactions with the proline ring of the substrate. The residues of the basic triad (K63, R68, and R69) form electrostatic interactions with the bound multivalent anion, in this case a sulphate ion. In the native substrate, a phosphate group would be bound at this site. Diagram made using PyMol (PDB ID: 1PIN).

Alternatively, a second, non-covalent, catalytic mechanism was proposed based on biochemical studies. It was found that mutation of Cys113 to an aspartate residue did not impact PPIase activity *in vivo* or *in vitro*, indicating that catalysis may not rely on this residue behaving as a nucleophile (Behrsin *et al*, 2007). Instead, it was suggested that Cys113 acts to destabilise the double bond character of the substrate peptide bond. The partial double bond character of the peptide bond originates from charge delocalisation from the amide nitrogen to the carbonyl carbon, subsequently leading to the carbonyl oxygen assuming a negative charge

(see figure 20a). As a result of this configuration, the peptide bond has restricted rotation. In the current catalytic model, it is suggested that the partial negative charge of Cys113 disfavors this charge delocalisation, destabilising the double bond character and hence lowering the energetic barrier for rotation. In this model, the observed hydrogen bonding network surrounding Cys113, involving the other members of the catalytic tetrad, could act to stabilise the ionised state of Cys113 and the proposed twisted amide transition state. Further support for this mechanism was obtained through studies of Pin1 transition state analogues (TSAs). TSA molecules which acted as electrophilic acceptors of a potential nucleophilic attack by Cys113 were found to be poor inhibitors (Xu *et al.*, 2012). By contrast, TSA inhibitors resembling a twisted amide transition state were more effective in binding to Pin1 and inhibiting its activity (Xu *et al.*, 2011). The crystal structures of Pin1 in complex with these TSAs revealed that an additional active site residue, Gln131, was within hydrogen bonding distance of the prolyl C=O group. The coordination of this group by Gln131, as well as the phosphate group by the basic triad could explain how the enzyme forces the prolyl peptide bond into a stretched conformation prior to rotation (see figure 20b) (Mercedes-Camacho *et al.*, 2013; Xu *et al.*, 2011).

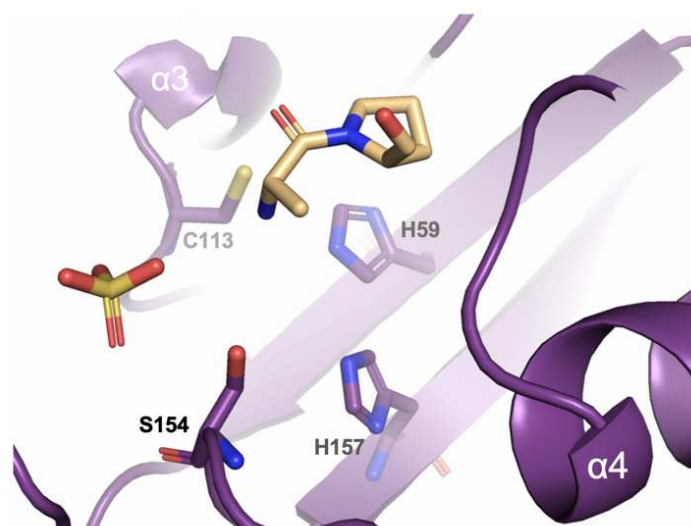


Figure 19: The catalytic tetrad of Pin1. The catalytic tetrad is comprised of His59, Cys113, Ser154 and His157 which are all arranged in close proximity to the proposed binding site of the prolyl peptide bond (here the Ala-Pro dipeptide is displayed in beige, and the bound sulphate ion which represents the phosphate binding site). Diagram created using PyMol (PDB ID: 1PIN)

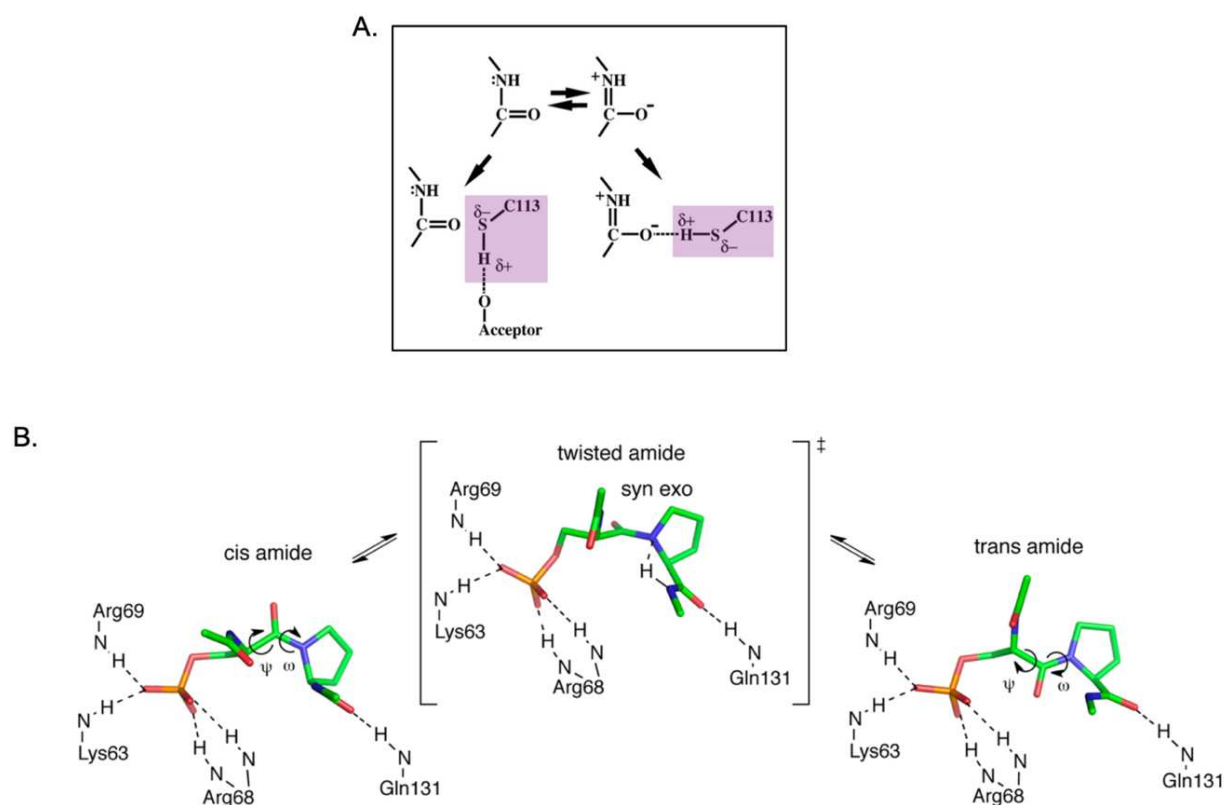


Figure 20: The proposed non-covalent catalytic mechanism of Pin1 prolyl isomerisation. **A.** Charge delocalisation from the amide nitrogen to the carbonyl carbon promotes the double bond character of the peptide bond, leading to the adoption of a negative charge by the carbonyl oxygen. In the case of the Pin1 active site (purple box), if Cys113 were to donate a hydrogen bond to the carbonyl oxygen atom, this state would be stabilised. However, the placement of this peptide bond into a negatively charged environment, as is the case if the partial negative charge of the Cys113 residue was stabilised, would stabilise the double bond character of the carbonyl C=O bond and hence the single bond character of the peptide bond (Behrsin et al., 2007) **B.** The proposed catalytic mechanism of Pin1 was modelled based on crystal structures of Pin1 in complex with reduced amide TSAs. Here, a pSer/Pro peptide bond was modelled in the active site of Pin1, in an extended conformation due to hydrogen bonding to the basic triad at the phosphate group, and the opposing C-terminal prolyl carbonyl group coordinated by Gln131. This allows the passage through a twisted amide TS Figure taken from Mercedes-Camacho et al., 2013.

3.4 Substrate binding models

One of the key features differentiating Pin1 from its other PPIase family members is its specificity for phosphorylated Ser/Thr-Pro motifs. Despite the WW domain and PPIase domain both recognising the same motifs, each domain has been shown to display different substrate binding preferences in terms of proline conformation. The WW domain exclusively recognises

pSer/Thr-Pro motifs in the *trans* conformation, while the PPIase domain can bind to motifs in either conformation, albeit with a notable preference for *cis* motifs. Interestingly, for several important Pin1 substrates, including mitotic proteins Cdc25C and Myt1, the WW domain is seen to act as the predominant binding module (Verdecia *et al*, 2000). The observed higher affinity of the WW domain for Pin1 substrates *in vitro* has suggested that this domain acts as a substrate-targeting module (Lu *et al*, 1999b; Smet *et al*, 2005).

As previously mentioned, a remarkable feature of Pin1 is that while both domains are known to recognise and bind to the same substrate motifs, only the PPIase domain has the ability to catalyse their isomerisation. Given this observation and other lines of evidence, such as the WW domain's apparent higher affinity and preference for binding to motifs in the *trans* conformation, it remains unclear how Pin1 interacts with substrates in a coordinated manner. Several models have been proposed to illustrate the function of the two domains in substrate binding (summarised in figure 21) (Innes *et al*, 2013; Lee & Liou, 2018). Firstly, the '**sequential binding model**' was suggested, taking into account that the WW domain shows a higher affinity for substrates than the PPIase domain (Lee & Liou, 2018). This model suggests that the WW domain first binds to a pSer/Thr-Pro motif within a substrate target, increasing the local concentration of the PPIase domain, and allowing it to bind to and isomerise a second pS/T-P motif on the same substrate in the immediate vicinity (multiple-motif binding) (figure 21a). In a subtle variation of this model, the binding of a pSer/Thr-Pro motif to the WW domain triggers conformation changes that allow the PPIase domain to compete with the WW domain. The same pSer/Thr-Pro motif can then be isomerised by the PPIase domain (single-motif binding). Alternatively, the '**multimeric binding model**' has been proposed for substrates belonging to a multimeric protein complex containing an active Pro-directed kinase (figure 21b). Here, it is suggested that the WW domain would first bind to a pSer/Thr-Pro motif in the kinase. The kinase then phosphorylates a Ser/Thr-Pro motif in the Pin1 substrate, allowing the PPIase domain, which is already in high local concentration, to bind and catalyse isomerisation. This model is supported by the observation that many Pin1 substrates are themselves kinases. The '**catalysis first**' model for Pin1 substrate binding is based on the evidence that the WW domain has a strong binding preference for motifs in the *trans* conformation. This model thereby states that the PPIase domain binds to motifs in the *cis* conformation, catalysing their isomerisation (figure 21c). Following the isomerisation reaction, the WW domain recognises and binds to the resulting *trans* motifs, in order to prevent the reverse reaction from occurring.

This model implies that the reaction proceeds in a *cis*-to-*trans* direction, mediated by the intricate interplay between Pin1's PPIase and WW domains. Finally, a ‘**simultaneous binding model**’ was proposed for Pin1 substrates containing more than one pSer/Thr-Pro motif, whereby both domains of Pin1 interact simultaneously with neighbouring motifs (figure 21d). Nevertheless, there is not enough evidence at present to undeniably confirm any of these models. Moreover, recent studies suggest that the interaction of Pin1 with substrates may be influenced by an allosteric communication between the two domains and is hence likely more complex than originally thought.

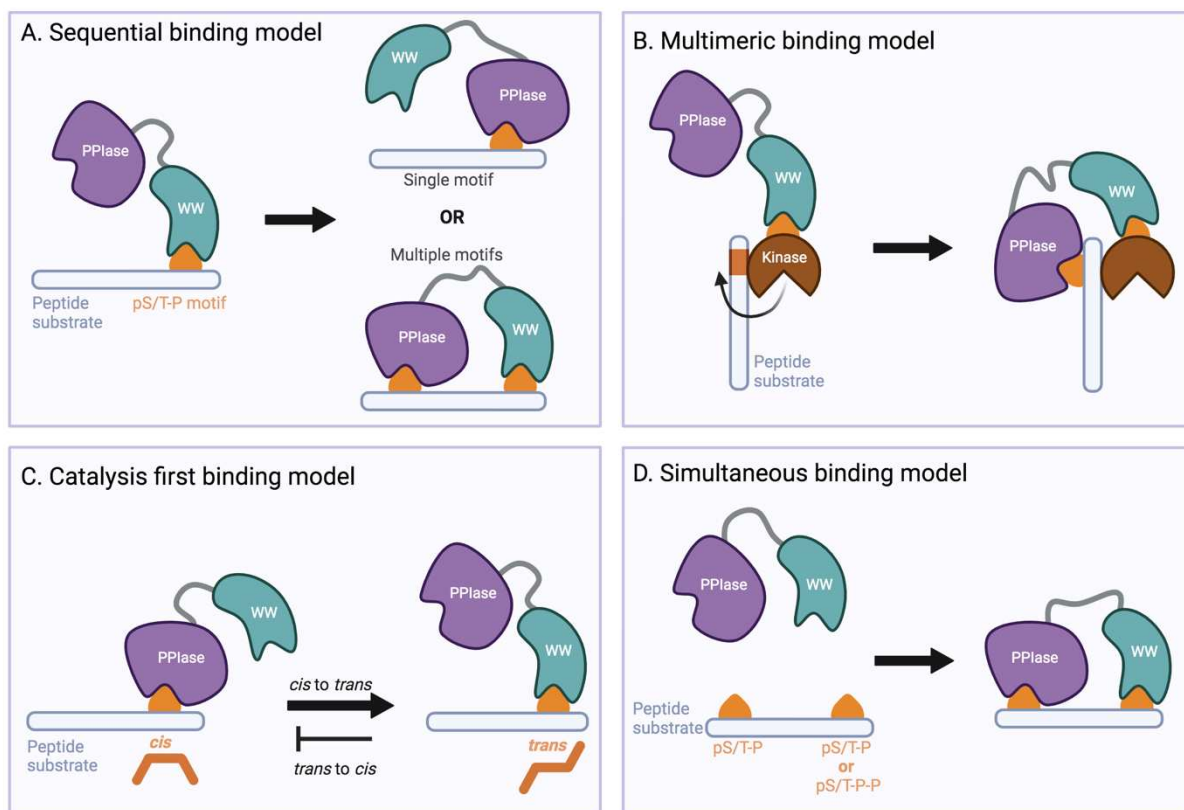


Figure 21: Pin1 substrate binding models. **A.** The sequential binding model states that, due to its higher affinity for pS/T-P motifs, the WW domain first binds to the target. The increased local concentration of the PPIase domain allows it to displace the WW domain or bind to a neighbouring recognition site in substrates containing multiple pS/T-P motifs. **B.** The multimeric binding model suggests that the WW domain binds to a pS/T-P motif within an active kinase within a multimeric complex containing the Pin1 substrate. The PPIase domain can then bind to pS/T-P motif within the substrate. **C.** The catalysis first binding model suggests that the PPIase domain binds and isomerises pS/T-P motifs in the *cis* conformation. The role of the WW domain would be to sequester motifs in the *trans* configuration to prevent the reverse reaction from occurring. **D.** The simultaneous binding model suggests that both domains of Pin1 can simultaneously bind to pS/T-P motifs.

3.5 Interdomain allostery

While the independent roles of the PPIase and WW domains in catalysis and substrate binding by Pin1 have been established overall, growing evidence suggests the potential for interplay between these domains in the form of interdomain allostery. Allostery is a widely studied regulatory mechanism found in many proteins, which enables ligand binding at one site to exert control over the activity of a distal functional site. Initially, the phenomenon of allostery was predominantly explained in terms of structural changes, whereby a conformational change between two states was observed upon ligand binding (Motlagh *et al*, 2014). However, more recent studies suggest that the concept of 'dynamic allostery' more accurately describes the process, indicating that allostery can occur even in the absence of significant structural alterations. This emphasises the critical role of protein motions and fluctuations in facilitating allosteric communication between different regions of a protein. There is mounting evidence to suggest that Pin1 undergoes such dynamic allostery upon ligand binding to its WW domain, provoking a change in the nature of the distal active site of the PPIase domain. As such, a comprehensive understanding of the proposed interdomain allostery within Pin1 is essential to fully unravel the mechanism of Pin1 binding and activity, also potentially holding significant implications for the development of therapeutic candidates targeting Pin1.

An interplay of the WW and PPIase domains can first be accounted for by the impact of the WW domain on PPIase activity. As previously mentioned, while the WW domain of Pin1 recognises phosphorylated substrate motifs, this domain cannot catalyse the *cis/trans* isomerisation of these motifs and hence its function within the full-length protein initially remained somewhat elusive. Despite lacking catalytic activity, its presence has been shown to be critical for Pin1 function *in vivo* (Behrsin *et al*, 2007). Notably, several single mutations in the WW domain were not only reported to weaken the structural stability of the WW domain, but also to alter the organisation of residues of the substrate binding pocket of the PPIase domain. Additional evidence suggested that the catalytic activity of full-length Pin1 *in vitro* ($k_{ex} = 1 \text{ s}^{-1}$) is ~ 10 x higher than that of the isolated PPIase domain ($k_{ex} = 0.8 \text{ s}^{-1}$), confirming the importance of the WW domain for PPIase activity (Landrieu *et al*, 2006). More specifically, studies have since indicated that the WW domain can both positively and negatively regulate

PPIase activity depending on the number of phosphorylated motifs in the studied peptide substrate (Peng, 2015; Smet *et al*, 2005). Hence, the binding of this domain to substrate motifs clearly impacts the function of the PPIase domain, supporting the presence of dynamic allostery within Pin1.

Nuclear Magnetic Resonance (NMR) spectroscopy has been widely adopted for exploring ligand binding and dynamic allostery within Pin1, and this choice of methodology is substantiated by several key factors. Firstly, Pin1's relatively small protein size makes it exceptionally compatible with NMR analysis. Moreover, NMR serves as a powerful tool for investigating molecular motions at the residue level, encompassing a broad range of timescales. This includes characterising protein motions on the μs to ms timescale which align closely with the rates of ligand binding events (ranging from 10^2 - 10^6sec^{-1}). Therefore, NMR proves to be a good technique for tracking interdomain dynamics triggered by Pin1-ligand interactions. Additionally, NMR offers the capacity to measure ligand binding affinities, particularly for those in the micromolar range, facilitating the extraction of quantitative insights at the single-residue level. These considerations collectively support the rationale behind employing NMR spectroscopy to explore the concept of interdomain allostery within Pin1.

To further explore the effect of substrate binding on the internal dynamics of the protein, NMR ^{15}N backbone relaxation studies were performed on Pin1 in the absence and presence of different substrates, to reveal the effects of substrate binding on backbone residue flexibility within Pin1. Concurrently, ligand binding affinities were estimated by probing ^1H - ^{15}N binding-induced chemical shift perturbations (CSPs) in Pin1. These studies firstly revealed that the WW and PPIase domains of apo-Pin1 display transient interdomain contact and can undergo almost independent tumbling motions in solution, with substrate binding able to reduce or increase interdomain flexibility (Bayer *et al*, 2003; Jacobs *et al*, 2003). These findings further suggested that the conformational space explored by Pin1 in solution is defined by two extreme conformers: a compact state in which the interdomain interface is well-defined, and an extended state in which the two domains appear to behave more independently (see figure 22). While both states exist in the apo-protein, the compact state has been found to be the dominant conformation, with a population ratio of 70:30 compared to the extended state (Born *et al*, 2021).

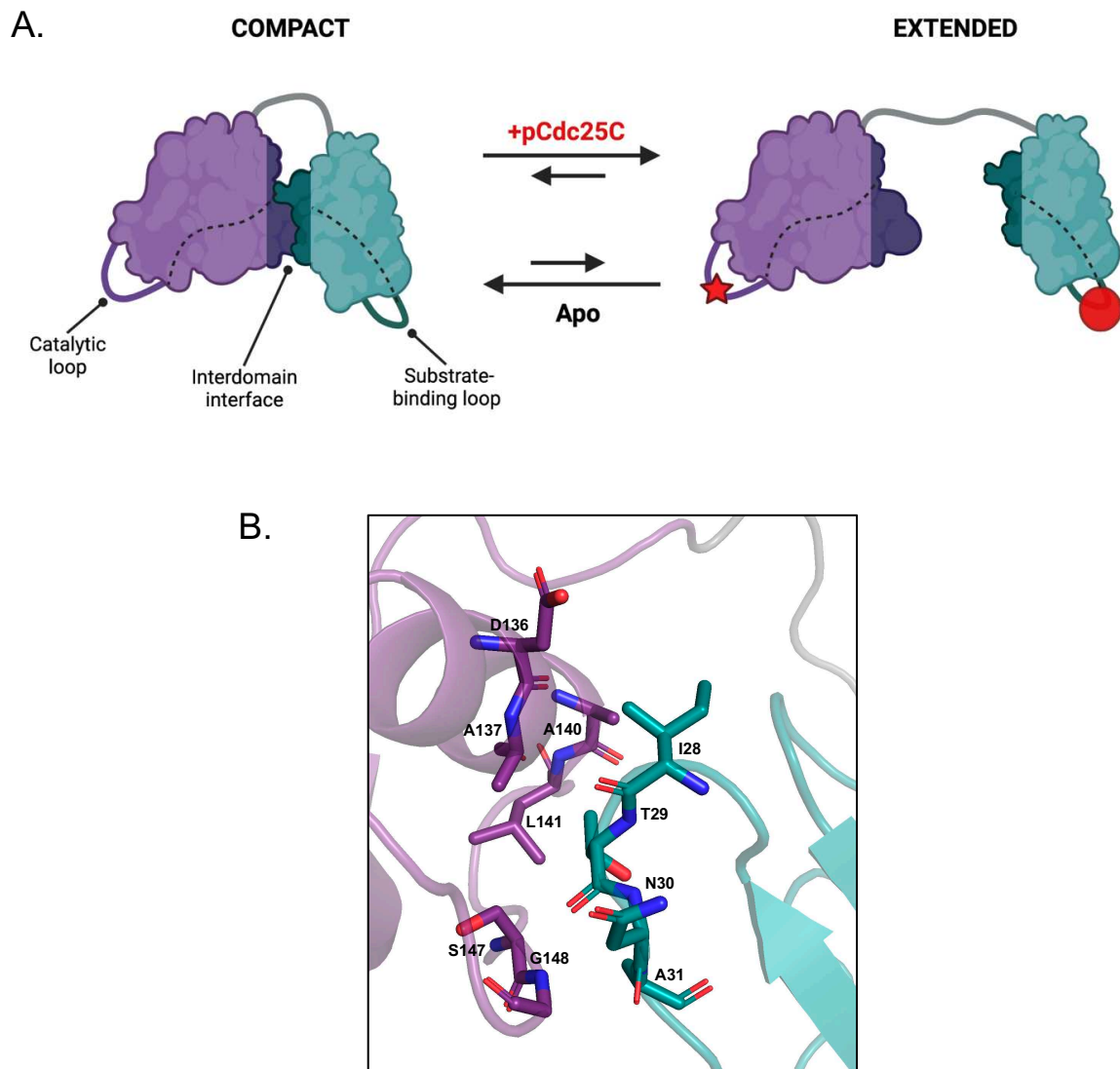


Figure 22: Schematic representation of the two states of Pin1 (A) and the residues involved in transient interdomain contacts (B). **A.** Apo Pin1 populates two states in solution, a compact state whereby the interdomain interface is well-defined, and the extended state which displays a loss of interdomain contact. The binding of different substrates can promote the stabilisation of one state over the other (e.g. binding of pCdc25C promotes the extended conformation). This suggests the presence of an allosteric pathway (dashed lines) which allows the propagation of a signal from substrate binding (red circle) in the WW domain to the distal catalytic loop (red star). **B.** Residues Ile28/Thr29/Asn30/Ala31 of loop 2 of the WW domain, residues Asn136/Ala137/Ala140/Leu141 of the α 4 helix and Ser147/Gly148 of the β 6-strand of the PPIase domain form transient interdomain contacts in the compact state of Pin1.

Interestingly, the binding of different substrate sequences can have antagonistic effects on the interdomain flexibility of Pin1. For example, the binding of the WW domain to pCdc25C, a

short peptide sequence mimicking the pT48-P49 site of Cdc25C (EQPLpTPVTDL), weakens interdomain contacts and promotes adoption of the extended state (Jacobs *et al*, 2003). On the other hand, no clear interdomain rearrangement occurs upon the binding of the WW domain to an optimal Pin1 peptide sequence known as the pintide (WFYpSPR) (Jacobs *et al*, 2003; Born *et al*, 2022). The transient contact between the PPIase and WW domain can hence be modulated by substrate binding, illustrating a direct method by which interdomain allostery in Pin1 can be described.

With accumulating evidence suggesting the presence of dynamic allostery within Pin1, the question arises as to exactly how this interdomain communication is transmitted between two distal sites. NMR deuterium (^2D) and ^{13}C relaxation studies probed methyl-bearing residues within Pin1 and identified a decrease in sidechain flexibility and an associated reduction in intrinsic conformational exchange processes within the PPIase domain upon pCdc25C binding to the WW domain (Namanja *et al*, 2011, 2007). Significantly, it was noted that the regions displaying altered sidechain mobility upon substrate binding are not random, but in fact form a continuous conduit of hydrophobic residues spanning from the active site to the interdomain cleft (see figure 23a). This conduit is comprised of residues from three distinct regions, firstly of the ID interface: Ile28 of loop 2 of the WW domain and residues Leu141 of the $\alpha 4$ -helix and Val150 of the $\beta 6$ -strand of the PPIase domain. Additional residues of the conduit, Ile93, Ile96 and Leu106 belong to the $\alpha 1$ -helix of the PPIase domain. The hydrophobic path terminates at a conserved hydrophobic cluster, Leu60/Leu61/Val62, buried in the PPIase domain core, next to the active site. The presence of this extended hydrophobic path could help explain how a signal from the interdomain interface could be propagated upon ligand binding to affect functional groups in the distal active site, which is located approximately 12 Å away. Many of the residues of this conduit are conserved across Pin1 from various species, emphasising their presumed importance for Pin1 function (see figure 23b). Complementary experiments focused on binding of *cis*- and *trans*-locked ligands, as well as their parent compound, pintide, to Pin1 revealed a similar modification of sidechain mobility throughout this conduit, confirming that such dynamic responses are due to substrate binding as opposed to *cis/trans* isomerisation of the bound peptide (Namanja *et al*, 2011). This study also indicated that the exact residues participating in the conduit can vary for different substrate sequences. For instance, the binding of the pCdc25C peptide to the WW domain induced greater rigidity of the conserved hydrophobic cluster (Leu60/Leu61/Val62) of the active site, whereas pintide binding provoked

greater stiffening of residues of the substrate-binding pocket of the PPIase domain (Leu122/Ala124/Met130) (see figure 23a). Additionally, the extent of the dynamic response to substrate binding was dependent not only on the substrate sequence, but also on the conformation of the pS/T-P moiety, with the *cis*-locked inhibitor inducing a more significant conduit response (Namanja *et al*, 2007, 2011).

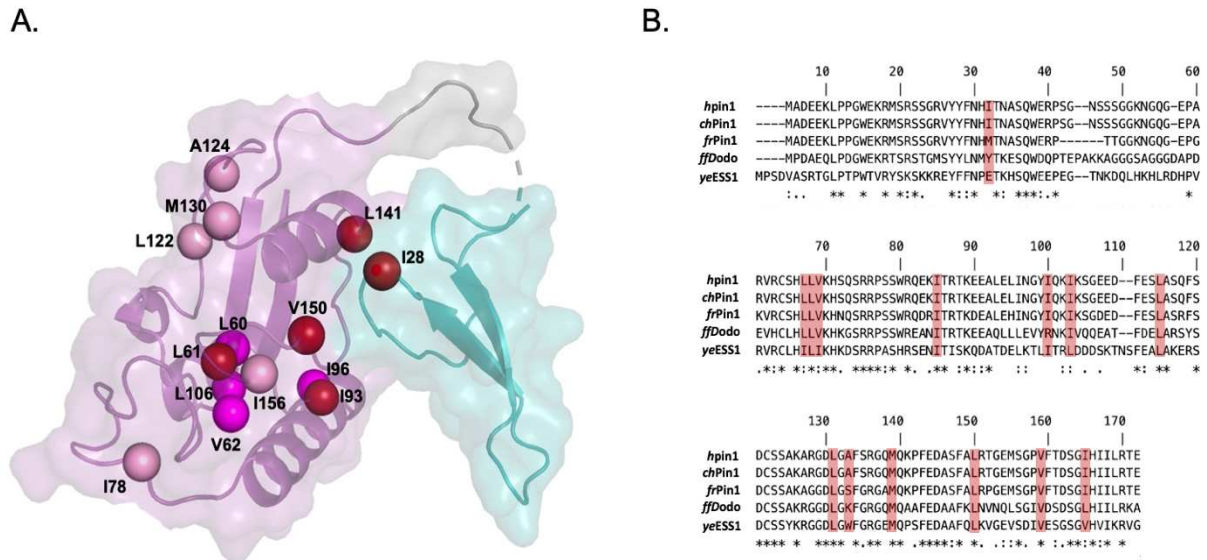


Figure 23: The proposed allosteric hydrophobic conduit of Pin1. **A.** Residues showing increased rigidity upon substrate binding to the WW domain are represented as spheres (light pink: only in pintide conduit, magenta: only in Cdc25C conduit and red: present in both conduits.) **B.** Multiple sequence alignment of Pin1 homologs from various species (hPin1: human, chPin1: chimpanzee, frPin1: Western clawed frog, ffDodo: fruit fly and yeESS1: Brewer's yeast). The residues of the proposed hydrophobic conduit are highlighted in red, displaying an overall high level of conservation.

Further evidence scrutinising the importance of interdomain communication came in the form of mutagenesis studies of residues of the proposed conserved hydrophobic conduit. A significant member of this conduit, I28, located at the contact interface between the WW and PPIase domains, has been established as critical for transmission of interdomain communication (Namanja *et al*, 2007; Wilson *et al*, 2013). The I28A mutant displayed weakened interdomain contacts and an associated increase in isomerase activity ($k_{tc+ct}^{WT} = 31.3 \pm 0.5 \text{ s}^{-1}$ vs $k_{tc+ct}^{I28A} = 73 \pm 2 \text{ s}^{-1}$) (Wilson *et al*, 2013). Notably, a reduction in WW domain binding affinity for the pCdc25C peptide was also observed at the residue level, with K_D^{WT} values of 3-9 μM and K_D^{I28A} values of 43-48 μM , representing a 5-10-fold affinity

decrease. Considering that neither Ile28 nor loop 2 of the WW domain, to which it belongs, directly contact the substrate, the decrease in binding affinity suggested the presence of intradomain allostery connecting the interdomain interface to the substrate binding loop of the WW domain. Further studies of Pin1 mutants separately perturbing interdomain contact, PPIase activity and WW domain substrate binding also identified an associated increase in PPIase activity and related disruption in interdomain contact in response to pCdc25C binding to the WW domain, confirming the apparent negative allosteric mechanism in Pin1 triggered by interdomain contact (Wang *et al*, 2015). The enhancement of PPIase activity upon loss of interdomain contact in the extended conformation of Pin1 can be explained by the fact that the PPIase and WW domains behave as separate entities in this state (Peng, 2015). It was calculated that, for the pCdc25C substrate, the catalytic activity of the isolated PPIase domain was higher than that of the full-length protein by a factor of ~ 4.5 (Namanja *et al*, 2011; Peng, 2015). Therefore, it is theorised that upon pCdc25C binding to the WW domain, the associated increase in interdomain flexibility allows the catalytic pocket to sample conformations that are more similar to those of the isolated PPIase domain, leading to higher PPIase activity (Peng, 2015).

The summary of these results indicated that the presence of both intradomain dynamic allostery within both domains of Pin1, as well as interdomain conformational rearrangements, help to propagate a signal from the WW domain substrate binding loop to the distal active site of the PPIase domain (Peng, 2015). The creation of a suitable model to incorporate this inherent allostery into the mechanism of Pin1 substrate binding and catalysis is necessary for complete understanding of its internal mechanisms and to guide therapeutic strategies. A process of conformational selection has been proposed to explain the existence of both the extended and compact forms in apo and ligand-bound Pin1 (Peng, 2015; Born *et al*, 2022). In terms of the structures of these states, Pin1 is consistently observed in the compact state within all resolved crystal structures. To unveil the intricate structural interplay between Pin1's compact and extended states in solution, recent studies utilised advanced NMR and electron paramagnetic resonance (EPR) techniques (Born *et al*, 2019, 2022). In particular, as the two states are in fast exchange, traditional structural techniques might struggle to distinguish and fully describe both states simultaneously. The study used exact Nuclear Overhauser Effect (eNOE) measurements, allowing for remarkably accurate proton-proton distance determinations, even at distances of up to 8 Å (Born *et al*, 2019). Measuring highly accurate distances between protons within the

protein's domains (intradomain distances) and between protons from different domains (interdomain distances) was instrumental in distinguishing and characterising the compact and extended states of Pin1. Additionally, Residual Dipolar Couplings (RDCs) were employed to discern the relative orientations of Pin1's two domains. To extend the scope of distance restraints, paramagnetic spin labels were introduced into different regions of Pin1. The subsequent use of Paramagnetic Relaxation Enhancement (PRE) and Double Electron-Electron Resonance (DEER) techniques enabled the probing of distances spanning 25-80 Å between spin labels. By using this combined approach, the authors could determine the structure of the compact and extended states at atomic resolution (see figure 24). The resolution of pCdc25C- and pintide-bound solution structures of Pin1 in the same way revealed that the binding of different substrate sequences, which can have differential effects on interdomain contacts, can alter the equilibrium between the two states. The significant interdomain reorientation in response to certain ligand binding is seemingly driven by slight reorientation of hydrophobic residues at the interdomain interface of each domain (Born *et al*, 2022) (see figure 24).

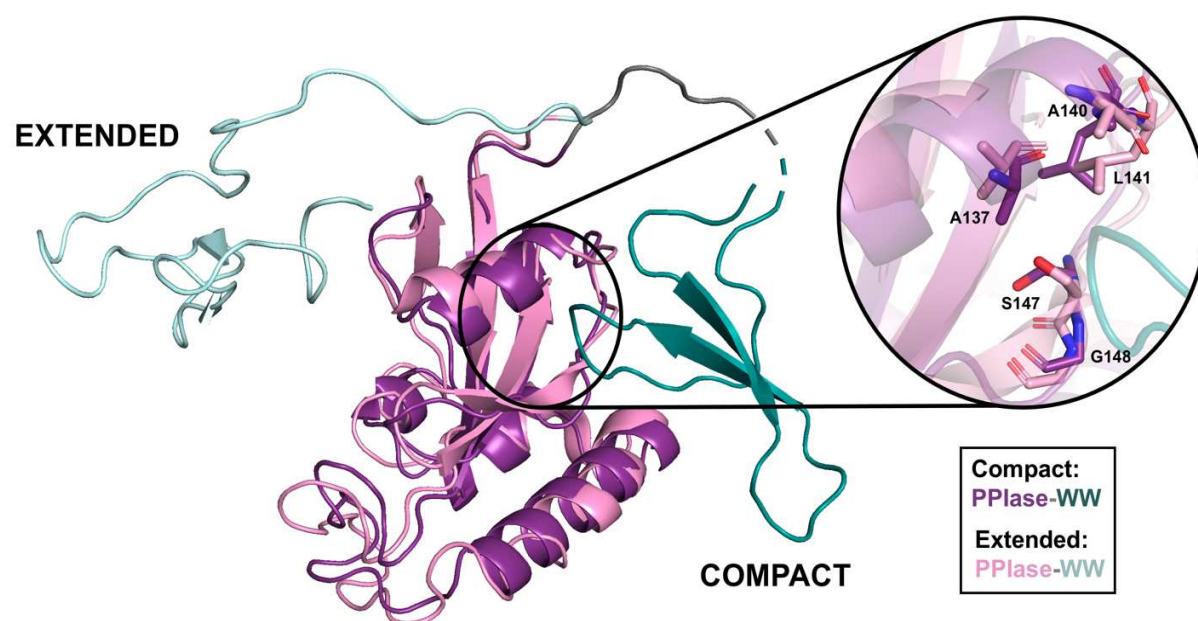


Figure 24: The structure of the compact and extended states of Pin1. The solution structure of the extended state (PDB ID: 7SA5) compared to the crystal structure of the compact state of Pin1 (PDB ID: 1PIN) reveals an overall different organisation of the two domains. In particular, hydrophobic residues of the PPIase domain ID interface are oriented towards the ID interface in the compact form (purple) and towards the PPIase domain in the extended state (pale pink). Diagram made using PyMol.

These hydrophobic residues include key residues of the PPIase which mediate ID contacts: Ala137/Ala140/Leu141 of the α 4 helix and Gly148 of the β 6-strand of the PPIase domain. Notably their respective hydrophobic sidechains are oriented towards the interdomain interface in the compact state, but towards the core of the PPIase domain in the extended state of apo- and substrate-bound Pin1. Hence, it is hypothesised that in the compact state, these residues make interdomain contacts, however, their subsequent occlusion from the interface in response to pCdc25C binding allows the stabilisation of the extended state via the hydrophobic effect. Since it was suggested that the level of interdomain contact has a direct effect on catalysis, this indicated the presence of an intradomain allosteric signalling mechanism within the PPIase domain (Peng, 2015; Wang *et al*, 2015).

Molecular dynamics studies have indicated that the proposed hydrophobic conduit within the PPIase domain cannot describe the full picture as it is only representative of methyl-bearing sidechains (Guo *et al*, 2015). Instead, these studies have put forward the idea that two pathways mediate the allosteric communication between the two domains (figure 25). The first pathway involves the propagation of a signal from the back of the WW domain, through the interdomain interface and the PPIase domain core β -sheets to two key loops neighbouring the active site (β 5- α 4, residues 142-145, and β 6- β 7, residues 152-154). On the other hand, path 2 begins at the front of the WW domain and signal propagation primarily occurs via the α 1 helix to the catalytic loop (residues 63-80) of the PPIase domain. It was first assumed that both pathways are essential for signal transmission, however, recent studies have suggested that different substrates can selectively utilise one of the pathways dependent on their differential effects on interdomain contacts (Born *et al*, 2022). As interdomain contact is disrupted upon the binding of pCdc25C, the α 1 helix of the PPIase domain is inaccessible, and hence the signal propagates via path 1, through the core β -sheet element of the PPIase domain (see figure 25). Alternatively, as the interdomain interface is maintained upon pintide binding to the WW domain, the substrate can interact directly with the α 1 helix to propagate a signal from the WW domain substrate binding pocket to the catalytic loop via the activation of path 2 (see figure 25).

This model of interdomain allostery could help shed light on the observed binding promiscuity of Pin1, as the binding of different substrate sequences provokes distinct changes in the dynamics of Pin1, offering an explanation as to how Pin1 is able to differentially regulate such

a large interactome (Born *et al.*, 2022). Additionally, since many Pin1 recognition motifs (pSer/Thr-Pro modules) occur in intrinsically disordered regions of their substrates, the two-state model's inherent interdomain dynamics facilitate adaptive binding to such flexible ligands (Bouchard *et al.*, 2018). Interdomain mobility could also facilitate the simultaneous binding of each domain of Pin1 to substrates containing more than one phosphorylated motif (for instance in the case of Cdc25C which contains six Ser/Thr-Pro motifs in its *N*-terminal regulatory region). Thus, incorporating interdomain allostery into Pin1's binding mode provides additional explanations for substrate interaction. Further studies of Pin1 with substrates, including full-length proteins containing multiple motifs, should reveal more about the binding mechanism. In particular, it would be interesting to see if substrate binding to the PPIase domain reciprocally modulates the binding to the WW domain via an allosteric mechanism.

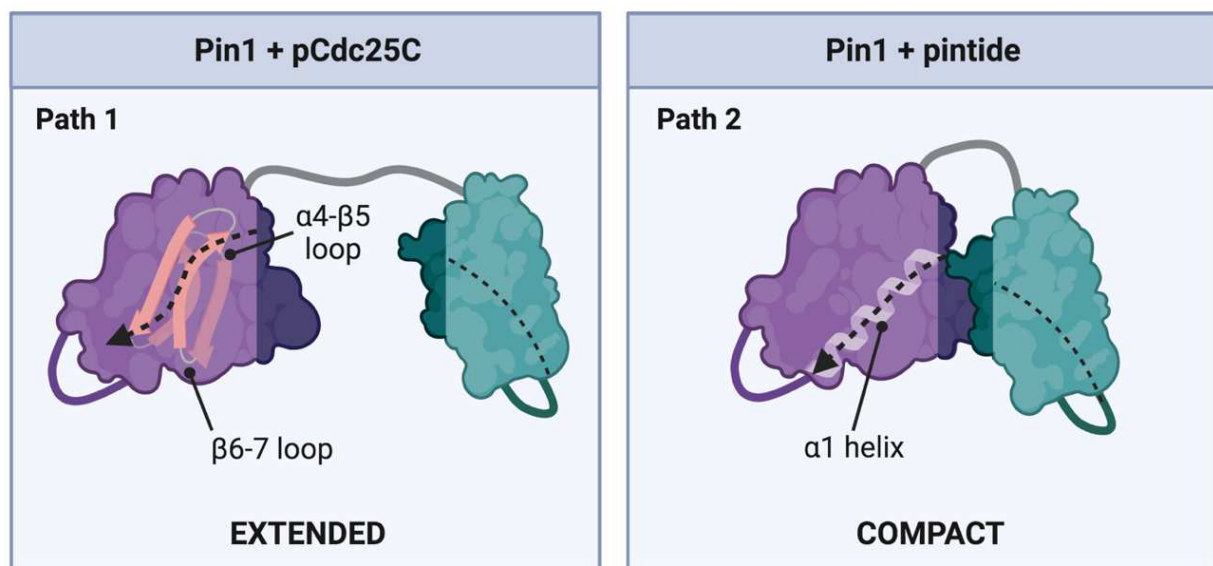


Figure 25: Schematic representation of the effect of substrate sequence on inter- and intradomain allosteric signalling within Pin1. Upon pCdc25C binding (left box), interdomain contact is disrupted and the extended form of Pin1 is stabilised. In this state, as the $\alpha 1$ helix is inaccessible to the substrate, the signal is propagated via the core β -sheet of the PPIase domain (path 1). Upon pintide binding, interdomain contact is maintained, hence the compact form of Pin1 is populated. The maintenance of ID contacts allows the activation of path 2, as the substrate can directly interact with the $\alpha 1$ helix.

3.6 The interaction between Pin1 and Cdc25C

To better understand the molecular interaction between Pin1 and its substrates, further studies are required to explore the specific mechanisms and consequences of Pin1's binding with its various target proteins. In particular, it is unclear how Pin1 recognises and interacts with protein partners containing multiple pS/T-P binding motifs. One significant binding partner of Pin1 is the mitotic protein Cdc25C, which contains multiple Pin1 recognition motifs within its IDR. However, despite being a well-studied Pin1 substrate, the details of this interaction and its regulatory consequences on Cdc25C activity remain largely unexplored.

As formerly described, Pin1 activity is heavily intertwined with cell cycle progression, with many key regulators of each stage of the cell cycle identified as Pin1 binding partners. Amongst these, the Cdc25C phosphatase is a notorious substrate of Pin1 (Shen *et al*, 1998). The IDR of Cdc25C contains six Ser/Thr-Pro motifs, all of which could be phosphorylated to form potential Pin1 substrates (T48, T67, S122, T130, S168 and S214) (Liu *et al*, 2020). However, as of yet, there remains little convincing evidence confirming which of these motifs Pin1 interacts with. Although no structural evidence has been released to help characterise the interaction between Pin1 and Cdc25C, it is widely accepted that Pin1 interacts exclusively with mitotic Cdc25C, whereby it is known to be phosphorylated at five key sites (T48, T67, S122, T130 and S214) (Sur & Agrawal, 2016). To further investigate which phosphorylated sites are important for Pin1 binding, biochemical and mutagenesis studies have been performed. *In vitro* studies demonstrated that a quintuple mutant of *Xenopus* Cdc25C, where the three threonine residues were substituted with valine and the two serine residues were replaced with alanine (T3S2), resulted in complete inhibition of mitotic phosphorylation and Pin1 binding (Izumi & Maller, 1993; Shen *et al*, 1998). Moreover, a triple mutant targeting only the threonine residues (T3) also exhibited significantly diminished binding to Pin1, suggesting these residues were essential for the interaction. Based on comprehensive peptide scanning analyses, it was determined that the primary Pin1 binding sites within Cdc25C are pT48-P49 and pT67-P68 (Lu *et al*, 1999b). This evidence was further supported by the fact that a Cdc25C^{T48V,T67V} mutant did not appear to interact with Pin1 *in vitro* (Zhou *et al*, 2000). However, it is crucial to note that the Cdc25C^{T48V,T67V} mutant used in this study did not exhibit a clear molecular weight shift characteristic of 'complete' phosphorylation (as seen in the WT protein) upon

phosphorylation using *Xenopus* mitotic extract. The lack of this characteristic shift raises questions about the phosphorylation status of the remaining S/T-P sites, including T130, S122, and S214. It is worth emphasising that the assessment of the phosphorylation status of Cdc25C and its mutants in this study relied solely on SDS-PAGE gel shifts, a method that lacks the resolution to pinpoint individual phosphorylation events. Therefore, the observed lack of interaction of Pin1 with this mutant could potentially be explained by incomplete phosphorylation at the remaining S/T-P sites, indicating that the interaction is potentially not solely dependent T48 and T67.

From the structural point of view, the details of the interaction between Pin1 and Cdc25C, specifically the binding mode of Pin1, also remain unclear. Much of the research concerning the Pin1-Cdc25C interaction has been centred around investigations of the binding of a short pCdc25C peptide (EQPLpTPVTDL), which is a derivative of the pT48-P49 site of *Xenopus* Cdc25C (Born *et al*, 2022; Jacobs *et al*, 2003; Namanja *et al*, 2007, 2011; Peng, 2015; Wang *et al*, 2015). It is important to highlight that such studies primarily focusing on the short peptide sequence, which contains only a single phosphorylated motif (pT-P), are likely to overlook critical aspects of the interaction. This is particularly significant when considering that the full-length IDR of Cdc25C harbours six potential Pin1 binding motifs. Nevertheless, structural studies using the pCdc25C peptide revealed that the WW domain of Pin1 acts as the principle binding site for pCdc25C (Jacobs *et al*, 2003; Verdecia *et al*, 2000). This was confirmed by tracking ^1H - ^{15}N amide resonance shifts in 2D NMR spectra upon titration with the pCdc25C peptide into full-length Pin1 and its isolated domains. The peptide displayed no binding to the isolated PPIase domain, while the isolated WW domain and full-length protein displayed binding ($K_D = 10 \mu\text{M}$). Specifically, residues of the phosphate-binding loop (16-20) and proline binding site (33-35) of the WW domain displayed binding-induced chemical shift changes upon binding of the isolated WW domain of FL Pin1 to pCdc25C. In full-length Pin1, binding to pCdc25C additionally induced changes in residues of the α 4-helix (132-140), residues of the β 4-sheet (56-61) and residues of the α 4- β 6 loop (residues 152-154). Since the pCdc25C peptide does not appear to bind to the isolated PPIase domain, these effects are postulated to be an allosteric consequence of pCdc25C binding to the WW domain. This is supported by many studies which have characterised that pCdc25C binding to the WW domain reduces interdomain contacts and promotes the population of the extended state of Pin1 (as described in detail in section 3.5) (Born *et al*, 2022). While no detailed, residue-level information is

available regarding the interaction between Pin1 and full-length Cdc25C, pull-down assays of mitotic cell lysates using the full-length and isolated domains of Pin1 have similarly confirmed that the WW domain is the predominant binding module for mitotic Cdc25C (Lu *et al*, 1999b).

Moreover, the regulatory effect of Pin1 binding on Cdc25C activity is complex and poorly understood. Initially, it was believed that Pin1 isomerisation provokes an inhibitory effect on Cdc25C by promoting its dephosphorylation by PP2A (see figure 26) (Hutchins & Clarke, 2004). Cdc25C is a physiological substrate of the PP2A phosphatase, which negatively regulates Cdc25C activity throughout interphase and the mitotic exit via its dephosphorylation of residue Thr130 (Forester *et al*, 2007). PP2A has been shown to display *trans*-specific Pro-directed phosphatase activity (Zhou *et al*, 2000). Hence it was theorised that the *cis-trans* isomerisation by Pin1 affects the recognition of Cdc25C by PP2A, confirmed by the fact that Pin1 isomerisation facilitates the dephosphorylation of Cdc25C by PP2A *in vitro*.

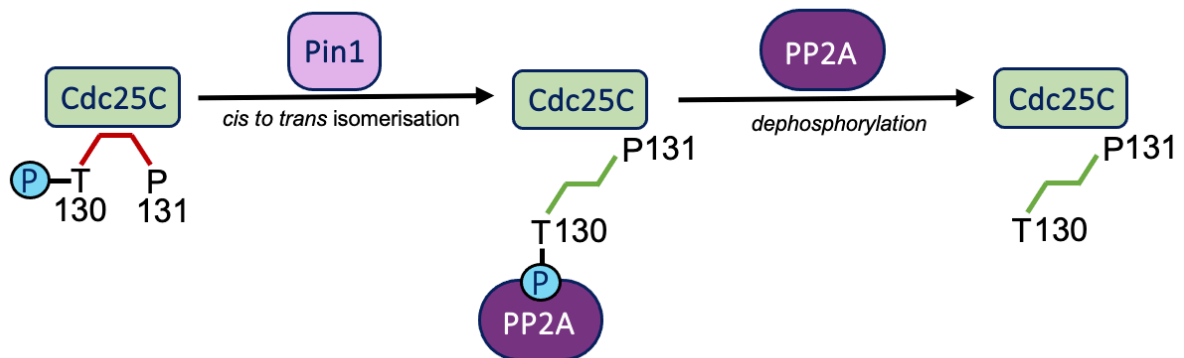


Figure 26: Schematic representation of the proposed PP2A hypothesis of Pin1 regulation of Cdc25C. Phosphorylation of Thr130 contributes to the active form of Cdc25C by promoting dissociation from 14-3-3 proteins (see chapter 2.3). The *cis* conformation of pThr130-Pro131 (red) is not recognised by PP2A. It is proposed that Pin1 can isomerise, in the direction *cis* to *trans*, this phosphorylated motif. Then, PP2A can recognise the *trans* pThr130-Pro131 motif (green) and catalyse its dephosphorylation. The removal of this activating phosphorylation is hypothesised to inhibit Cdc25C (perhaps by promoting binding to 14-3-3 proteins).

The apparent inhibition of Cdc25C phosphatase activity upon regulation by Pin1 may initially seem counterintuitive given that the two proteins interact during mitosis when the activity of

Cdc25C is at its peak. However, subsequent studies have shed light on an alternative perspective, indicating that Pin1 can also positively regulate Cdc25C activity (Stukenberg & Kirschner, 2001). When recombinant *Xenopus* Cdc25C was phosphorylated *in vitro* by the CDK1 kinase, Pin1 isomerisation appeared to have an inhibitory effect on Cdc25C phosphatase activity, corroborating previous results. However, when Cdc25C was phosphorylated by both CDK1 and Plx1 *in vitro*, an increased phosphatase activity was observed in response to Pin1 isomerisation. It has been previously stated that both CDK1 and Plk1 phosphorylate on Cdc25C during mitosis, hence the activation of Cdc25C by Pin1 perhaps displays a more accurate picture of what occurs during the mitotic transition. Although this study did not analyse which sites of the Cdc25C sample were phosphorylated by CDK1 and Plx1, it has been previously noted that CDK1 is responsible for the hyperphosphorylation of S/T-P sites and Plk1 appears to phosphorylate Cdc25C at residue S198 of its NES.

A potential biological implication for the ability of Pin1 to both activate and inactivate Cdc25C dependent on its phosphorylation status could be to further fine-tune the entry to mitosis, contributing to the gradual and complete activation of CDK1-Cyclin B by Cdc25C at the G₂-M transition. The activation of CDK1-Cyclin B at the mitotic transition habitually displays a lag period followed by rapid activation which quickly reaches a peak, marking entry to mitosis (see figure 27) (Stukenberg & Kirschner, 2001). However, in cells depleted of Pin1, a similar lag period is observed but the subsequent complete activation of CDK1-Cyclin B is much slower (see figure 27). One potential explanation for this could be that during the lag period prior to the mitotic transition, Cdc25C is only phosphorylated by CDK1, hence Pin1 maintains Cdc25C in an inhibited state as to not fully activate CDK1-Cyclin B too early. Upon the rapid transition to mitosis, where both CDK1 and Plk1 phosphorylate Cdc25C, Pin1 isomerisation activates Cdc25C, allowing the full and rapid activation of CDK1-CyclinB. This hypothesis could explain why, in the absence of Pin1, CDK1 activation is much slower, and incomplete. Overall, the consequence on the Pin1-Cdc25C interaction on Cdc25C activity is seemingly dependent on the phosphorylation status of Cdc25C, adding further complexity to the effect of multi-site phosphorylation on Pin1 regulation which warrants further investigation. The remarkable ability of Pin1 to both activate and inactivate Cdc25C is likely crucial to its role as a key regulator of the mitotic transition. However, the full investigation of the Pin1-Cdc25C interaction has been limited by the lack of structural information on the *N*-terminal regulatory region of Cdc25C.

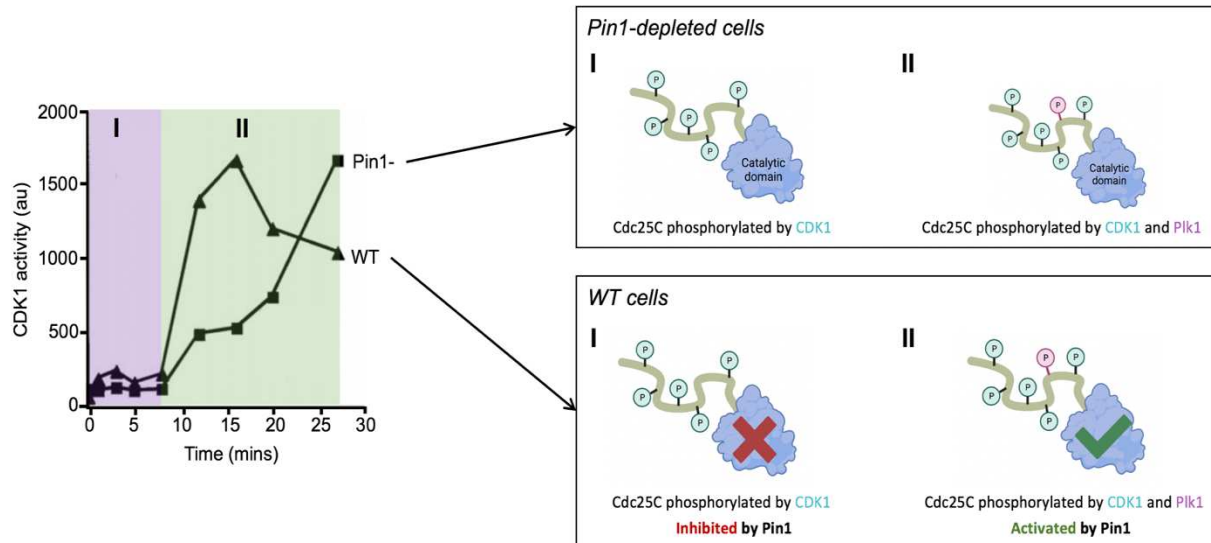


Figure 27: A model for Cdc25C regulation by Pin1 based on the activity of CDK1 *in vivo* in the presence and absence of Pin1. The activity of CDK1 can be used as an indicator of the effect of Pin1 depletion on Cdc25C activity *in vivo* since Cdc25C is its predominant activator. These results can help to build a hypothesis of Pin1-mediated Cdc25C activation. In the presence of Pin1 (WT cells), Cdc25C activity is, at first, inhibited when solely phosphorylated by CDK1. However, upon additional phosphorylation by Plk1, Pin1 activates Cdc25C rapidly and completely. In the absence of Pin1 this activation is much slower.

Chapter 4 Pin1 as a Therapeutic Target

Due to its crucial role in regulating a variety of essential cellular processes, it comes as no surprise that Pin1 dysregulation can have severe pathological consequences. Extensive research has highlighted Pin1's involvement in diseases such as Alzheimer's and cancer and it has since emerged as a promising therapeutic target. The development of Pin1 inhibitors has gained substantial attention, offering new avenues for therapeutic interventions. This section will focus on the diverse therapeutic implications of Pin1, exploring its association with Alzheimer's disease, its involvement in cancer progression and the classes of Pin1 inhibitors currently under investigation. By elucidating Pin1's therapeutic potential and the challenges associated with its inhibition, this chapter aims to highlight the crucial importance of fully understanding the binding mechanism of Pin1, as it is a fundamental prerequisite for the development of effective inhibitors.

4.1 Dysregulation of Pin1 in disease

The effect of Pin1 dysregulation in human cells can manifest as a variety of pathogenic phenotypes, including cancer, cardiovascular disease, metabolic disease, neurodegenerative disease and even increased susceptibility to viral infection (Li *et al*, 2021). In particular, the involvement of Pin1 has garnered significant attention in two pivotal areas of research: cancer, as one of the leading causes of mortality worldwide, and Alzheimer's disease, a rapidly growing concern in an aging population.

4.1.1. The involvement of Pin1 in Alzheimer's disease

The involvement of Pin1 in the development of Alzheimer's disease (AD) is relatively well-characterised, with its downregulation strongly associated with an increased likelihood of developing the disease. Pin1 plays a significant role in healthy aging, as evidenced by studies on Pin1 knockout (KO) mice (Liou *et al*, 2002, 2003). These mice exhibit a relatively normal phenotype until later stages of life when they manifest signs of premature aging, including neurodegeneration. When analysed further, these mice more specifically show neuropathological signs of AD (Liou *et al*, 2003). The apparent protective role of Pin1 against

development of AD is twofold, as it appears to safeguard against the formation of both A β plaques and neurofibrillary tangles (NFTs), the two key pathological hallmarks of AD (summarised in figure 28).

The deposition of insoluble amyloid fibrils as extracellular plaques stands as the principal and most extensively studied characteristic of the brains of patients with AD (Serpell, 1999). The A β peptide, which constitutes these fibrils, is generated through the cleavage of a larger protein known as the amyloid precursor protein (APP). APP is a type I integral membrane protein expressed in neuronal cells (Zhang *et al*, 2012a). Although necessary for proper neuronal development, the precise physiological role of APP is not well-understood. Nevertheless, the post-translational processing of APP carries significant implications for the health of neuronal cells, with Pin1 emerging as a pivotal regulator in this process (Bianchi & Manco, 2018). Pin1 promotes the non-amyloidogenic processing pathway, which leads to the production of shorter and less toxic peptides. In the absence of Pin1 or its reduced activity, the amyloidogenic pathway becomes favoured, ultimately resulting in the accumulation of A β plaques. The processing of APP is regulated via a series of phosphorylation events, where hyperphosphorylation serves as a defining characteristic of the pathogenic pathway (Driver *et al*, 2015). Phosphorylation at one particular site, Thr668-Pro, can have significant impacts on the outcome of APP processing, dependent upon whether it adopts the *cis* or *trans* conformation. The build-up of *cis*-pT688-P leads to APP accumulation and toxic A β production. Significantly, through the *cis* to *trans* isomerisation of phospho-Thr668-Pro, Pin1 helps to prevent the accumulation of the *cis* conformation and hence promotes the non-amyloidogenic pathway (Pastorino *et al*, 2006). The second neuropathological hallmark of AD is the formation of intracellular NFTs composed of the hyperphosphorylated tau protein. In contrast to APP, the physiological role of tau is well-defined, forming a key component of microtubules in neuronal scaffolds. However, upon hyperphosphorylation of Ser/Thr-Pro motifs, the tau protein has a propensity to aggregate and form neurofibrillary tangles, resulting in impaired neuronal function (Driver *et al*, 2015). It is now understood that, similarly to APP, the accumulation of such phosphorylated motifs in the *cis* conformation precedes tau filament formation (Nakamura *et al*, 2012). Pin1 is seen to specifically interact with the phosphorylated Thr231-Pro motif of tau and catalyse its isomerism from the *cis* to *trans* conformation (Lu *et al*, 1999a) This process plays a crucial role in the restoration of tau function and facilitates its

dephosphorylation by the *trans*-specific phosphatase, PP2A (Liou *et al*, 2003; Lu *et al*, 1999a; Zhou *et al*, 2000).

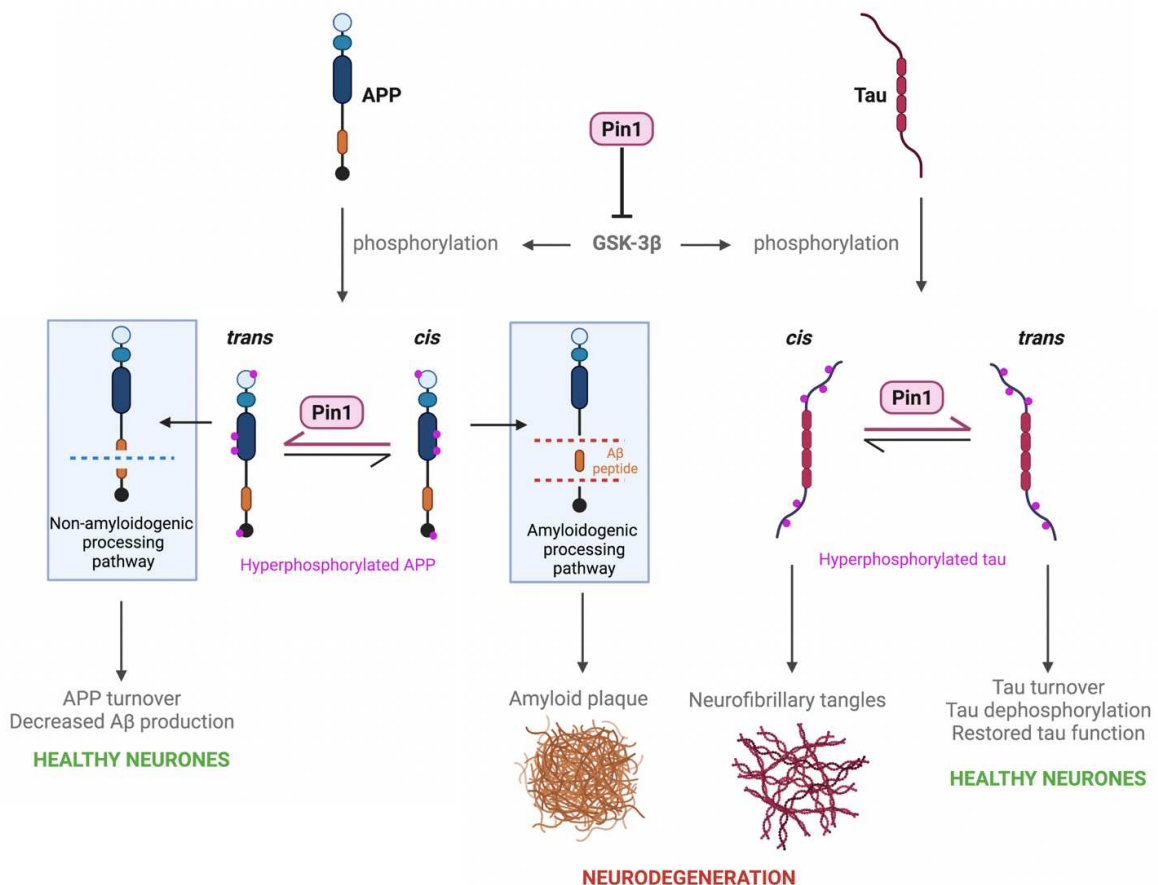


Figure 28: The protective role of Pin1 against the development of Alzheimer's disease. Pin1 has been seen to promote neuronal health by reducing APP and tau hyperphosphorylation (via its direct inhibition of the GSK β kinase), and by inducing a conformational change of the *cis* phospho-forms of APP/tau to the *trans* structure. The *trans* conformers of hyperphosphorylated tau/APP are functional. In the absence of Pin1, the pathogenic *cis* conformers of hyperphosphorylated tau and APP would accumulate. For APP, this provokes amyloidogenic post-translational processing, leading to the production of toxic AB peptides which aggregate to form amyloid plaques. For tau, the hyperphosphorylated *cis* form aggregates to form NFTs. Both NFTs and amyloid plaques contribute to neurodegeneration observed in AD.

Furthermore, Pin1 is known to interact with a crucial kinase involved in the characteristic phosphorylation of both tau and APP, a process that precedes the accumulation of their toxic *cis*-conformers. Glycogen synthase kinase-3 (GSK- β 3) is prominently expressed in neuronal

cells and plays a crucial role in neuronal cell growth (Castaño *et al*, 2010). Notably, it functions as one of the key kinases involved in the phosphorylation of tau and APP, and its overactivity has hence been strongly associated with elevated tau phosphorylation and increased production of toxic A β (Aplin *et al*, 2002; Flaherty *et al*, 2000; Ma *et al*, 2012). Pin1 is seen to directly interact with GSK- β 3 and negatively regulate its activity through isomerisation of its Thr330-Pro motif (Ma *et al*, 2012).

Overall, the protective role of Pin1 in AD holds great significance as prevents the accumulation of the pathogenic *cis* species of both phosphorylated tau and APP, which are associated with neurodegeneration. Pin1 is proposed to achieve this by directly isomerising the toxic *cis* forms of these proteins to their functional *trans* structure, restoring their normal cellular function. Additionally, Pin1 inhibits the key kinase, GSK- β 3, involved in the prior phosphorylation of tau and APP. These combined actions highlight the significant impact of Pin1 in counteracting the pathological processes underlying AD, making it an intriguing target for potential therapeutic interventions.

4.1.2. The overexpression of Pin1 in cancer

In striking contrast to its characteristic downregulation in Alzheimer's disease, Pin1 dysregulation prominently manifests in cancer, where it is frequently overexpressed and strongly associated with poor prognosis. (Li *et al*, 2021). The overexpression or overactivation of Pin1 has been implicated in multiple cancer types, including breast, prostate, gastric, oesophageal, lung, oral, cervical, brain, colon and melanoma (Ai *et al*, 2009; Bao *et al*, 2004; Cheng *et al*, 2016; Jin *et al*, 2011; Leung *et al*, 2009; Shi *et al*, 2015; Wulf, 2001; Zhu *et al*, 2016). The link between Pin1 activity and cancer progression was further established by *in vitro* and *in vivo* studies of Pin1 knockout (KO) mutants, whereby Pin1 abolition inhibited tumorigenesis and promoted apoptosis (Wulf *et al*, 2004).

The mechanism by which Pin1 exacerbates malignant phenotypes is complex, with the enzyme being known to activate over 50 oncogenes and inhibit over 20 tumour suppressors (Chen *et al*, 2018). Particularly, it has been highlighted that Pin1 plays a role in all 10 characteristic hallmarks of cancer which were summarised by Hanahan and Weinberg as sustained proliferative signalling, evasion of growth suppressors, enabling replicative immortality, tissue

invasion and metastasis, sustained angiogenesis, evasion of apoptosis, inflammation, chromosomal instability, metabolic dysregulation and evasion of the immune system (Chen *et al*, 2018; Hanahan & Weinberg, 2000, 2011) (see figure 29).

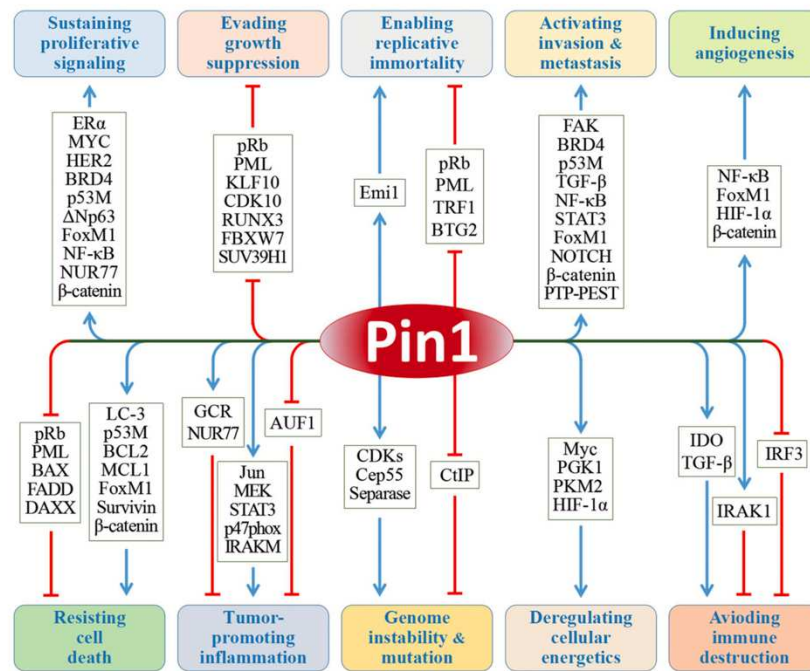


Figure 29: The involvement of Pin1 in promoting the 10 hallmarks of cancer. Summary of the multiple contributions of Pin1 in activation (blue arrow) of oncogenes or inhibition (red arrows) of key tumour suppressors which influence cancer-driving pathways. (Figure taken from Chen *et al*, 2018).

The complexity of the network of cancer-driving pathways in which Pin1 is involved comes as no surprise considering its extensive interactome. However, the details of these specific processes are not the focus of this thesis and have been extensively reviewed elsewhere (see Chen *et al*, 2018). Nevertheless, as an example, Pin1's significance in breast cancer (BC) has been well-documented, establishing its crucial role in BC tumour development and progression (Rustighi *et al*, 2017). Notably, Pin1 emerges as a key player in promoting proliferative signalling, which is unsurprising considering its central involvement in regulating cell cycle progression. Pin1 displays comparable levels as Cyclin D in BC cells since Pin1 is seen to upregulate its expression and increase its cellular stability, thereby contributing to enhanced proliferation (Lu & Hunter, 2014; Rustighi *et al*, 2017). It is suggested that one mechanism through which Pin1 increases Cyclin D levels involves a direct interaction with the

pThr286-Pro motif of Cyclin D (Liou *et al*, 2002). The precise mechanism by which Pin1 stabilises Cyclin D is not yet fully understood; however, it is known that phosphorylation of this motif is linked to the targeting of Cyclin D for degradation by the UPS. Therefore, Pin1's isomerisation of this motif may influence its recognition by Ubiquitin ligases, potentially preventing Cyclin D degradation (Diehl *et al*, 1997). Another Pin1 target, the estrogen receptor α (ER α), which acts as a ligand-activated transcription factor, promotes proliferation in the majority of breast cancers. Pin1 appears to increase the activity of ER α via several mechanisms, including increasing its expression, decreasing its degradation and enhancing its binding affinity (Rajbhandari *et al*, 2012, 2014, 2015). Pin1 directly binds to the phosphorylated Ser118-Pro119 motif of the intrinsically disordered *N*-terminal transactivation domain (AF1) of ER α (Rajbhandari *et al*, 2015). Catalysis by Pin1 increases the population of the *trans* isomer of this motif, associated with increased transcriptional activity. Furthermore, coexistence of HER2 and Pin1 overexpression is frequently observed in HER2-positive breast cancers (Lam *et al*, 2008). It has been observed that Pin1 activity can stabilise HER2 via preventing its ubiquitination, however the exact motif of HER2 to which Pin1 binds has yet to be determined. These are just a few examples of the vast interactome of Pin1 and the diverse regulatory consequences it exhibits in a single cancer type, highlighting the intricate and multifaceted nature of the involvement of Pin1 in cancer.

4.2 Pin1 inhibitors

The well-established link between Pin1 overexpression and cancer progression has made the hunt for inhibitors of Pin1 an attractive field of study for the production of multiuse anti-cancer drugs. Several types of inhibitors have been developed, including those based on natural compounds, small molecule inhibitors, and peptide drugs. The literature surrounding the discovery of Pin1 inhibitors is expansive, encompassing numerous significant findings. However, despite the notable potency demonstrated by many candidates in biochemical assays, their efficacy in cellular systems has often fallen short, thus limiting their potential application as effective anti-cancer drugs. In the following section, several notable discoveries (summarised in Table 2) will be examined in greater detail, highlighting the ongoing struggle in developing Pin1 inhibitors that possess not only potent activity, but also exhibit specificity and favourable pharmacokinetic/pharmacodynamic properties. Moreover, considering Pin1's

protective role in Alzheimer's disease, it is essential to consider the possibility of unintended consequences of such drugs on neurodegeneration.

4.2.1. Therapeutic application of Pin1 inhibitors

Pin1 represents an attractive anticancer target due to its distinctive characteristics. Not only does it possess a druggable profile with clear substrate specificity and a well-defined active site, but it is also frequently overexpressed or overactivated in several significant human cancers (Zhou & Lu, 2016). Furthermore, what makes Pin1 particularly intriguing is its ability to regulate multiple oncogenes, including those that are considered non-druggable. This unique feature holds great promise in addressing aggressive or drug-resistant tumours, as targeting multiple pathways simultaneously can improve treatment outcomes (Luo *et al*, 2009; Zhou & Lu, 2016). By targeting Pin1, it could be possible to impact multiple pathways involved in cancer progression, providing a comprehensive approach to tackling the complexity of the disease. To address the question of potential unintended consequences of Pin1 inhibition on neurocognitive function, insights from studies on Pin1 KO mice revealed a relatively normal phenotype with the exception of neurodegeneration later in life, suggesting that the short-term use of Pin1 inhibitors for cancer treatment may not exhibit significant long-term consequences (Liou *et al*, 2003). In addition, it should be noted that these drugs may not even cross the blood-brain barrier (Zhou & Lu, 2016). Considering these factors, the potential benefits of targeting Pin1 in cancer treatment warrant further exploration of Pin1 inhibitors in order to fully understand their therapeutic potential.

4.2.2. Classes of Pin1 inhibitors

The development of therapeutic inhibitors from natural compounds has gained significant attention in drug discovery, harnessing the rich diversity of plant, marine, and microbial sources. Indeed, the first parvulin-specific PPIase inhibitor discovered was **juglone**, a natural brown dye isolated from walnut skin (Hennig *et al*, 1998). Juglone targets the PPIase domain of Pin1 and inhibits catalysis by irreversibly modifying the catalytic Cys113 residue. Specifically, the inhibitory mechanism involves the nucleophilic attack of the thiolate of cysteine by the 1,4-naphthoquinone moiety of juglone, to form a covalent enzyme-inhibitor complex. Juglone has been shown to be effective at reducing cancer cell proliferation (Wang *et al*, 2015; Yu *et al*, 2020). However, due to the simplicity of its molecular structure (see table

2), juglone lacks specificity, increasing the risk of off-target effects, as it has the potential to modify exposed cysteine residues of other proteins, which hence prevents its application as an anti-cancer drug.

During the pursuit of additional Pin1 inhibitors, **PiB** was discovered through low-throughput screening of compounds exhibiting anti-cancer activities, yet without a previously identified target (Uchida *et al*, 2003). PiB exhibits a similar double ring structure to juglone, but it differs in its mode of action, binding to the substrate binding sites of Pin1 and functioning as a competitive inhibitor ($IC_{50} = 1.5 \mu M$). Docking simulations of PiB binding to Pin1 suggested that its aromatic rings establish hydrophobic interactions with specific residues within the hydrophobic proline binding pocket of the PPIase domain (including Leu122, Met130 and Phe134). Furthermore, these computational models demonstrate that the oxygen atoms in PiB likely form hydrogen bonds with the sidechains of Arg69 of the phosphate binding site, and His157, a member of the catalytic tetrad. This suggests that PiB occludes the substrate-binding site of Pin1, masking the catalytic site, hence grounding it as a competitive inhibitor. However, despite inhibiting Pin1 *in vitro* and *in vivo*, PiB does not display adequate potency nor specificity for use in a multicellular system.

Another natural compound with known anticancer properties, epigallocatechin-3-gallate (**EGCG**), was also discovered to directly inhibit Pin1 *in vitro* ($IC_{50} = 20 \mu M$) (Urusova *et al*, 2011). EGCG was the first inhibitor revealed to target both domains of Pin1, and was found to delay tumour development in mice models. However, as with the other compounds, EGCG lacks specificity and is known to target many signalling pathways of the cell. Additionally, it displays relatively poor pharmacokinetic and pharmacodynamic properties, which pose challenges for its therapeutic application (Ferrari *et al*, 2022). As a result, its current use is primarily as an optional supplement with potential anticancer properties, rather than a prominent chemotherapeutic drug.

All-trans retinoic acid (**ATRA**), a metabolite of vitamin A, stands out as the sole FDA-approved inhibitor of Pin1, and is applied to the treatment of human acute promyelocytic leukaemia (APL) (Wei *et al*, 2015). Initially, the success of ATRA in the treatment of APL was attributed to its ability to bind and activate retinoic acid receptors (RARs), which in turn

induces myeloid differentiation. This mechanism was considered crucial as APL is characterised by defective myeloid differentiation which causes an accumulation of myeloid precursors cells (Di Martino & Welch, 2019). This impairment is driven by the oncogene RML-RAR α , which acts as a transcriptional repressor, inhibiting the expression of genes essential for proper differentiation. However the therapeutic concentrations of ATRA for APL treatment (1-10 μ M) are significantly higher than the concentrations required for RAR activation (10 nM), indicating the involvement of an additional mechanism (Nasr *et al*, 2008; Yu *et al*, 2020). This can be explained by the ability of ATRA to bind to the PPIase domain of Pin1 and promote its degradation (Wei *et al*, 2015). The X-ray crystal structure of the isolated PPIase domain in complex with ATRA has been resolved (PDB ID: 4TNS), revealing that the molecule binds in the substrate-binding pocket of the PPIase domain (see figure 30a). More specifically, the cyclohexane ring of ATRA is coordinated by residues Leu122, Met130 and Phe134 of the proline-binding pocket (see figure 30b).

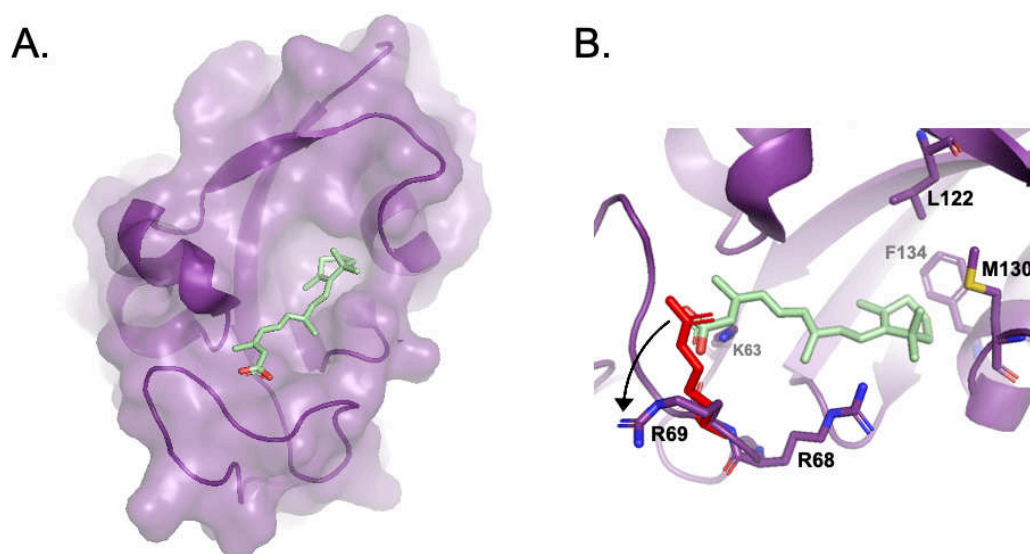


Figure 30: X-ray crystal structure of the Pin1 PPIase domain complexed with ATRA. **A.** The ATRA compound is seen to bind to the substrate-binding pocket of the PPIase domain. **B.** Residues of the proline-binding site (L122/M130/F134) coordinate the cyclohexane ring, which is a good proline mimic. Residues of the phosphate-binding sites (K63/R68/R69) coordinate the carboxylate group. The positions of these residues mimic those of Pin1 bound to the pSer-Pro peptide, except for R69 whose sidechain rotates away from the binding site to accommodate a larger substrate. R69 is depicted in red to illustrate its orientation observed during pSer-Pro binding, while in purple, it demonstrates its position when binding ATRA. Diagram made using PyMol (PDB ID: 4TNS; 1PIN).

The carboxylate group of ATRA was sufficient for binding residues Lys63, Arg68 and Arg69 of the phosphate-binding site of the PPIase domain. The position of the substrate binding residues of Pin1 are similar as in the protein when complexed to an pSer-Pro motif, though notably, the ATRA molecule is longer than the active pSer/Thr-Pro binding motif. To accommodate this, the Arg69 side-chain undergoes reorientation, pivoting away from the phosphate-binding site to prevent steric clashes. The binding of ATRA to the substrate binding site of the PPIase domain makes ATRA a potent competitive inhibitor ($IC_{50} = 820$ nM). Notably, RML-RAR α is a substrate of Pin1, and Pin1-mediated stabilisation contributes to its oncogenic activity. Consequently, the inhibition of Pin1 by ATRA results in the degradation of oncogenic RML-RAR α . The effective use of ATRA to inhibit Pin1 in the treatment of APL highlights the significant therapeutic potential of targeting Pin1 in anti-cancer therapies. This success underlines the need for continued exploration and development of novel Pin1 inhibitors to advance our understanding and expand treatment options for various cancer types.

In addition to the pursuit of potential anti-cancer candidates, various ligands targeting Pin1 have been developed, offering valuable insights into the mechanisms underlying its action, as well as highlighting key strategies for design of a potent Pin1 inhibitor. Through a structure-based rational design approach, a series of peptidomimetic inhibitors targeting Pin1 have been synthesised. These ligands characteristically incorporate a phosphate group to target the phosphate-binding sites of Pin1 (Dong *et al*, 2010; Guo *et al*, 2009). These ligands (e.g., benzothiolene-based peptidomimetics) bind directly to the substrate-binding site of the PPIase domain, as revealed by X-ray crystal structure in complex with the ligands (see figure 31a). The aromatic naphthalene and phenyl rings occupy the proline-binding pocket (making hydrophobic contacts with residues Leu122, Met130, Phe134 and His157 of the active site), while the oxygens of the phosphate group form hydrogen bonds with Lys63/Arg68/Arg69 of the phosphate-binding site. Notably, substitution of a fluorine atom into the C3 position of the phenyl ring dramatically increased inhibitory activity (C_3H $K_i = 179$ nM, C_3F $K_i = 6$ nM). Despite these inhibitors exhibiting high potency in *in vitro* assays, they often demonstrate limited or no activity in cellular environments, likely due to their poor cell permeability (Zhou & Lu, 2016). Such studies highlighted that the substitution of the phosphate group with a carboxylate moiety in peptidomimetics improved cell permeability but had a negative impact on the potency of the inhibitors. This was attributed to the fact that the carboxylate group does not fully occupy the phosphate-binding site of the PPIase domain (see figure 31b). While one

of the oxygens of the carboxylate group makes a direct hydrogen bond to Lys63, the other oxygen interacts indirectly with the Arg69 sidechain via a water molecule. Notably, no interaction with the other arginine (Arg68). This leaves half of the phosphate-binding pocket unfilled.

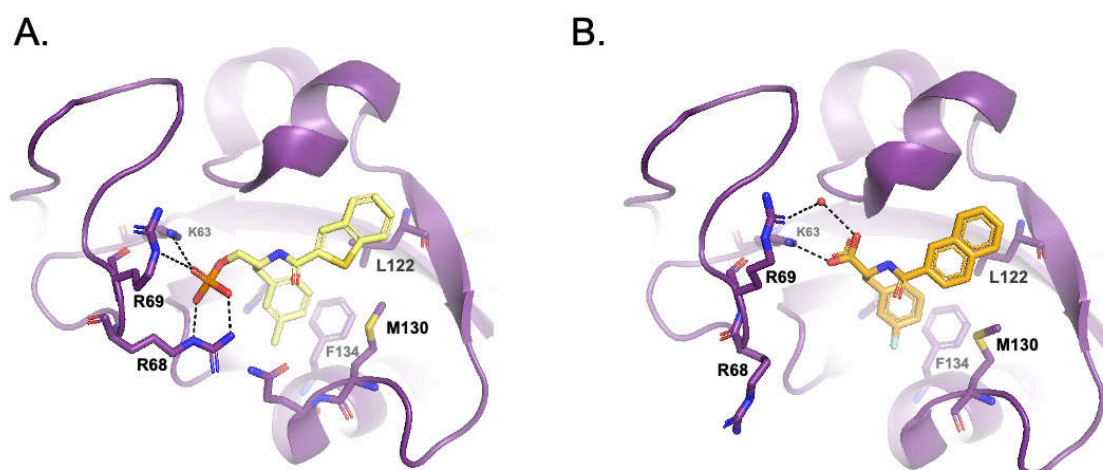


Figure 31: X-ray crystal structure of the PPIase domain of Pin1 complexed with phosphate- (A) and carboxylate-containing (B) benzothiolene-based peptidomimetics. The naphthalene and phenyl rings of both compounds form hydrophobic interactions with residues L122/M130/F134 of the proline-binding pocket. The notable difference between the binding of phosphate-containing peptidomimetic (yellow) compared to the carboxylate-containing compound (orange) is that the carboxylate group does not fully occupy the phosphate-binding pocket, not interacting at all with R68, and only forming a water-mediated interaction with Arg69. Hydrogen bonds displayed as dashed lines, water molecules as red spheres. Diagram created using PyMol, PDB ID: 3IKG and 3I6C).

Although currently not directly useful to the pharmaceutical industry, such peptidomimetics have been of great use in further scrutinising Pin1 function. For instance, conformationally locked peptides provided insights into the binding preference of the PPIase domain towards *cis* pSer-Pro motifs (Wang *et al*, 2004). This discovery highlighted the potential of these peptides as more effective inhibitors of PPIase activity, as evidence of a significantly lower IC_{50} value (*in vitro*) of a *cis*-locked alkene peptidomimetic ($IC_{50} = 1.3 \mu M$) compared to the corresponding *trans* analogue ($IC_{50} = 28 \mu M$). While thus far only one Pin1 inhibitor has been successfully applied to the treatment of human cancer, there is still promise that with good potency,

specificity and pharmacokinetic/pharmacodynamic properties, Pin1 inhibitors could be a vastly useful novel avenue for the treatment of cancer.

	Name	Structure	Mode of action	Approved for cancer treatment?
Natural compounds	Juglone		Selective and irreversible inhibition of PPlase domain (No IC ₅₀ determined)	No
	PiB		Competitive inhibition of Pin1 (IC ₅₀ = 1.5 μM)	No
	EGCG		Binds to WW and PPlase domains (IC ₅₀ = 20 μM)	No
	ATRA		Binds to PPlase domain and causes degradation (IC ₅₀ = 820 nM)	Yes Acute Promyelocytic Leukaemia
Peptides	Benzothiolene-based peptidomimetic		Binds to the active site of Pin1 (IC ₅₀ = 6 μM)	No
	Cis-locked peptidomimetic		Selective binding to the PPlase domain as a substrate mimic (IC ₅₀ = 1.3 μM)	No

Table 2: The structure, properties and application of selected Pin1 inhibitors.

4.3 The incorporation of fluorine into pharmaceutical compounds

The struggles associated with the production of a potent Pin1 inhibitor for use as an anticancer drug are not unique to this enzyme. The production of a successful pharmaceutical compound relies not only on effective binding and inhibition of the targeted enzyme but also on optimising its pharmacokinetic and pharmacodynamic properties, ensuring appropriate interactions within the complex multicellular environment of the human body. These interactions can be described in terms of the drug's absorption, distribution, metabolism, and excretion (ADME) properties (Lin *et al*, 2003). These properties play a crucial role in determining the bioavailability of the drug compound, i.e., the amount that reaches its intended target destination.

In the pharmaceutical industry, small molecule drugs with a molecular weight (MW) below 500 g/mol have traditionally dominated the market (Czechtizky & Hamley, 2016). These small molecules offer several advantages, such as good cell permeability, which enhances their bioavailability. Additionally, they are often easier to synthesise compared to larger molecules, making them more accessible for development and production. However, in the past decade, a substantial shift has occurred in favour of producing biologic drugs (e.g., peptide-based drugs) which harbour advantages including increased target specificity, reduced side effects, and lower toxicity compared to small molecule drugs (Nelson *et al*, 2010). However, a common limitation of biologics is their challenge in penetrating cell membranes, as a result the majority of peptide-based drugs target cell membrane or extracellular proteins.

When it comes to enhancing the bioavailability of drugs, there are various strategies available. Among these, the incorporation of fluorine atoms into pharmaceutical compounds has gained popularity due to the favourable pharmacokinetic and pharmacodynamic properties it affords (Inoue *et al*, 2020). In fact, as many as 20 % of current pharmaceutical compounds contain at least one fluorine atom. Fluorine, being both small and highly electronegative, can have a remarkable influence on the pharmacological effect of a drug compound when substituted for hydrogen or hydroxyl groups at key positions (see table 3) (Shah & Westwell, 2007). Firstly, the C-F bond exhibits significantly greater stability than the C-H bond, making fluoropharmaceuticals less prone to metabolism and degradation within the body (Park *et al*, 2001).

In addition, fluorine's strong electron-withdrawing property allows it to significantly modify the pKa of neighbouring functional groups, thereby impacting the compound's acidity or basicity, a property which has the potential to enhance the membrane permeability of a drug (Shah & Westwell, 2007). Furthermore, many studies have indicated that the incorporation of fluorine atoms in aliphatic regions can increase the lipophilicity of a compound (Jeffries *et al*, 2020; Park *et al*, 2001). This effect is further amplified when groups such as CF₃ (trifluoromethyl) are utilised, as compared to single fluorine atoms. In the case of peptide molecules, the favourable effects of fluorine incorporation are particularly pronounced and well-translated. The hydrophobicity and stability of the peptide structure is indeed increased upon substitution with fluorinated amino acids (Salwiczek *et al*, 2012). Most fluorinated amino acids integrated into peptide molecules employ CF₃ groups as replacements for native CH₃ groups, effectively mimicking their structure and function whilst affording additional favourable properties.

Element	VDW radius (Å)	Electro-negativity	Bond length (CH ₂ X) Å	Bond energy (kcal/mol)
H	1.2	2.1	1.09	99
F	1.47	4.0	1.39	116
O (OH)	1.52	3.5	1.40	85

Table 3: Physiochemical properties of fluorine.

In certain cases, such as with proline, the substitution of a CF₃ group can result in further notable effects on the peptide's characteristics. As previously introduced, the distinctive cyclic structure of proline gives rise to a unique peptide bond that can exist in both the *cis* and *trans* conformers. It has long been established that the production of oxazolidine- and thiazolidine-based pseudo-proline (ΨPro) analogues can enhance these properties of proline, lowering the transition state barrier between *cis* and *trans* conformers (Mutter *et al*, 1999). Specifically, Cδ-substituted ΨProlines (Cδ-ΨPro) have the extraordinary capacity to promote the adoption of the *cis* conformation of the prolyl peptide bond (see figure 32) (Dumy *et al*, 1997; Keller *et al*, 1998; Mutter *et al*, 1999). Despite its remarkable capacity to enhance the population of *cis* amide bonds, oxazolidine-based ΨPro exhibits limited stability in acidic environments due to

Chapter 5 General Aims of the Thesis

Pin1 serves as a master regulator of pivotal cellular processes including the cell cycle, DNA replication, transcription, cell development, and apoptosis (Liou *et al*, 2011). This regulatory function arises from its capacity to catalyse the isomerisation of pSer/Thr-Pro motifs within its protein substrates, leading to alterations in their stability, post-translational modifications, cellular localisation, protein interactions, and activity (Zhou & Lu, 2016). Given its pivotal role in governing an extensive interactome of proteins, dysregulation of Pin1 has been implicated in various human pathologies, including cancer and neurological disorders (Li *et al*, 2021). Notably, the overexpression of Pin1 has been observed in multiple human cancers, thus highlighting its potential as a promising therapeutic target. However, despite its status as a well-studied and crucial regulatory protein, the precise mechanisms enabling Pin1's selective binding and regulation of such a vast interactome remain incompletely understood. Furthermore, despite numerous attempts to develop Pin1 inhibitors as anticancer drugs, existing compounds have generally exhibited limited potency, specificity, and bioavailability (Wu *et al*, 2022).

To expand on the existing knowledge of Pin1 binding mechanisms, this thesis is divided into two distinct parts:

- i. Characterisation of the binding and inhibitory activity of novel peptide ligands of Pin1 (in collaboration with the BioCis laboratory of Cergy-Pontoise)
- ii. Structural characterisation of the interaction between Pin1 and the intrinsically disordered region of Cdc25C, which hosts six potential Pin1 binding sites

By pursuing these objectives, the primary aim is to further the understanding of Pin1-mediated substrate regulation, with a specific focus on unravelling the underlying molecular mechanisms that govern its substrate binding. This investigation will not only shed light on the intricate interplay between Pin1 and its substrates but also has the potential to pave the way for the development of innovative therapeutic interventions targeting cancer.

5.1 Characterisation of the binding and activity of novel peptide ligands of Pin1

The limited understanding of the Pin1 interaction with its target proteins has broader implications beyond the elucidation of the enzyme's mechanism. It also hampers the ability to design effective ligands as potential anti-cancer candidates. Pin1 overexpression in cancer is extensively documented and associated with poor prognoses (Pu *et al*, 2020). This is attributed to Pin1's extensive interactome, which involves various oncogenic and tumour-suppressing proteins. The ability of Pin1 to target several proteins involved in cancer, including those that are seen as non-druggable, enhances its attractiveness as an anticancer candidate (Zhou & Lu, 2016). Although, a potential limitation of Pin1 inhibition is the neuropathogenic phenotype associated with Pin1 abolition. The protective role of Pin1 against neurodegeneration has been well-documented in studies of patients with AD and Pin1 KO mice (Li *et al*, 2021; Liou *et al*, 2003). However, designing inhibitors that do not cross the blood-brain barrier can help to avoid potential issues, ensuring selective and safe targeting of Pin1 in cancer cells while preserving the neuroprotective function of Pin1 in non-cancerous tissues.

Despite significant efforts, the development of Pin1 inhibitors as anti-cancer therapies has encountered several challenges, with existing compounds demonstrating inadequate potency, lack of specificity, and poor cell permeability, hindering their therapeutic potential. In collaboration with chemists of the BioCis laboratory at Cergy-Pontoise, an ongoing project focuses on the design of a new generation of fluorinated peptide ligands targeting Pin1. This innovative approach, supported by the ANR grant F-LAIR, aims to address the limitations associated with existing compounds by combining two emerging trends in pharmacology: the utilisation of peptide drugs, known for their lower toxicity profiles and higher specificity, and the incorporation of fluorine atoms, which have been shown to enhance the bioavailability of drugs (see chapter 4.3). The design strategy for these ligands is based on extensive literature studies. Templates with a minimum backbone length of four residues (P1-4) have been selected, incorporating the optimal pSer-Pro Pin1 recognition motif at positions 2 and 3. The peptides also include aromatic residues at the *N*-termini and aromatic or basic residues at the *C*-termini based on observed preferential binding of such groups to Pin1 (Wildemann *et al*, 2006). More specifically, the optimal consensus sequence for Pin1 binding was found to be

(W/F/Y)-(F/I)-Y/R/F/W)-**pS-P**-(R/F/Y/W)-(L/I). Several strategies for the incorporation of fluorine are being explored. The incorporation of a CF₃-pseudoproline group at the C δ carbon of Ser ($\Psi^{\text{H,CF}_3}\text{Pro}$) was shown to significantly decrease the uncatalysed *cis/trans* rotational barrier (ΔG^\ddagger from ≈ 20 Kcal/mol for the Xaa-Pro peptide bond to ≈ 15 Kcal/mol for the Xaa- $\Psi^{\text{H,CF}_3}\text{Pro}$ peptide bond) (Greenwood *et al*, 2011). This is close to the free energy barrier measured on WT peptides in the presence of Pin1 ($\Delta G \approx 13$ Kcal/mol). As a result, the incorporation of $\Psi^{\text{H,CF}_3}\text{Pro}$ at P3 position of the peptide template is expected to favour the isomerisation reaction catalysed by Pin1. Additionally, incorporating this group is expected to enhance the *cis* population, aiding the development of PPIase domain-specific ligands. Alternatively, fluorine could be incorporated via a CF₂-phosphonate group, in the hopes of improving the stability of the phosphate moiety and enhancing the metabolic profile of the ligands.

Based on these templates, a first-generation library of ligands (table 4) was generated, enabling the identification of critical features which contribute to the binding affinity. The binding of these ligands to FL Pin1 and to each isolated domain was previously assessed by the lab. As displayed in table 4, the highest affinity ligands displayed K_D values in the tens of micromolar range. Notably, ligands with a C-terminal tryptamine group, inspired by Etzkorn's ligand library, displayed the highest affinity for Pin1 (Xu *et al*, 2012). N-terminal acetylation and Fmoc group protection, additional techniques employed by the Etzkorn group, were also integrated into the initial ligand designs in this project. N-terminal Fmoc protection was found to improve affinity compared to acetylation. The first generation of fluorinated ligands focussed on the incorporation of fluorine via the CF₃-pseudoproline group (see figure 33). It was found that the presence of the group indeed decreased the *cis/trans* rotational barrier and enhanced the *cis* population. Consequently, peptides containing this group, for example CZ115, exhibited preferential binding to the PPIase domain (figure 34a). On the other hand, peptides lacking this group, for instance CZ55, adopted the *trans* conformation and displayed specific binding to the WW domain (figure 34b). This approach has thus enabled the creation of domain-specific Pin1 ligands. However, while the presence of an N-terminal Fmoc group improved binding affinity, it was also observed that it could compromise the single-domain specificity of the ligands, as evidenced in the comparisons of the K_D values calculated for (Fmoc)-CZ40 vs. (Ac)-CZ55 and (Fmoc)-CZ34 vs. (Ac)-CZ52 (see table 4).

Ligand name	Sequence	Kd with FL Pin1 (μM)		Kd with Isolated domains (μM)	
		WW	PPIase	WW	PPIase
CB06	Ac-pSer-Pro-Trp-NH ₂	429 \pm 70	1232 \pm 350	ND	ND
CB07	Ac-Ser-Pro-Trp-NH ₂	> 1mM	> 1mM	ND	ND
CB08	Ac-pSer-Pro-Phe-NH ₂	518 \pm 114	1625 \pm 332	ND	ND
CZ55	Ac-pSer-Pro-Tryptamine	32 \pm 5,8	> 1mM	69	> 1mM
CZ52	Ac-pSer- $\Psi^{\text{CF}_3, \text{H}}$ Pro-Tryptamine	> 1mM	490 \pm 50	> 1mM	540 \pm 65
CZ40	Fmoc-pSer-Pro-Tryptamine	26 \pm 6	438 \pm 80	ND	ND
CZ34	Fmoc-pSer- $\Psi^{\text{CF}_3, \text{H}}$ Pro-Tryptamine	7 \pm 4	41 \pm 8,6	ND	ND
CZ115	Ac-Cha-pSer- $\Psi^{\text{CF}_3, \text{H}}$ Pro-reduced Tryptamine	> 1mM	28 \pm 6	> 1mM	20 \pm 9
CZ178	Ac-pSer- $\Psi[\text{CH}(2)\text{N}]$ -Pro-Tryptamine	> 1mM	15 \pm 4	1730	4,3 \pm 2,2
CZ181	CF ₃ -pSer- $\Psi[\text{CH}(2)\text{N}]$ -Pro-Tryptamine	> 1mM	44 \pm 6,9	1445	14,4 \pm 2,4

Table 4: The first-generation library of Pin1 peptide ligands. The first-generation library tested peptides containing the classical Pin1 consensus sequence pSer-Pro (aside from CB07 which was unphosphorylated). Various *N*-terminal (Ac- or Fmoc-) and *C*-terminal (Phe, Trp or tryptamine) moieties were tested. The incorporation of fluorine via the CF₃-pseudoproline group ($\Psi^{\text{CF}_3, \text{H}}$ Pro) was explored (CZ52, CZ34 and CZ115), enabling the specific targeting of the PPIase domain. For ligand CZ115, the originally intended sequence was Ac-Phe-pSer- $\Psi^{\text{CF}_3, \text{H}}$ Pro-Tryptamine however a reduction reaction during synthesis resulted in the formation of a cyclohexane moiety (Cha) at position P1 and a reduced tryptamine at the P4 position. The $\Psi[\text{CH}(2)\text{N}]$ group of CZ178 and CZ181 refers to a reduced amide bond. Table taken from the thesis of Soha Abou-Ibrahim.

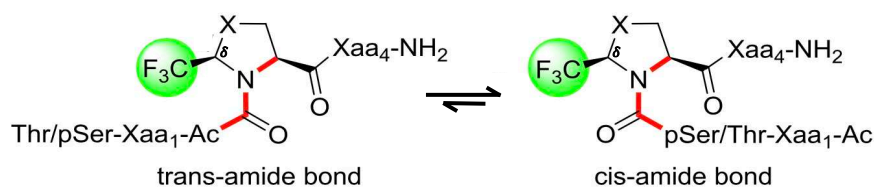


Figure 33: General structure of ligands containing the CF₃-pseudoproline group. In the case of the first-generation ligand library X = O. The CF₃-pseudoproline residue favours the *cis* conformation of the peptide bond.

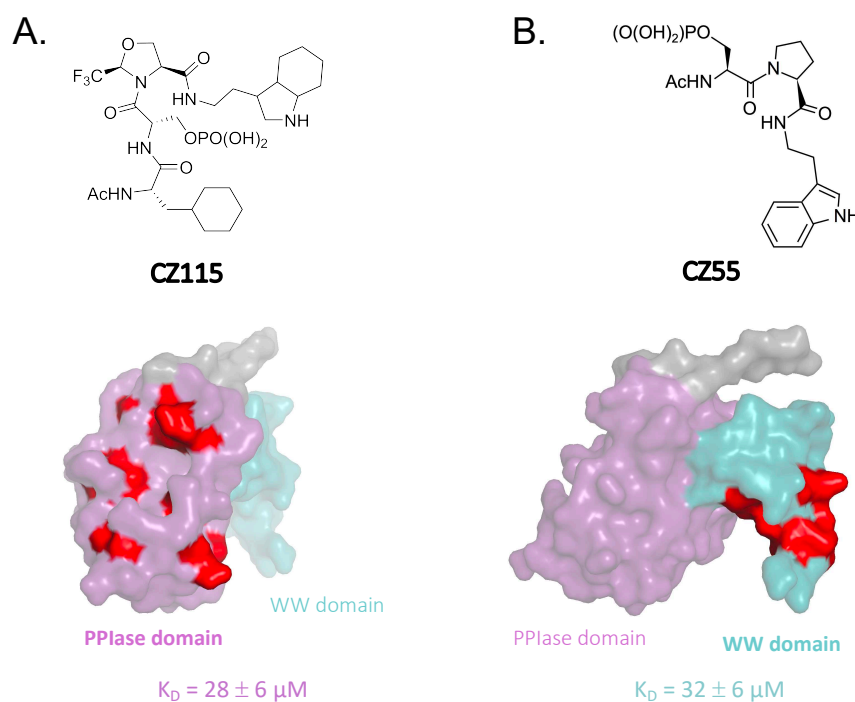


Figure 34: Domain-specific ligands synthesised in the first-generation library of fluorinated peptide ligands of Pin1. **A.** CZ115 is an example of one of the ligands synthesised which contains the CF₃-pseudoproline group. The ligand showed preferential binding to the PPIase domain (the interaction surfaces highlighted in red). This was the highest affinity PPIase-specific ligand from the first-generation library (K_D PPIase domain = $28 \pm 6 \mu\text{M}$). **B.** The CZ55 ligand, which was synthesised without incorporation of the CF₃-pseudoproline group, displayed specific binding to the WW domain (interaction surfaces highlighted in red). This ligand displayed the highest specificity and affinity for the WW domain (K_D WW domain = $32 \pm 6 \mu\text{M}$).

With the promising results obtained from the first-generation library of ligands, a second-generation library of ligands will be generated, and their binding to Pin1 assessed in this work. This second-generation library will test additional strategies for Pin1 targeting, incorporating key groups identified in the first-generation and their synergetic effects for

optimal binding. The overall objective is to generate higher affinity domain-specific ligands which incorporate the CF₃-pseudoproline group, making them potential candidates for application as anticancer drugs. Additionally, the enhancement of the pharmacokinetic and pharmacodynamic properties of Pin1 ligands will be explored through two distinct strategies: the removal of the dianionic phosphate group, either by complete elimination or replacement with a glutamate phosphomimetic or via its substitution with a CF₂-phosphonate group. Both approaches aim to improve the ligands' characteristics, with phosphate group removal primarily targeting enhanced cell permeability, while the CF₂-phosphonate group sought to enhance the metabolic stability of the ligands, leading to more promising drug candidates. Due to the challenges in achieving nanomolar affinity, an alternative approach for Pin1 inhibition utilising Proteolysis-Targeting Chimeras (PROTAC) technology was also investigated by chemists of the BioCis laboratory. This strategy involves the design of Pin1 ligand/ubiquitin ligase chimeras, which would effectively hijack the native ubiquitin proteasome system (UPS) of the cell, resulting in the degradation of Pin1. As described in detail in chapter 7.1.3, this technique should not require nanomolar affinities to achieve significant reduction of cellular Pin1 levels, making it a promising alternative to catalytic inhibition. In addition to characterisation binding properties of the ligands, the development of a robust Pin1 activity assay is being pursued. This assay aims to assess the ability of the synthesised molecules to inhibit Pin1's catalytic activity. Indeed, although the first-generation peptides were identified to bind to Pin1, preliminary activity assays in the presence of these ligands yielded inconclusive results regarding their inhibitory effect. The lack of inhibition data was attributed to unreliability of the fluorometric assay employed. Hence, the optimisation of a Pin1 activity assay based on UV spectroscopy or NMR EXSY experiments will be explored to compare the inhibitory potential of the lead compounds from the first- and second-generation ligands.

5.2 Structural characterisation of the interaction between Pin1 and the IDR of Cdc25C


To gain a comprehensive understanding of Pin1's interaction with its protein substrates, further structural investigations are required. Specifically, studies examining Pin1 binding to substrates containing multiple binding motifs hold the potential to unveil the coordinated substrate binding mechanisms of the WW and PPIase domains. As previously introduced, the

Cdc25C phosphatase is a well-known Pin1 target containing six Ser/Thr-Pro motifs in its IDR which could be phosphorylated to form Pin1 substrates (Liu *et al*, 2020). Indeed, the combined phosphorylation of at least five of these sites has been confirmed in the mitotic form of Cdc25C, which serves as the preferential binding partner of Pin1 (Sur & Agrawal, 2016). Mutagenesis studies have established the indispensable role of these five sites in Pin1 binding, with evidence suggesting a heightened significance of three phosphorylated Thr-Pro sites, namely Thr48-Pro, Thr67-Pro, and Thr130-Pro (Izumi & Maller, 1993; Shen *et al*, 1998). Among these sites, Thr48 and Thr67 have been proposed as the primary binding sites for Pin1 (Lu *et al*, 1999b); (Zhou *et al*, 2000). However, it is essential to emphasise that no structural studies have yet supported this proposition. In addition, the regulation of Cdc25C by Pin1 has been seen to be more complex than originally thought, exhibiting a high dependency on the phosphorylation status of Cdc25C (Stukenberg & Kirschner, 2001). Pin1 has been observed to exhibit dual and opposing effects on Cdc25C activity, acting as an inhibitor when Cdc25C is phosphorylated solely by CDK1 and as an activator when it is phosphorylated by both CDK1 and Plk1. However, the specific phosphorylation sites of Cdc25C involved in each context remain undetermined.

The current understanding of the interaction between Pin1 and Cdc25C raises several unanswered questions that warrant further investigation. Firstly, it is crucial to structurally confirm which motifs of Cdc25C are preferentially bound by Pin1. In particular, to further understand the regulatory effect of Pin1-catalysed prolyl isomerisation on Cdc25C activity, the differential phospho-states of Cdc25C which provoke its inhibition or activation should be characterised. Currently, it can only be hypothesised which sites are phosphorylated by CDK1 and Plk1 when acting in unison. This could help shed light onto how Pin1 differentially recognises and interacts with these two phospho-forms to exert an antagonistic regulatory role. Furthermore, the interaction between Pin1 and Cdc25C has predominantly been investigated from the perspective of Cdc25C, mostly focusing on the effects of Pin1 binding on Cdc25C activity. However, the binding behaviour of Pin1 itself, specifically the precise binding domains of Pin1 which interact with Cdc25C, remains largely unexplored. While studies have suggested that the WW domain exhibits a higher affinity for the Cdc25C protein and serves as the primary binding module for a phosphorylated peptide derived from Cdc25C (pCdc25C), the interaction of the PPIase domain with Cdc25C lacks sufficient evidence (Jacobs *et al*, 2003; Zhou *et al*, 2000). Considering that the PPIase domain catalyses the isomerisation of Cdc25C

motifs, it is imperative to determine the extent of its interaction with Cdc25C and understand its contribution to the overall Pin1-Cdc25C interaction. The only existing structural investigations of the interaction between Pin1 and Cdc25C have primarily focused on the binding of the pCdc25C peptide (EQPLpTPVTDL), which contains only a single Pin1 binding motif (Born *et al*, 2022; Jacobs *et al*, 2003; Namanja *et al*, 2007, 2011; Peng, 2015; Wang *et al*, 2015). Consequently, these studies likely overlook crucial aspects of the interaction. Investigating the interaction between Pin1 and the IDR of Cdc25C, rather than relying solely on short peptide sequences, would provide valuable insights into the broader substrate binding mechanisms of Pin1. Expanding structural investigations to encompass Pin1's interaction with full-length proteins or protein regions, particularly those with multiple pSer/Thr-Pro motifs, holds the potential to significantly advance our understanding of Pin1's substrate recognition.

In this work, NMR spectroscopy will be employed to characterise the molecular interaction between Pin1 and the IDR of Cdc25C. NMR spectroscopy offers the advantage of providing atomic-level insights into the interaction which will allow the identification of residues of each protein that are involved in the interaction. The presence of a putative 230-residue IDR in the *N*-terminal region of Cdc25C poses a challenge for its structural investigation by X-ray crystallography which relies on the presence of a well-defined three-dimensional structure. By contrast, solution state NMR can be employed to characterise both the conformational dynamics and interactions of IDPs at the atomic level, with the solution state being more similar to the native cellular environment. Furthermore, while the NMR investigation of well-folded globular proteins becomes challenging for proteins larger than 30-40 kDa, the rapid tumbling of IDPs and IDRs in solution improves their relaxation properties, leading to sharp and narrow peaks in NMR spectra, even for larger proteins. However, the inherent structural flexibility of IDPs strongly reduces signal dispersion, in particular in the ^1H dimension, which can result in extreme spectral crowding and requires the use of spectrometers with high magnetic fields to enhance signal resolution. To study the IDR of Cdc25C and its interaction with Pin1, a 500 MHz NMR spectrometer hosted by the laboratory will be used, complemented by a 950 MHz spectrometer facilitated through financial support from the Infranalytics research federation of the CNRS. In order to investigate the impact of Cdc25C phosphorylation on the interaction with Pin1, both phosphomimetic mutants and *in vitro* phosphorylated forms of the Cdc25C IDR will be produced.



*RESULTS &
DISCUSSION*

Chapter 6 Production of Samples for NMR Analysis

To investigate the interaction of Pin1 with the IDR of Cdc25C and the second generation of fluorinated peptide ligands, it was essential to generate isotopically labelled and unlabelled samples of both Pin1 and the IDR of Cdc25C. This chapter outlines the process of expression and purification employed to obtain these protein samples. Detailed protocols are provided in the Material and Methods section (chapter 9.3-4).

6.1 Expression and purification of recombinant Pin1

The expression of human Pin1 (18.5 kDa) in *E. coli* has been well-documented in various studies and has undergone optimisation within our laboratory. A pET-15b plasmid was employed for this purpose. This plasmid contained a gene conferring ampicillin resistance and encoded the full-length (FL) Pin1 protein as a fusion construct. The fusion protein consisted of an *N*-terminal (His)₆ tag separated from the Pin1 sequence by a PreScission cleavage site. The pET-15b plasmid was introduced into *E. coli* BL21-Gold (DE3) PlysS competent cells, which possessed an inherent resistance to chloramphenicol. Transformants were selected based on their dual antibiotic resistance phenotype, indicating successful plasmid uptake. Glycerol stocks were prepared from the transformed *E. coli* cells carrying the Pin1-encoding pET-15b expression vector, allowing for long-term storage and availability of cell-cultures harbouring the recombinant plasmid.

6.1.1. Expression of unlabelled Pin1

E. coli cells harbouring the pET-15b^{Pin1} plasmid were grown as large-scale cultures in LB media supplemented with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. The cultures were allowed to grow at 37°C until reaching an optical density at 600 nm (OD₆₀₀) of 0.6-0.8. Protein expression was induced with the addition of 0.5 mM IPTG, followed by further incubation at 37°C for 4 hours. Subsequently, the cells were harvested through centrifugation and stored at -80°C in Pin1 binding buffer (see Materials and Methods section) overnight.

6.1.2. Expression of ^{15}N -labelled Pin1

The Marley protocol (Marley *et al*, 2001), a well-established method known for its high yield in producing uniformly labeled protein samples, was employed to produce isotopically labelled (^{15}N) Pin1 samples. The production approach began by cultivating *E. coli* cells containing the pET-15b^{Pin1} plasmid in a high volume of LB media (typically 4 L) to generate a high cell mass. Subsequently, the cells were pelleted and transferred to 1 L of M9 media (lacking glucose and NH_4Cl), resulting in a 4-fold increase in cell density. Following a short period to allow for growth recovery and enable the consumption of residual LB media, the M9 media was supplemented with isotopically-labelled nutrients (^{15}N NH_4Cl and ^{12}C -glucose). Protein expression was induced using 0.5 mM IPTG at 37°C for 4 hours. Finally, the harvested cells were stored at -80°C in Pin1 binding buffer to maintain sample integrity prior to purification.

6.1.3. Purification of FL Pin1 Ni-NTA chromatography and gel filtration

Cell pellets obtained from centrifugation of LB and M9 cultures were lysed by sonication and the soluble cell-free extract (CFE) isolated from the cell debris by centrifugation. The purification of labelled and unlabelled samples of FL Pin1 was carried out in 3 steps. Firstly, the CFE was passed through an Ni-NTA column equilibrated with Pin1 binding buffer using an AKTA purifier apparatus. The flow-through (FT) containing the bulk of proteins from *E. coli* cells was collected, and the column washed with buffer to remove weakly-interacting proteins. Protein elution was achieved by applying an increasing gradient of imidazole concentration over 40 mL volume. Each step of this process was monitored by a chromatogram measuring UV absorption at 280 nm and the collected fractions were analysed by SDS-PAGE (see figure 35). High-intensity bands with apparent molecular weight corresponding to FL Pin1 (18.5 kDa) are mostly observed for fractions corresponding to Peak 2, indicating that the protein is well-expressed.

The fractions containing Pin1 were pooled and incubated with PreScission enzyme at 4°C overnight to cleave the (His)₆ tag. The (His)₆-tagged PreScission enzyme used in this study was produced in-house as described in section 6.3. To separate fully-cleaved Pin1 from the (His)₆ tag fragments and (His)₆-tagged PreScission enzyme, the sample underwent a second

round of purification using an Ni-NTA column. Before application to the column, the mixture underwent a 3-fold dilution in Pin1 binding buffer to reduce the concentration of imidazole. Initially, dialysis employing a 3 kDa membrane was implemented at this stage to achieve a complete buffer exchange; however, a notable protein loss (up to 25%) was noted, necessitating a modification of this procedure. After injection of the sample into the nickel column, the cleaved protein was no longer retained by the column and was present in the FT, confirming successful cleavage (see figure 36).

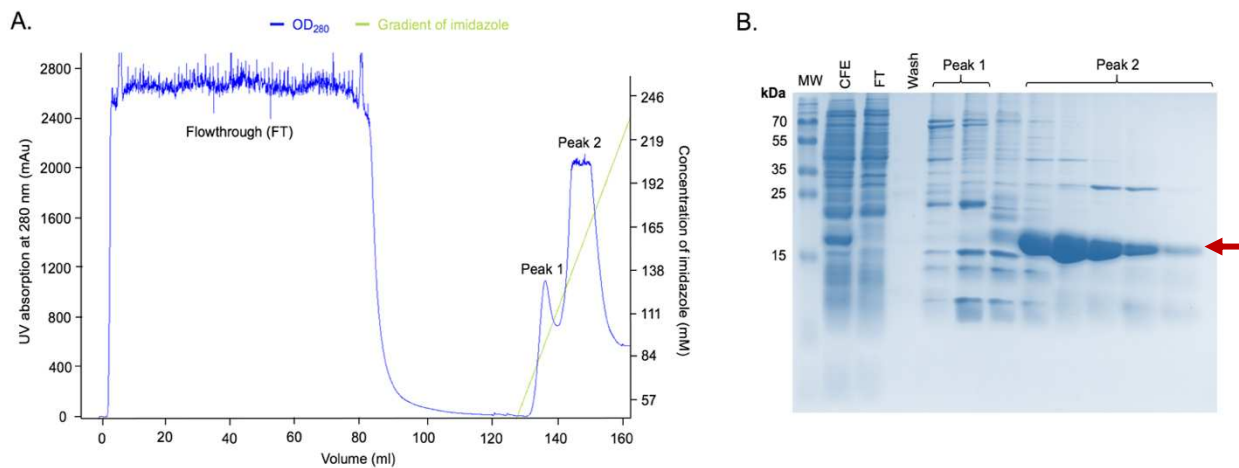


Figure 35: Chromatogram (A) and SDS-PAGE analysis (B) of the 1st passage of FL Pin1 through the Ni-NTA column. Pin1 was eluted from the column at imidazole concentration 138-165 mM (peak 2) as identified by SDS-PAGE analysis whereby a peak at 18.5 kDa is observed in these fractions (highlighted by the red arrow).

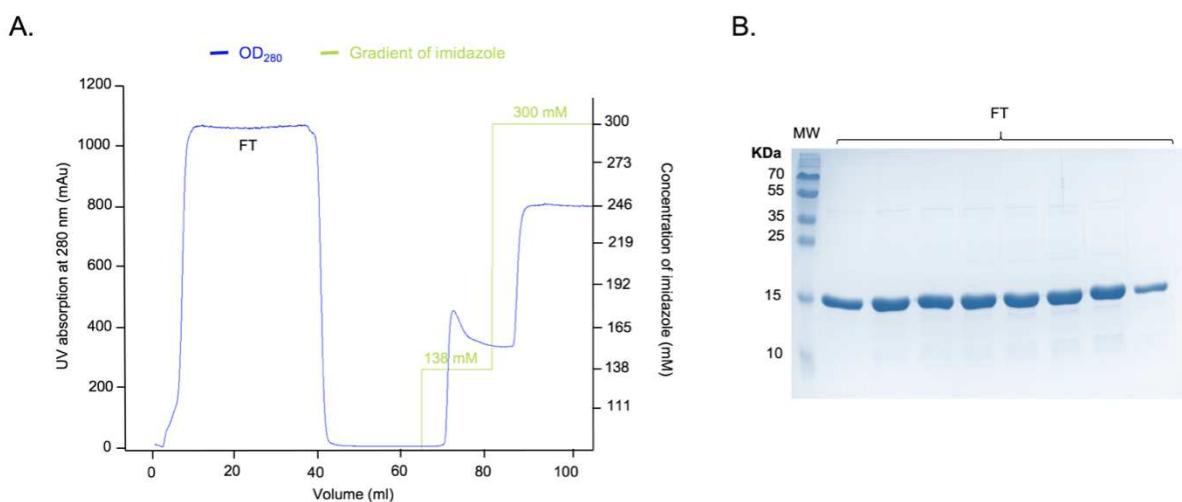


Figure 36: Chromatogram (A) and SDS-PAGE analysis (B) of the 2nd passage of FL Pin1 through the Ni-NTA column. After cleavage of the (His)₆ tag by PreScission enzyme, Pin1 was collected in the FT after passing through the column. Fractions of the FT were collected and analysed by SDS-PAGE, confirming the presence of Pin1 (band at 18 kDa).

A final purification step was carried out using size exclusion chromatography (SEC) by passing the Pin1-containing sample through a Superdex 75 16/600 column using an AKTA purifier. SEC, also known as gel filtration chromatography, operates on the principle of separating molecules based on their hydrodynamic size and shape, allowing for fractionation, isolation, and purification of proteins according to their molecular size. Before injection into the column, the sample was first concentrated to a volume of 5 mL using an Amicon stirred ultrafiltration cell equipped with a 3 kDa ultrafiltration membrane. A pressure-based concentration method was chosen over more conventional centrifugal filtration devices as previous studies by the lab had noted a significant protein loss (up to 60%) associated with the use of centrifugal filters. To enhance separation efficiency and ensure simultaneous entry, the sample was injected as two distinct 2.5 mL fractions into the gel filtration column, aligned with the injection tubing's capacity. This SEC step not only enhanced the purity of Pin1 but also facilitated the transfer of Pin1 into the suitable buffer for NMR experiments, which varied depending on the specific requirements of each experiment. The Pin1 samples obtained after SEC show a very high level of purity with the absence of additional bands in the SDS-PAGE analysis of the elution fractions (figure 37).

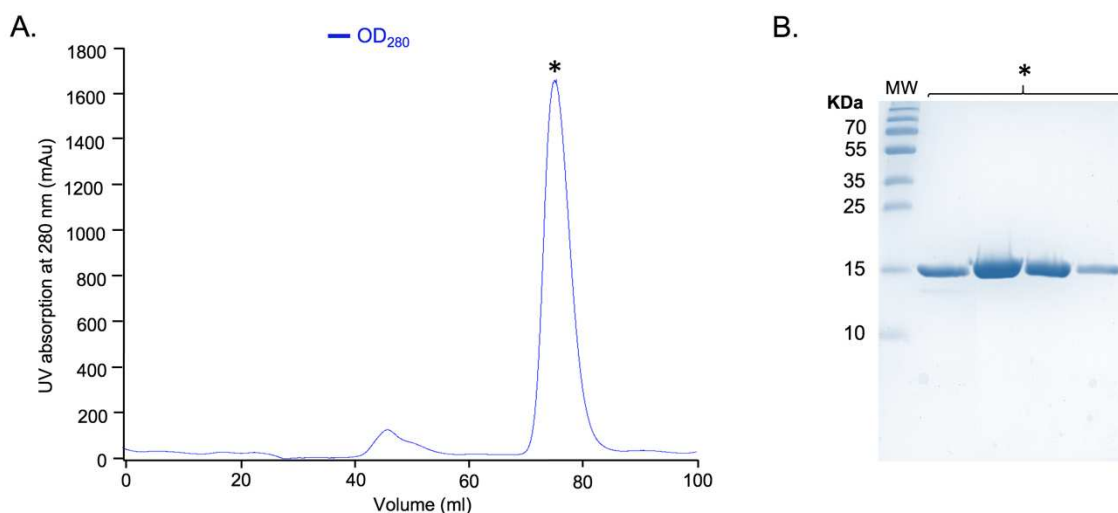


Figure 37: Chromatogram (A) and SDS-PAGE analysis (B) of Pin1 purification by SEC

6.1.4. *Optimisation of the protocol*

Although the production and purification protocol employed in this study was a well-established approach previously utilised by the laboratory, it still benefited from some optimisation. Several batches of unlabelled and ^{15}N -labelled Pin1 were prepared throughout

the course of this thesis and optimisation steps were required to increase both yield and purity. The yields of Pin1 purified from LB and M9 media for each batch were compared and are summarised in Table 5 and 6 respectively. Notably, application of the Marley method for the production of ^{15}N -Pin1 was indeed successful, as the yield of single-labelled samples was not diminished in comparison to unlabelled samples.

Batch	Yield (mg/L of LB)
1	8
<hr style="border-top: 1px dashed green;"/>	
2	10
<hr style="border-top: 1px dashed red;"/>	
3	16
4	15
5	18

Table 5: Yields (mg of purified protein per litre of LB medium) of unlabelled Pin1 acquired throughout this thesis. The dashed lines represent the implementation of optimisation steps to enhance cleavage by PreScission enzyme (green) and to limit degradation (red).

Batch	Yield (mg/L of M9)
1	9
<hr style="border-top: 1px dashed green;"/>	
2	9
<hr style="border-top: 1px dashed red;"/>	
3	8
4	17
5	10
6	16
7	14

Table 6: Yields (mg of purified protein per litre of M9 medium) of ^{15}N -labelled Pin1 acquired throughout this thesis. The dashed lines represent the implementation of optimisation steps to enhance cleavage by PreScission enzyme (green) and to limit degradation (red).

The initial batches of both labelled and unlabelled Pin1 proteins exhibited relatively low yields (~9 mg/L of bacterial culture). Moreover, upon performing SDS-PAGE analysis after gel filtration, the presence of two bands close to 18 kDa was observed (see figure 38a). This observation led to the hypothesis that there were two species present in the purified protein samples, which could be attributed to incomplete cleavage of (His)₆-Pin1 by the PreScission enzyme or protein degradation during the purification process. To investigate this further, a 2D ^1H - ^{15}N HSQC NMR spectrum of the first batch of ^{15}N -Pin1 was recorded, revealing several unexpected peaks (see figure 38b). Based on these findings, it was postulated that the sample contained both (His)₆-Pin1 and fully-cleaved Pin1. This hypothesis suggested that the Ni-NTA column used may not have efficiently retained the bound protein, potentially contributing to the lower yield. Notably, the first batch of unlabelled and ^{15}N -Pin1 were produced using an old

batch of PreScission enzyme and a reused Ni-NTA column. To enhance yield and purity, fresh PreScission enzyme was prepared (as detailed in section 6.3) and a new Ni-NTA column was used. Markedly, this increased the purity of the sample leading to a single population of cleaved Pin1 (confirmed by SDS-PAGE analysis and NMR spectroscopy). However, the yield did not exhibit a significant increase.

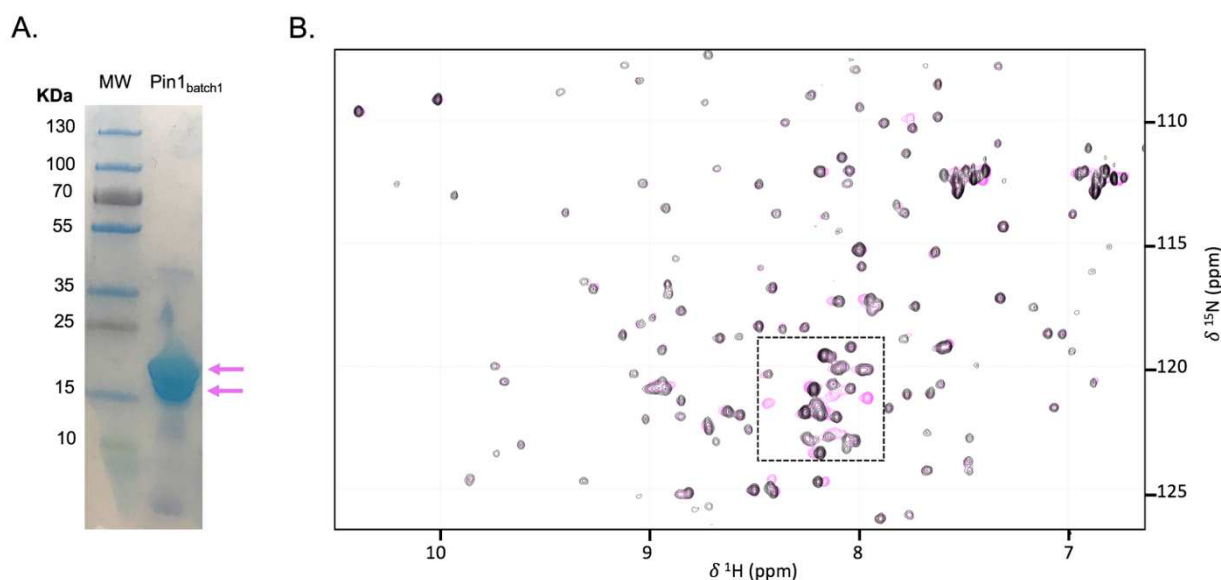


Figure 38: Characterisation of the first batch of Pin1 samples using SDS-PAGE analysis (A) and a 2D ^1H - ^{15}N HSQC NMR Spectrum. Both the SDS-PAGE and NMR analysis of the first batch of Pin1 purified indicated the presence of two species. From the SDS-PAGE gel it is clear that these two species are both very close in molecular weight ($\sim 18\text{kDa}$). The ^1H - ^{15}N HSQC of fully cleaved/purified Pin1 (black) and the first batch of purified Pin1 (magenta) are compared. Several additional peaks are observed in the spectrum of Pin1_{batch1}, particularly in the highlighted region.

In light of the persistently low yield, it was postulated that protein degradation might be occurring during the purification process. To investigate this possibility, antiproteases were introduced to the samples at each purification step and freeze-thaw cycles were avoided throughout the purification in attempt to limit protein degradation. After gel filtration, the protein was concentrated and aliquoted before flash-freezing and storage at -80°C . By implementing these measures, a notable increase in yield was observed, reaching approximately 15 mg of purified protein per litre of bacterial culture. This represents a significant 66% increase in yield compared to the earlier batches. The optimised purification protocol for the production of unlabelled and ^{15}N -labelled samples of Pin1 is summarised in figure 39.

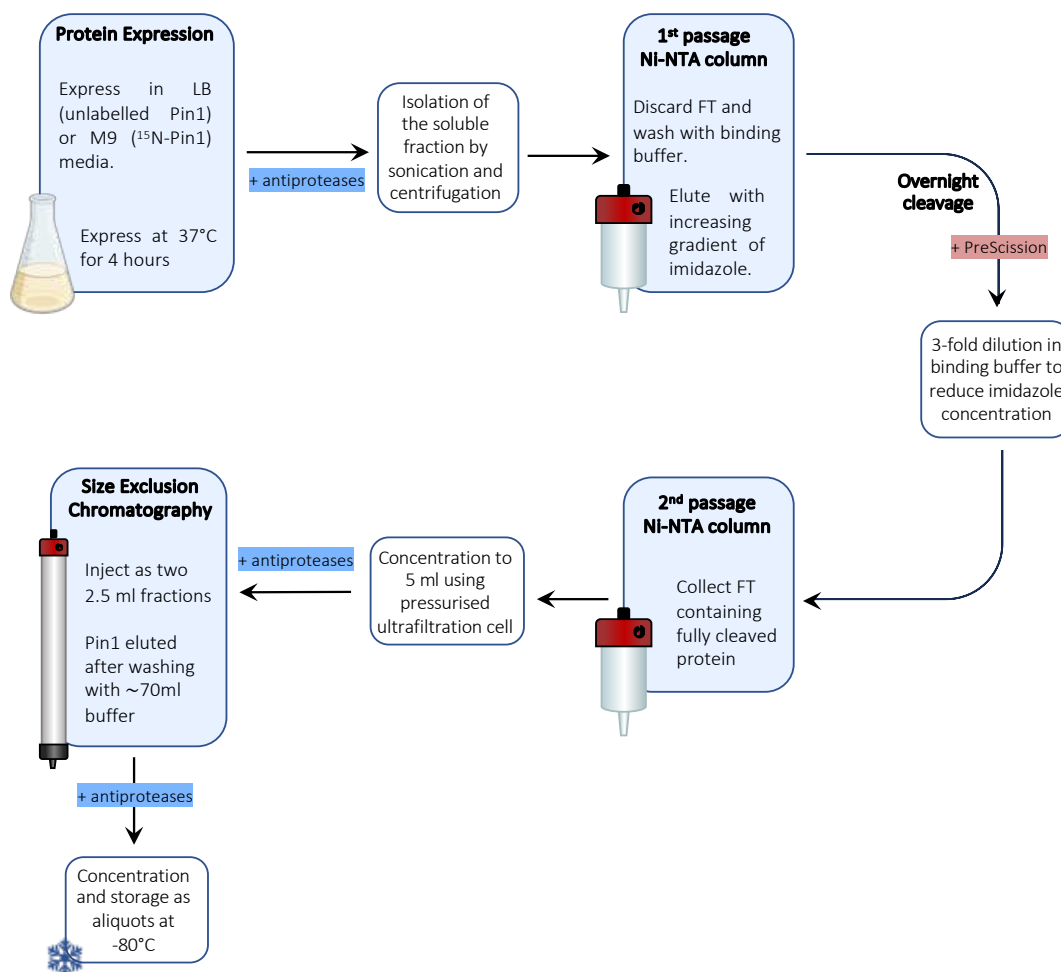


Figure 39: Flow chart summarising the steps to produce NMR samples of FL Pin1.

6.2 Expression and purification of the IDR of Cdc25C and its mutants

The production and purification of the *N*-terminal IDR of human Cdc25C (residues 1-259) had not been previously established, making its optimisation a time-consuming aspect of this thesis. A 185-residue fragment of the *N*-terminal IDR of Cdc25C was selected for study, comprising residues 44-228 (see figure 40a). This specific sequence was chosen due to its inclusion of all six pS/T-P motifs while maintaining a protein fragment less than 200 residues. Indeed, as this region is predicted to be intrinsically disordered, a typical weak dispersion of NMR signals in the proton dimension, leading to strong spectral crowding, was anticipated. To facilitate the expression and purification of the selected Cdc25C⁴⁴⁻²²⁸ sequence, a pGEX-6P-1 plasmid was designed to encode a 47 kDa fusion protein containing the selected Cdc25C⁴⁴⁻²²⁸

fragment (20.5 kDa) separated from an *N*-terminal GST protein (26.5 kDa) by a PreScission protease cleavage site (figure 40b). Incorporation of a GST fusion partner was favoured over a small (His)₆ tag as it is known to promote solubility of disordered proteins and to offer protection against protein degradation during expression. Additionally, the selected expression vector contained ampicillin resistance and was regulated by the lac operon. The pGEX-6P-1(Cdc25C⁴⁴⁻²²⁸) plasmid was introduced into competent *E. coli* BL21-Gold (DE3) PlyS cells. Successful transformants were selected based on their dual antibiotic resistance (ampicillin and chloramphenicol) phenotype and glycerol stocks were subsequently prepared and stored at -80°C to ensure long-term preservation.

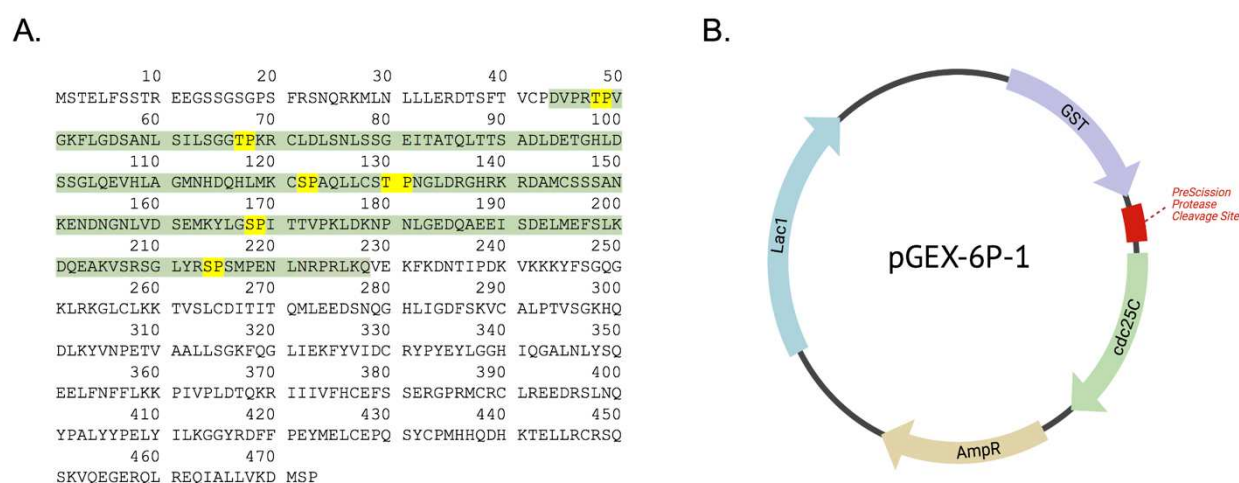


Figure 40: Selection of the Cdc25C⁴⁴⁻²²⁸ fragment for study (A) and expression vector design (B).

A. The sequence highlighted in green represents the 185-residue sequence of the Cdc25C IDR selected for study with the 6 S/T-P sites highlighted in yellow.

6.2.1. Expression of Cdc25C⁴⁴⁻²²⁸ in LB media

In order to optimise a protocol for the production of GST-Cdc25C⁴⁴⁻²²⁸ using a bacterial expression system, different conditions of protein expression were trialled. *E. coli* BL21-Gold (DE3) PlyS cells harbouring the pGEX-6P-1(Cdc25C⁴⁴⁻²²⁸) expression vector were grown as large-scale cultures in LB media supplemented with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. The cultures were allowed to grow at 37°C until reaching an OD₆₀₀ of 0.6-0.8. Protein expression was induced by the addition of 0.5 mM IPTG. The optimal temperature for protein expression was investigated by incubating cultures for either 4 hours at 37°C or 12 hours at 20°C after addition of IPTG. SDS-PAGE analysis of non-induced and induced cultures

confirmed that expression was successful in both cases, with the observation of high-intensity bands at ~50 kDa for the induced cultures (figure 41). The subsequent purification process was closely monitored for each trial to determine if either temperature generated a higher protein yield. After the final purification step, it was determined that protein expression at 20°C resulted in a 62.5% increase in yield compared to at 37°C (see table 7).

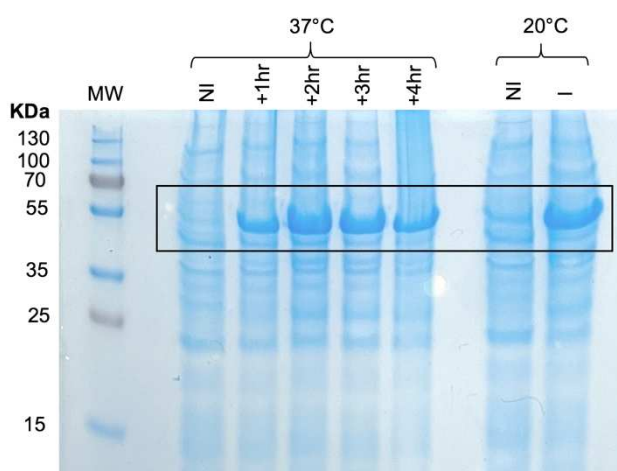


Figure 41: SDS-PAGE analysis of cultures after induction of GST-Cdc25C⁴⁴⁻²²⁸ expression at 37°C or 20°C. The black box highlights the band observed at 47 kDa corresponding to GST-Cdc25C⁴⁴⁻²²⁸.

Induction temperature (°C)	Yield (mg/L of LB)
37	8
20	13

Table 7: Comparison of protein yields (mg of purified protein per L of growth media) obtained after purification of Cdc25C⁴⁴⁻²²⁸ following induction at 37°C or 20°C.

6.2.2. Expression of Cdc25C⁴⁴⁻²²⁸ in M9 media

In order to investigate the structural properties of the Cdc25C⁴⁴⁻²²⁸ fragment using NMR, it was necessary to generate single (¹⁵N) and double (¹³C/¹⁵N) labelled samples of Cdc25C⁴⁴⁻²²⁸. Since the Marley protocol was successfully applied for the production of labelled Pin1 samples without effecting the yield, it was also employed for the production of single and double labelled Cdc25C⁴⁴⁻²²⁸ samples (Marley *et al*, 2001). *E. coli* cells containing the pGEX-6P-1 (Cdc25C⁴⁴⁻²²⁸) plasmid were grown at 37°C in LB media before transfer to a reduced volume of M9 media. Following a short period to allow for growth recovery and enable the consumption of residual LB media, the M9 media was supplemented with

isotopically-labelled nutrients ($^{15}\text{NH}_4\text{Cl}$ and ^{12}C - or ^{13}C -glucose). Protein expression was induced using 0.5 mM IPTG at 20°C for 12 hours. Finally, the harvested cells were stored at -80°C in Cdc25C binding buffer to maintain sample integrity prior to purification.

6.2.3. Purification of WT Cdc25C⁴⁴⁻²²⁸

Before proceeding with purification, solubility tests were conducted to determine the fractionation of the GST-Cdc25C⁴⁴⁻²²⁸ fusion protein following lysis and centrifugation. SDS-PAGE analysis confirmed the presence of GST-Cdc25C⁴⁴⁻²²⁸ in the soluble fraction (CFE) obtained after sonication and centrifugation (see figure 42).

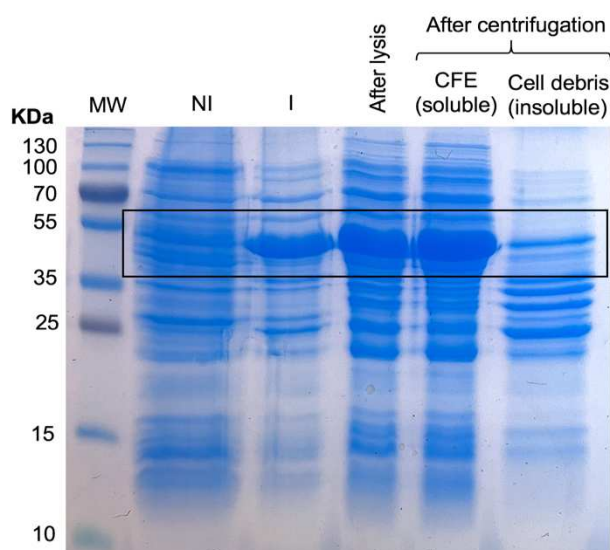


Figure 42: Solubility tests of GST-Cdc25C⁴⁴⁻²²⁸ after lysis and centrifugation of cells. The black box highlights the band observed at 47 kDa corresponding to GST-Cdc25C⁴⁴⁻²²⁸, confirming the construct is present in induced cells (I), cells after lysis and in the soluble fraction following subsequent centrifugation.

A three-step purification approach was employed to isolate Cdc25C⁴⁴⁻²²⁸, utilising affinity chromatography followed by gel filtration (summarised in figure 43). The affinity chromatography steps were carried out using a gravity-flow method. Initially, the CFE containing GST-Cdc25C⁴⁴⁻²²⁸ was incubated with a glutathione resin pre-equilibrated with Cdc25C binding buffer. This allowed the GST tag to bind to the resin during a 4-hour incubation period. Subsequently, the resin-CFE mixture was packed in a glass column and the FT fraction containing non-interacting proteins was collected. To remove weakly-interacting proteins, the resin was washed with buffer. Next, the resin was removed from the column and incubated at 4°C overnight with PreScission enzyme to facilitate an on-resin cleavage of the GST-tag. Following cleavage, the resin was reintroduced to the column, and the FT fraction

containing fully-cleaved Cdc25C⁴⁴⁻²²⁸ was collected. To regenerate the resin, it was treated with excess glutathione to elute the GST tag bound to the resin. Each purification step was monitored using SDS-PAGE analysis. As shown in Figure 44, overnight incubation with the PreScission enzyme resulted in the observation of two high-intensity bands with apparent molecular weights corresponding to the cleaved Cdc25C⁴⁴⁻²²⁸ fragment (20.5 kDa) in the FT fraction and the GST fusion partner (26.5 kDa) in the elution fraction. Furthermore, the GST-Cdc25C⁴⁴⁻²²⁸ fusion protein is no longer observed after overnight incubation with PreScission, indicating successful and complete cleavage.

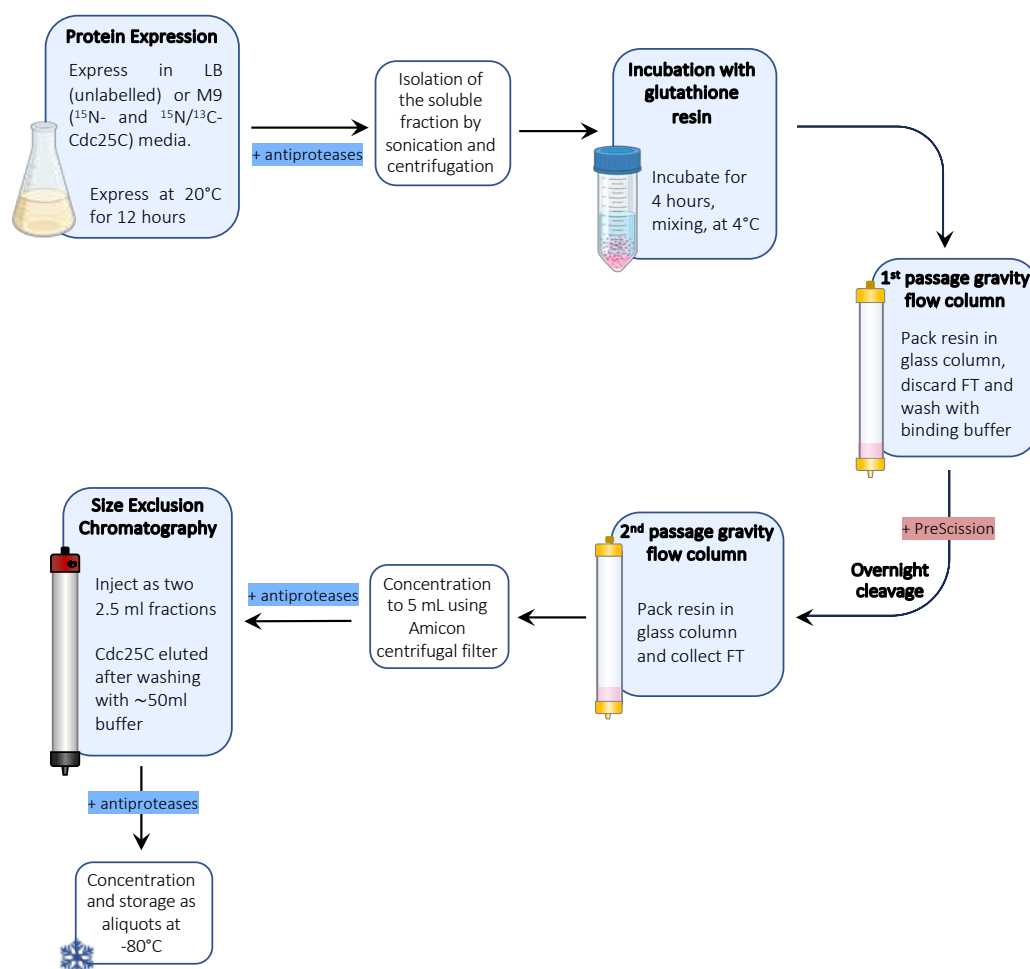


Figure 43: Flow chart summarising the steps to produce high-purity NMR samples of Cdc25C⁴⁴⁻²²⁸.

A final purification step was carried out using SEC by passing Cdc25C⁴⁴⁻²²⁸ through a Superdex 75 16/600 column using an AKTA purifier to enhance the purity and transfer the protein into appropriate buffer for each NMR experiment. The sample was first concentrated to 5 mL using a centrifugal filtration device with a 3 kDa MW cut-off before injection as two 2.5 mL fractions

into the column. The purification was monitored by chromatogram and SDS-PAGE analysis (see figure 45). Three distinct peaks were observed in the resulting gel filtration chromatogram, indicating the presence of three different species. SDS-PAGE analysis confirmed the presence of the Cdc25C⁴⁴⁻²²⁸ protein, indicated by a distinct band at approximately 20 kDa, specifically within the fractions of the second peak. The first peak, which partially overlaps with peak 2, seemingly contains a high molecular weight protein (~70 kDa). Notably, SDS-PAGE analysis of fractions of the third peak yielded no discernible bands. This could be attributed to the presence of nucleic acids, which absorb UV at 280 nm but are not visible on SDS-PAGE gels.

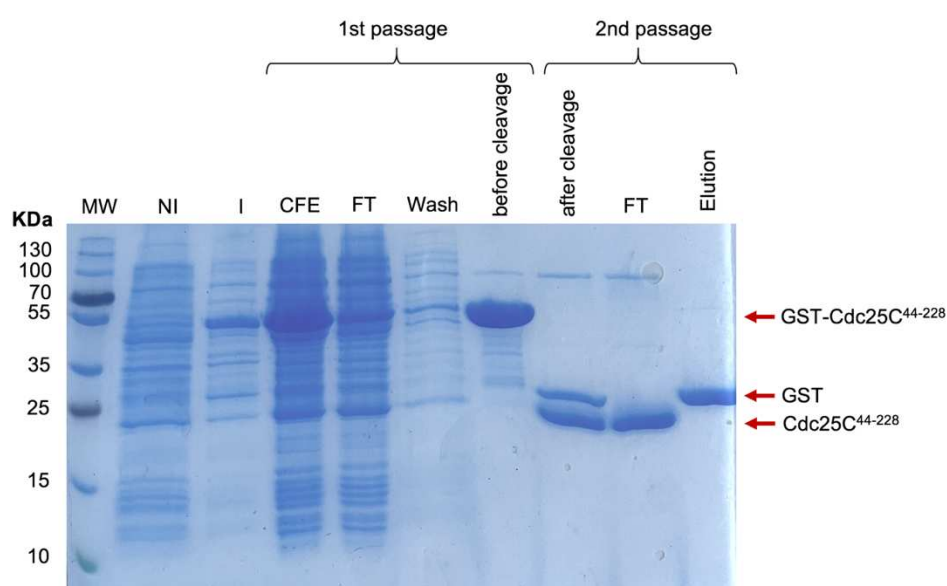


Figure 44: SDS-PAGE analysis of affinity chromatography steps in the purification of Cdc25C⁴⁴⁻²²⁸. The first two steps of purification of Cdc25C⁴⁴⁻²²⁸ from *E. coli* cells were monitored using SDS-PAGE analysis. The GST-tagged construct (47 kDa) is present in the induced cells (I), CFE and the sample bound to the glutathione resin before cleavage. After overnight incubation with PreScission enzyme, two bands representing GST (26.5 kDa) and Cdc25C⁴⁴⁻²²⁸ (20.5 kDa) are observed. Cdc25C⁴⁴⁻²²⁸ and the GST tag are separated by a second passage through the glutathione column, with Cdc25C⁴⁴⁻²²⁸ collected in the FT and GST eluted from the resin using excess glutathione.

MALDI-TOF analysis of an unlabelled sample was performed to validate the expression and purification protocol and verify the integrity of the purified Cdc25C⁴⁴⁻²²⁸ protein (figure 46). The analysis confirmed the presence of the full-length protein, with a measured molecular

weight that perfectly matched the expected value (20.5 kDa), providing a more accurate assessment of its molecular mass compared to SDS-PAGE.

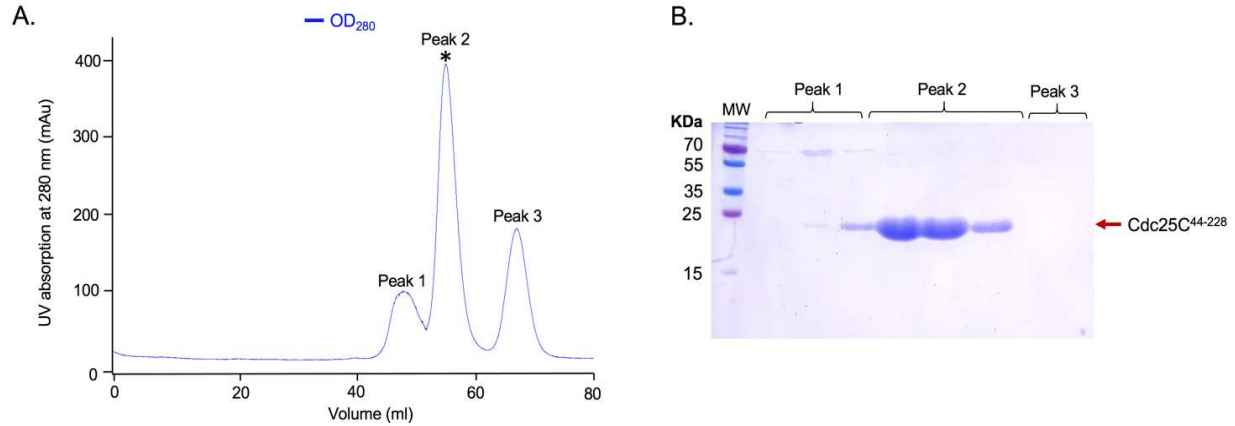


Figure 45: Chromatogram (A) and SDS-PAGE analysis (B) of the Cdc25C⁴⁴⁻²²⁸ purification by SEC.

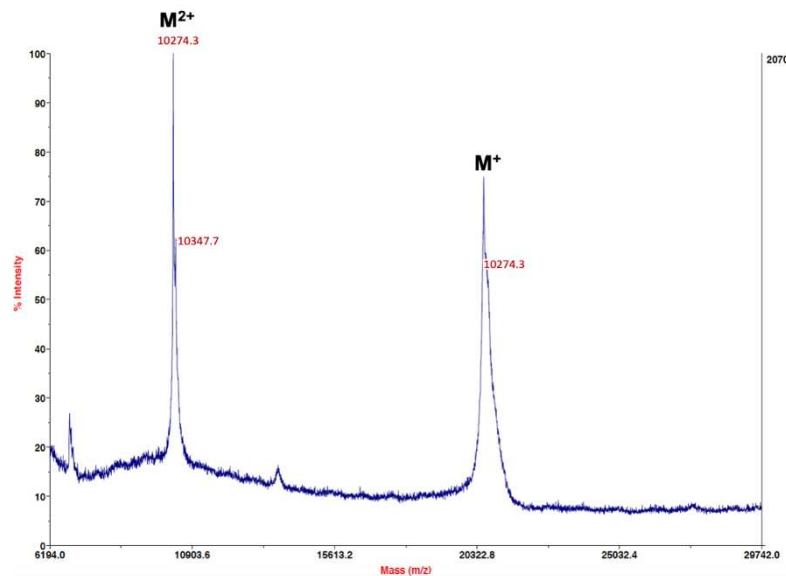


Figure 46: MALDI-TOF analysis of unlabelled Cdc25C⁴⁴⁻²²⁸. Two distinct peaks were observed corresponding to the singly charged Cdc25C⁴⁴⁻²²⁸ molecular ion (M⁺) and the doubly charged Cdc25C⁴⁴⁻²²⁸ (M²⁺) species. The M⁺ peak perfectly matches the expected mass (20.5 kDa) of Cdc25C⁴⁴⁻²²⁸.

Table 8 presents a comparison of the yields obtained for unlabelled, single ¹⁵N-labelled, and double ¹⁵N/¹³C-labelled Cdc25C⁴⁴⁻²²⁸ samples throughout the course of this thesis. The labelled samples produced in M9 media showed slightly lower average yields (5 to 10 mg/L of purified

protein per L of M9 medium) compared to the unlabelled protein (6 to 16 mg of purified protein per L of LB medium). However, it is important to highlight that the production yields varied across different batches, likely indicating inconsistencies in the expression of protein in *E. coli*. This variability can be attributed to the long-term storage of *E. coli cells* as glycerol stocks at -80°C , which can result in a decline in expression yields.

<i>Unlabelled Cdc25C⁴⁴⁻²²⁸</i>		<i>¹⁵N-Cdc25C⁴⁴⁻²²⁸</i>		<i>¹³C/¹⁵N-Cdc25C⁴⁴⁻²²⁸</i>	
Batch	Yield (mg/L of LB)	Batch	Yield (mg/L of M9)	Batch	Yield (mg/L of M9)
1	13	1	10	1	10
2	16	2	5.5	2	7
3	10	3	12	3	10
4	6	4	5	4	10
		5	7		

Table 8: Yields of unlabelled, single- and double-labelled Cdc25C⁴⁴⁻²²⁸ acquired throughout this thesis. Values are given in mg of purified protein per litre of culture medium.

6.2.4. Purification of phosphomimetic mutant of Cdc25C⁴⁴⁻²²⁸

In order to study the interaction of Pin1 with the IDR of Cdc25C, it was necessary to investigate its interaction with the six *phosphorylated* Ser/Thr-Pro motifs. As a first step before phosphorylation, a phosphomimetic mutant of Cdc25C⁴⁴⁻²²⁸ (pM-Cdc25C⁴⁴⁻²²⁸) was produced whereby each of the six serine or threonine residues were substituted with glutamate. The choice of glutamate as a phosphomimetic was based on its deprotonated state at physiological pH, which allows it to provide a partial electrostatic mimic of the native phosphate group, despite carrying a single negative charge rather than the double negative charge of the phosphate group. While still not a perfect mimic, glutamate was selected over aspartate due to its longer side chain, which offers a better structural mimic of the larger pSer/pThr groups. The mutant gene was inserted into the same pGEX-6P-1 expression vector by the Genscript company and expressed as a GST-fusion protein. The same expression and purification protocol was followed for pM-Cdc25C⁴⁴⁻²²⁸ as for the WT protein. The yields of unlabelled, ¹⁵N- and ¹⁵N/¹³C-labelled pM-Cdc25C obtained are compared in table 9. Notably, a significantly low yield of ¹⁵N-labelled pM-Cdc25C was obtained, which may be indicative of

challenges related to protein degradation during purification, such stability issues are often observed when working with IDPs.

Yield (mg/L of media)		
Unlabelled	¹⁵ N-labelled	¹³ C/ ¹⁵ N-labelled
11.5	3.5	11

Table 9: Yields of unlabelled, single- and double-labelled pM-Cdc25C⁴⁴⁻²²⁸ acquired throughout this thesis. Values are given in mg of purified protein per litre of culture medium.

6.3 PreScission enzyme

The purification protocols of both Pin1 and Cdc25C relied on the cleavage of their respective tags by PreScission enzyme. This enzyme was produced and purified in-house for these purposes. A pET-15b plasmid encoding a fusion protein consisting of the PreScission fragment and an *N*-terminal (His)₆ tag was already available in the lab. First, more plasmid DNA was prepared using a miniprep plasmid preparation kit to obtain 100 µg plasmid DNA. Once sufficient plasmid DNA was obtained, the pET-15b(PreScission) plasmid was introduced into *E. coli* BL21-Gold (DE3) PlysS competent cells and transformants were selected thanks to their dual antibiotic resistance. For the subsequent productions, a fresh glycerol stock of transformed cells was made and stored at -80°C.

6.3.1. Expression of PreScission enzyme in LB media

An expression protocol for the production of PreScission enzyme had already been established by the lab. *E. coli* cells harbouring the pET-15b(PreScission) plasmid were grown as large-scale cultures in LB media supplemented with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. The cultures were allowed to grow at 30°C until reaching an OD₆₀₀ of 0.6-0.8. Protein expression was induced by the addition of 0.5 mM IPTG, followed by incubation at 20°C for 12 hours. Subsequently, the cells were harvested through centrifugation.

6.3.2. Purification of PreScission enzyme

First, the cell pellets obtained from centrifugation of cultures were lysed by sonication and the soluble CFE containing PreScission enzyme was isolated from the cell debris using centrifugation. Active PreScission enzyme was then obtained in a single affinity chromatography purification step (figure 47). The CFE was passed through an Ni-NTA His-trap column equilibrated with PreScission binding buffer and the FT containing non-interacting proteins collected. The column was then washed with buffer to remove weakly interacting proteins. The (His)₆-PreScission construct was eluted from the column using a gradient of increasing imidazole concentration from 0 mM to 192 mM during 40 minutes.

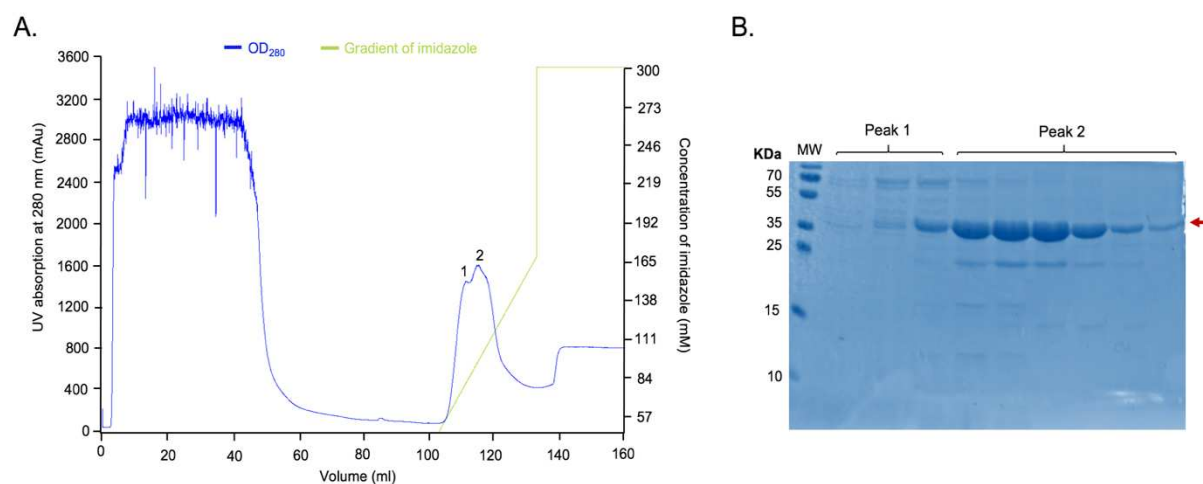


Figure 47: Chromatogram (A) and SDS-PAGE analysis (B) of the passage of PreScission enzyme through the Ni-NTA column. PreScission enzyme was eluted from the column at imidazole concentration of ~100 mM (peak 2) as identified by SDS-PAGE analysis whereby a peak at ~46 KDa (highlighted by the red arrow) is observed in fractions collected of peak 2.

Although the purity of the enzyme was not complete, it was sufficient for our needs, as this enzyme is used for cleavage purposes only and not for analysis. The resulting protein was stored as glycerol stock aliquots allowing for long-term storage and availability of PreScission enzyme for use in the purification of Pin1 and Cdc25C⁴⁴⁻²²⁸.

Chapter 7 Characterising the binding and activity of novel peptide ligands of Pin1

The first-generation series of fluorinated peptide Pin1 ligands provided valuable insights into key binding features and facilitated the development of domain-specific ligands. However, a significant limitation of the first-generation series was their affinity being in the tens of micromolar range, falling short of the desired sub-micromolar affinity. Additionally, their inhibitory potential remained largely unexplored. In light of these limitations, the present study aimed to address the shortcomings of the first-generation ligands by generating a new series of ligands displaying a higher potency and optimising an activity assay to assess their inhibitory potential.

The design of the second-generation ligands drew inspiration from the first-generation ligands, incorporating essential groups for affinity identified in the earlier series into a pseudo-tetrapeptide backbone. All ligands were synthesised by collaborators at the BioCis laboratory of Cergy-Pontoise. To overcome the issue of potency, two distinct strategies were pursued. Firstly, in the context of the ANR F-LAIR project, to further develop the fluorinated peptide series, different methods were trialled to increase their potency and cell-permeability (CZ279, CZ280, CZ281, CZ282, GG82, GG84, VB66, GG156, GG168, GG201, GG213 and GG199). Secondly, a new strategy for Pin1 inhibition was explored based on PROTAC (Proteolysis-Targeting Chimeras) technology (CZ360, CZ282, LME27, LME51, AGI24, LME66) and supported by ANR PRODIGE. In this approach, instead of catalytic inhibition, Pin1 would be targeted for proteolytic degradation within the cell, whereby micromolar affinity may be adequate to restore Pin1 levels to normal in cancerous cells.

In order to test the inhibitory potential of compounds of the first and second-generation ligands, an activity assay for prolyl-isomerisation by Pin1 first required optimisation. The literature points towards PPIase assays based on fluorescence, UV spectroscopy or NMR EXSY experiments and NMR lineshape analysis (Greenwood *et al*, 2011). Previously, fluorescence-based assays performed in the presence of ligands of the first-generation series yielded inconsistent results, hence here, the focus was turned towards optimising UV- and NMR-based

assays, aiming to provide more reliable and informative data for evaluating the inhibitory activity of the ligands.

7.1 The binding of Pin1 to the 2nd generation of peptide ligands

7.1.1. Quantifying ligand affinity: NMR-based K_D determination

Studying the interaction between Pin1 and the second-generation ligands by NMR first required the assignment of the backbone resonances of the protein in the free form. Assignment of backbone and $^{13}\text{C}\beta$ chemical shifts of apo Pin1 in phosphate buffer, pH7, has previously been published and deposited in the Biological Magnetic Resonance dataBank (BMRB) (entries 5305, 11557, 27579) (Born *et al*, 2019; Jacobs *et al*, 2002; Xu *et al*, 2014). The assignment of 144 out of 155 non-proline residues was previously confirmed by the lab upon performing a series of 3D assignment experiments using a 500 MHz spectrometer equipped with a cryoprobe. Notably, residues Gly39 – Gly45 of the disordered linker region, could not be assigned, likely due to the fast exchange of amide protons with water under these conditions. In addition, amide group resonances corresponding to residues Arg17, Ser18, and Ser19 of loop 1 of the WW domain, and residue Ser114 of the PPIase domain could not be assigned in the ^1H - ^{15}N HSQC spectrum of Pin1 at pH 7, 298K. For interaction studies, the phosphate buffer conditions posed potential problems in interfering with ligand binding, hence an additional assignment of the backbone amide groups of Pin1 was also performed by the lab in the presence of Tris buffer (see figure 48). Unfortunately, the transfer to Tris buffer resulted in the line broadening of three additional amide group resonances, which belong to the phosphate-binding site of the PPIase domain (Arg68, Arg69, Ser71) (see chapter 3.3). Nevertheless, interaction studies were pursued under these conditions.

To assess the interaction of Pin1 with the novel series of peptide ligands, proton-nitrogen chemical shift perturbations (CSPs) were monitored throughout series of ^1H - ^{15}N HSQC experiments acquired during titration of ^{15}N -labelled Pin1 with different unlabelled ligands. Tracking backbone amide CSPs is a powerful method for studying protein-protein or protein-ligand interactions at the single residue level as it can provide accurate information about the key residues involved in binding. Chemical shift values of atomic nuclei are highly sensitive to their local chemical environment, which is influenced by neighbouring atoms whether covalently bonded or close in space. Consequently, the formation of a non-covalent

biomolecular complex, such as a Pin1:ligand complex, is expected to induce changes in chemical shifts of nuclei located at the interaction interface. This approach additionally enables the identification of protein regions that undergo distant allosteric structural rearrangements in response to ligand binding. Hence, by analysing the CSPs, valuable information can be obtained regarding the specific molecular interactions occurring between Pin1 and the peptide ligands.

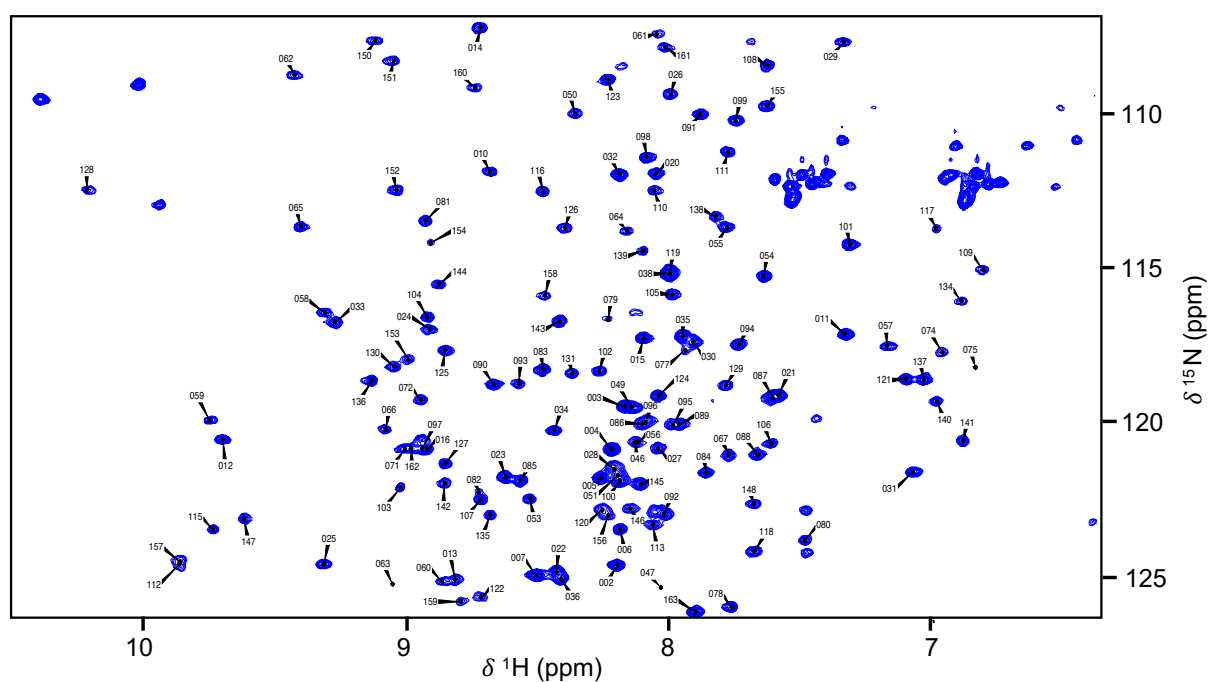
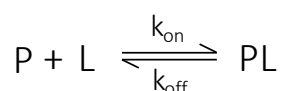


Figure 48: Assigned ^1H - ^{15}N -HSQC spectrum of apo Pin1 in Tris buffer (pH 7) recorded at 500 MHz, 298K. The label of each amide cross-peak corresponds to the residue number of Pin1.

In addition to elucidating the regions of Pin1 involved in ligand binding, CSP values can be a valuable tool for quantitatively determining the apparent dissociation constant (K_{Dapp}) of each residue implicated in the binding interaction (Williamson, 2013). These K_D values can then be used to estimate the binding affinity of ligands to the WW and PPIase domains by calculating the average K_D values of residues within each domain. To define the K_D , it is important to consider the underlying equilibrium of ligand binding. In a simple case of reversible binding between a protein and ligand, the equilibrium can be expressed as follows:



The association rate constant (k_{on}) is used to characterise the rate at which the protein (P) and ligand (L) combine to form the protein-ligand complex (PL) and is expressed in units of per

moles per second ($M^{-1}s^{-1}$). A higher k_{on} value signifies a more rapid association. On the other hand, the dissociation rate constant (k_{off}), is used to describe the rate of the backward reaction and is typically expressed in units of per second (s^{-1}). A higher k_{off} value indicates a faster dissociation rate. At equilibrium, the system is described by the dissociation constant, K_D , which is denoted as:

$$K_D = \frac{[P][L]}{[PL]} = \frac{k_{off}}{k_{on}}$$

Both equations for K_D calculation provide a quantitative measure of the ligand affinity and the equilibrium state of the protein-ligand interaction, with a lower K_D value reflecting stronger binding. In the simple case of diffusion-controlled binding, where the association of the ligand and protein is not hindered by additional factors such as conformational changes, the binding process is characterised by a rapid association dependent solely on the availability of binding sites and the collision frequency of the molecules (Williamson, 2013). This type of binding is characterised by a high k_{on} value (typically around $10^9 M^{-1}s^{-1}$) indicating fast association between the protein and ligand (Fersht, 2017). When K_D is expressed in M, this gives an approximate relationship between k_{off} and K_D , whereby:

$$k_{off} \cong k_{on} \cdot K_D$$

$$k_{off} \cong 10^9 \cdot K_D$$

This approximation of k_{off} has important implications for experimental measurements of K_D . In the context of NMR spectroscopy, the exchange regime refers to the relative rates of ligand binding and dissociation compared to the timescale of chemical shift measurements. The exchange rate is important to consider as it influences how the spectral peaks in NMR spectra change during ligand titration experiments. For instance, when the exchange rate is fast on the chemical shift timescale this means that k_{off} is significantly greater than the difference in Hz between the chemical shifts of free and bound protein. In this case, the HSQC spectrum of a protein as the ligand is titrated in, exhibits the signals smoothly transitioning from their positions in the free spectrum to those in the bound spectrum (see figure 49). The frequency of

the signal at any titration point represents the weighted average of the free and bound chemical shift values. Conversely, in the case of slow exchange, whereby k_{off} is much slower than the difference in Hz between the shifts of the free and bound protein, the signals exhibit the gradual disappearance of the free signal and the appearance of the bound signal. The intensities of these two peaks at each titration point reflect the concentrations of free and bound protein (see figure 49). In intermediate exchange, with the k_{off} value similar to the shift difference, the signals broaden and shift at the same time.

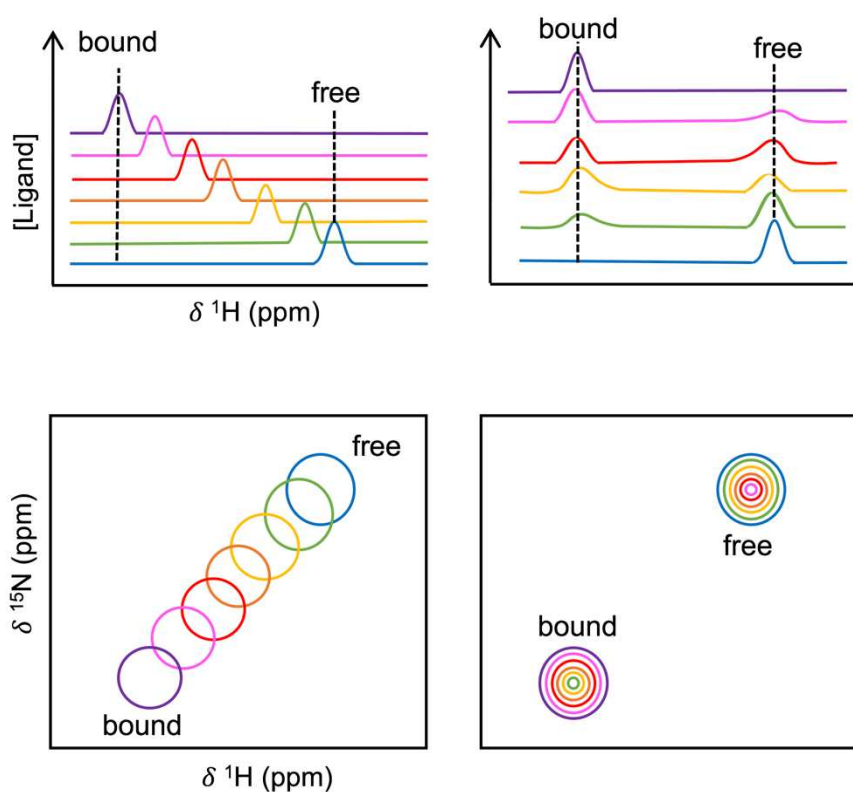


Figure 49: The dependence of 1D (upper) and 2D (lower) NMR spectra on the exchange rate. In fast exchange (left) as ligand concentration is increased, peaks move smoothly between free (blue) and bound (purple) forms. In slow exchange (right), as ligand concentration is increased, the intensity of the bound (purple) peak increases as the intensity of the free (blue) peak decreases.

In the case of Pin1, the analysis of ^1H - ^{15}N HSQC experiments recorded during titration with increasing concentrations of ligand provided insights into the exchange regime of ligand binding. The obtained spectra exhibited a smooth transition of signals from the free to the bound state at higher ligand concentrations, indicating a fast exchange regime (figure 50).

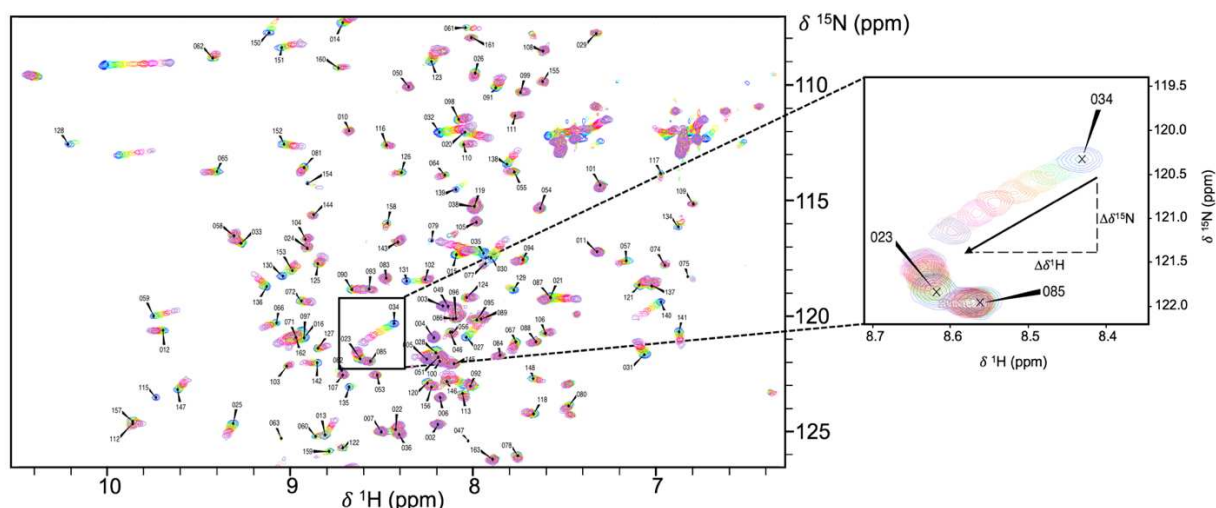


Figure 50: Overlay of nine ^1H - ^{15}N HSQC spectra recorded on ^{15}N Pin1 in the presence of increasing concentrations of ligand at 298K on a 500 MHz spectrometer equipped with a cryoprobe. The titration of increasing concentrations of peptide ligands (LME27 shown here) into a ^{15}N -labelled sample of Pin1 displayed a fast exchange regime of ligand binding. The label of each cross-peak corresponds to the residue number of Pin1. An expansion is shown for residue 34 where the direction of peak movement from the free form (blue) to the bound form (purple) is indicated by an arrow.

In the case of the fast exchange regime observed here, the observed chemical shift change (δ_{obs}) resulting from binding can be tracked as the weighted average of the shifts in the free (δ_f) and bound (δ_b) states:

$$\delta_{\text{obs}} = \delta_f f_f + \delta_b f_b$$

Whereby f_f and f_b represent the fractions of free and bound protein respectively, indicating the relative populations of the free and bound states in equilibrium. The relationship between f_f and f_b is governed by the condition that their sum equals 1, as indicated by the equation:

$$f_f + f_b = 1$$

This means that:

$$\delta_{\text{obs}} = \delta_f (1 - f_b) + \delta_b f_b$$

$$\delta_{\text{obs}} = \delta_f + f_b (\delta_b - \delta_f)$$

This can be used to calculate the overall CSP ($\Delta\delta_{\text{obs}}$) as:

$$\Delta\delta_{\text{obs}} = \Delta\delta_{\text{max}} \cdot f_b$$

Given that $\Delta\delta_{\text{obs}}$ and $\Delta\delta_{\text{max}}$ are defined by:

$$\Delta\delta_{\text{obs}} = \delta_{\text{obs}} - \delta_{\text{f}}$$

$$\Delta\delta_{\text{max}} = \delta_{\text{b}} - \delta_{\text{f}}$$

During the titration, the total concentrations of protein $[P]_{\text{t}}$ and ligand $[L]_{\text{t}}$ are known and can be described by the sum of the free and bound forms:

$$[P]_{\text{t}} = [P] + [PL]$$

$$[L]_{\text{t}} = [L] + [PL]$$

f_{b} can be expressed by:

$$f_{\text{b}} = \frac{[PL]}{[P]_{\text{t}}}$$

As a result:

$$\Delta\delta_{\text{obs}} = \Delta\delta_{\text{max}} \cdot \frac{[PL]}{[P]_{\text{t}}}$$

Furthermore, since the definition of K_{D} is known, it follows that:

$$K_{\text{D}} = \frac{[P][L]}{[PL]}$$

$$K_{\text{D}}[PL] = [P][L]$$

$$K_{\text{D}}[PL] = ([P]_{\text{t}} - [PL])([L]_{\text{t}} - [PL])$$

Thus, the following quadratic equation is obtained:

$$[PL]^2 - ([P]_{\text{t}} + [L]_{\text{t}} + K_{\text{D}}) \times [PL] \times [P]_{\text{t}} \times [L]_{\text{t}} = 0$$

By solving this equation,

$$[PL] = \frac{([P]_{\text{t}} + [L]_{\text{t}} + K_{\text{D}}) - \sqrt{([P]_{\text{t}} + [L]_{\text{t}} + K_{\text{D}})^2 - 4[P]_{\text{t}}[L]_{\text{t}}}}{2}$$

And hence, $\Delta\delta_{\text{obs}}$ can finally be defined as follows:

$$\Delta\delta_{\text{obs}} = \Delta\delta_{\text{max}} \times \frac{([P]_t + [L]_t + K_D) - \sqrt{([P]_t + [L]_t + K_D)^2 - 4[P]_t [L]_t}}{2 [P]_t}$$

To calculate the CSP ($\Delta\delta_{\text{obs}}$) from the experimental data recorded during a ligand titration, it is advantageous to take into account the chemical shift perturbations of both ^1H and ^{15}N amide nuclei for each peak to provide a comprehensive measure of the overall perturbation experienced by the amide groups upon ligand binding. To do this, an average amide CSP is calculated by introducing weighting factors, hence the ^{15}N chemical shift changes are reduced by a factor 10 (as the variation in the ^{15}N chemical shift is 10 orders of magnitude larger than that of the proton). The Euclidean distance is then given by:

$$\Delta\delta_{\text{obs}} = \sqrt{\left(\frac{(\delta_b - \delta_f)_{^{15}\text{N}}}{10}\right)^2 + \left((\delta_b - \delta_f)_{^1\text{H}}\right)^2}$$

A CSP curve can be established for each ligand by plotting $\Delta\delta_{\text{obs}}$ as a function of $[L]_t/[P]_t$ (representing the number of molar equivalents added). The titration is carried out until the protein is saturated, enabling the generation of experimental CSP curves (Figure 51). Only residues with significant perturbations exceeding the threshold value of $m + \sigma$ (where m is the average and σ is the standard deviation of the chemical shift change) are considered for quantitative analysis. Subsequently, a Python algorithm is employed to fit the experimental data to the above equation, enabling the calculation of K_D for each residue. The average K_D value for the ligand binding to each domain of the protein is obtained by calculating the mean of these individual K_D values. The K_D values of the second-generation series of peptide ligands for each domain of Pin1 are summarised in table 10. The specific targeting strategies explored and their effects on binding to Pin1 will be dissected in detail in the following sections.

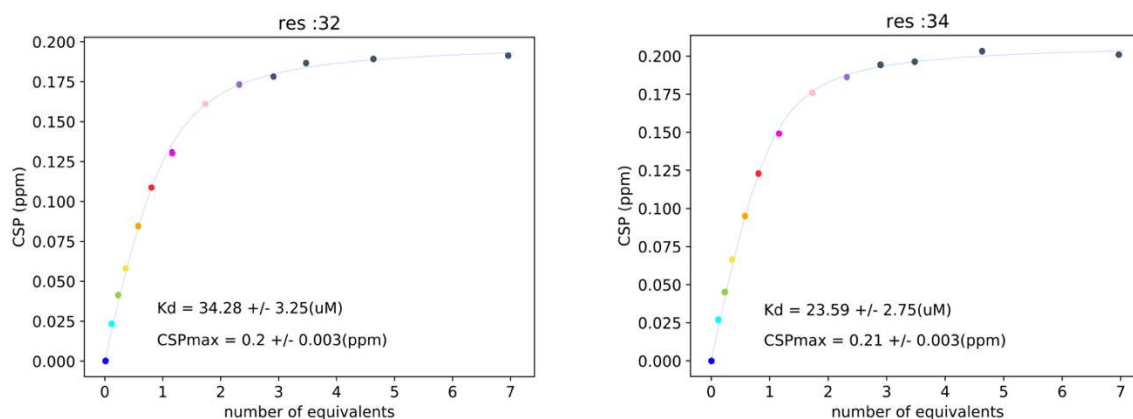


Figure 51: Experimental CSP curves obtained for two residues of Pin1 by plotting CSP as a function of $[L]_t/[P]_t$ (number of equivalents). Curves displayed for residues 32 and 34 upon binding to up to 7 equivalents of ligand LME27.

Series	Ligand name	Structure	Kd (μM)	
			WW	PPIase
Fluorinated	CZ279	Phe-Pro-Tryptamine	NB	NB
	CZ280	Phe- $\Psi^{\text{CF}_3, \text{H}}$ Pro-Tryptamine	NB	NB
	CZ281	Ac-Phe-Pro-Tryptamine	NB	NB
	CZ282	Ac-Phe- $\Psi^{\text{CF}_3, \text{H}}$ Pro-Tryptamine	NB	NB
	GG82	Ac-pyroGlu- $\Psi^{\text{CF}_3, \text{H}}$ Pro-Tryptamine	NB	NB
	GG84	Ac-Glu- $\Psi^{\text{CF}_3, \text{H}}$ Pro-Tryptamine	NB	NB
	GG156	Ac-p CF_2 -Ser-Pro-Tryptamine	291 \pm 50	436 \pm 137
	VB66	Ac-Cha-pSer- $\Psi^{\text{CF}_3, \text{H}}$ Pro-(Aminomethyl)piperidine	NB	386 \pm 72
	GG168	Ac-pSer-Pro-Nal-Gly-NH ₂	221 \pm 28	315 \pm 56
	GG201	Fmoc-pSer-Pro-Nal-Gly-NH ₂	25 \pm 5	135 \pm 23
PROTAC	CZ360	Ac-pSer-Pro-(O-propargyl)-serotonin	439 \pm 62	705 \pm 302
	LME27	FmocPra-pSer-Pro-Tryptamine	44 \pm 6	88 \pm 16
	LME51	AcPra-pSer-Pro-Tryptamine	285 \pm 32	598 \pm 193
	AGI24	PseudoFmocPra-pSer-Pro-Tryptamine	66 \pm 12	80 \pm 32

Table 10: The K_D values for the second-generation library of ligands calculated for each domain of Pin1. NB indicates no binding.

7.1.2. Fluorinated series of Pin1 ligands

Building upon the first-generation series of fluorinated peptide ligands of Pin1, a second-generation series was synthesised with the aim of investigating alternative targeting strategies to improve potency. Furthermore, efforts were made to optimise the compounds for their potential application as drug molecules by exploring approaches to streamline their synthesis and/or enhance their pharmacokinetic and pharmacodynamic properties. The strategies explored in the second-generation and their consequences on binding to Pin1 will be dissected in more detail below.

Removal of the phosphate group

One of the primary strategies investigated in the second generation of fluorinated ligands focused on evaluating the impact of removing the phosphate group on the ligand's affinity for Pin1. The presence of a phosphate group in drug molecules often leads to decreased cell permeability due to its 2⁻ negative charge at physiological pH, a problem encountered by many high affinity Pin1 peptide inhibitors reported in the literature (Wildemann *et al*, 2006; Zhang *et al*, 2007). Moreover, the incorporation of the phosphate group adds complexity to the synthesis of the ligands. Thus, the objective was to determine if sufficient affinity could be achieved without the phosphate group, making the ligands more amenable for use as drug compounds. More specifically, it was explored whether the presence of *N*-terminal and *C*-terminal aromatic residues was sufficient for binding to Pin1 in the absence of the pSer residue. To assess this, a series of four ligands (CZ279, CZ280, CZ281, and CZ282) were synthesised, where the pSer group was eliminated, and the *C*-terminal was capped with a tryptamine group, which had been shown to enhance binding in the first-generation ligands. Tryptamine, acting as a metabolically stable mimic of tryptophan, satisfied the *C*-terminal aromatic amino acid preference for Pin1 affinity. Phenylalanine was chosen as the *N*-terminal aromatic residue, similar to the Cha (cyclohexane) moiety present in the highest affinity fluorinated ligand of the first generation (CZ115) (see chapter 5). At position P2 the ligands included either a proline residue (CZ279 and CZ281) or a $\Psi^{\text{CF}_3, \text{H}}$ Proline residue (CZ280 and CZ282). Two of the ligands additionally incorporated an *N*-terminal acetyl group (CZ281 and CZ282) to mimic the peptide bond, resulting in pseudo-tetrapeptides. Each of the ligands encountered solubility issues due to their predominantly hydrophobic nature, rendering them soluble only in 100% dimethyl sulfoxide (DMSO) at 3 mM. For titration experiments, a total of 150 μL of the 3 mM stock solution was introduced into the NMR tube, which had a starting volume of 500 μL , in order to achieve a

final concentration of 6 molar equivalents of the ligand. Consequently, the DMSO concentration reached 23% in the final solution. Upon performing the four NMR titrations, weak CSPs were observed in proline binding sites of both domains (residues 23-35 and residues 127-135) (see chapter 3.3). Surprisingly, additional residues, including Arg54, Val55, Glu87, and Gly99, which are not typically associated with the canonical interaction sites of Pin1, were also affected. It is worth noting that all observed CSPs exhibited a linear pattern without reaching saturation, suggesting the lack of a specific interaction or very weak affinity. Furthermore, almost every residue of Pin1 showed minor perturbations, which is unusual for a specific binding event. Therefore, it was hypothesised that these perturbations were influenced by the titration of DMSO into the Pin1 sample, rather than the ligand itself. The effect of DMSO on protein folding and ligand interactions has been well-documented, indicating that DMSO concentrations as low as 3% can destabilise proteins, with concentrations above 40% commonly leading to complete denaturation or aggregation (Arakawa *et al*, 2007; Chan *et al*, 2017). To confirm if the observed CSPs were a consequence of DMSO influence, a control DMSO titration was performed, revealing weak linear perturbations at almost every residue as DMSO concentration increased. A comparison between the DMSO control titration and the ligand titrations (CZ279, CZ280, CZ281, and CZ282) revealed that the same regions of Pin1 exhibited perturbations in both cases (see Figure 52). Overall, this observation indicated that no significant ligand-induced CSPs were observed, suggesting the absence of a strong interaction between Pin1 and these ligands. Consequently, the removal of the pSer residue proved to be detrimental to the ligand's affinity, despite the presence of *N*-terminal and *C*-terminal aromatic groups.

In another attempt to explore phosphate group removal, a phosphomimetic ligand was synthesised, GG84 (Ac-Glu- $\Psi^{\text{CF}_3, \text{H}}$ Pro-Tryptamine), whereby the canonical phospho-serine residue at position P2 was replaced by a glutamate. Glutamate is commonly used as a structural and electrostatic mimic of phospho-serine, although its sidechain carries a 1⁻ charge compared to the 2⁻ charge of the phosphate group. Furthermore, a second ligand, GG82, was investigated, which was unintentionally generated as a by-product during the synthesis of GG84. The *N*-terminal deprotection of GG84 led to the cyclisation of the glutamate moiety, resulting in the formation of a pyroglutamic group in GG82 (Ac-pyroGlu- $\Psi^{\text{CF}_3, \text{H}}$ Pro-Tryptamine). This unexpected compound was also included in the evaluation to assess its potential binding properties. No CSPs were observed upon titration of 6 molar equivalents of either ligand into a

^{15}N -labelled sample of Pin1, indicating no interaction. Overall, these results demonstrate that the presence of the phosphorylated serine residue is crucial for the binding of ligands to Pin1.

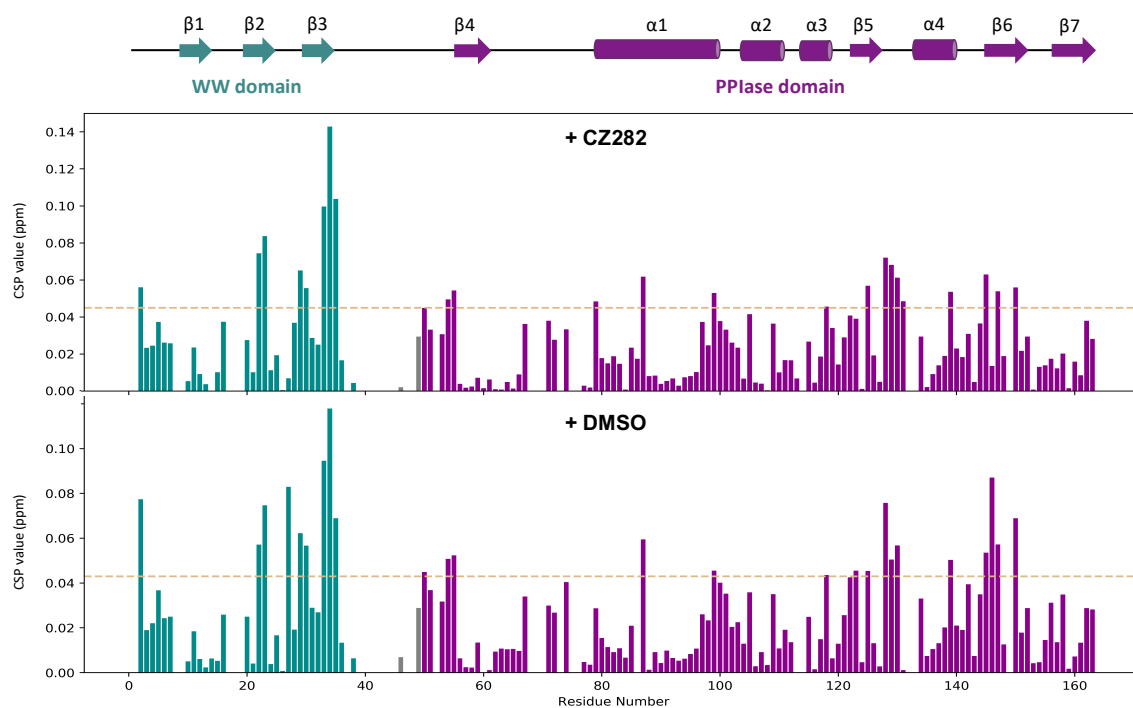


Figure 52: CSP analysis of Pin1 in the presence of CZ282 or DMSO control (lower). The upper panel shows the CSP chart of Pin1 upon titration with 6 molar equivalents of ligand CZ282 (in 100% DMSO), resulting in a final DMSO concentration of 23%. The lower panel represents the CSP chart obtained from a control experiment reaching 23% DMSO in the absence of ligand. Remarkably, the same regions of Pin1 exhibited perturbations in both cases, suggesting that these perturbations were caused by the presence of DMSO rather than ligand binding.

Production of CF_2 -phosphonate ligands

As efforts to remove or replace the phosphate group in the ligands proved unsuccessful, a novel approach involving fluorine incorporation via a CF_2 -phosphonate group, which serves as an alternative phosphoserine mimetic, was explored to enhance the ligands' metabolic properties (see figure 53). Difluoro-phosphonate and phosphonate groups hold the potential to improve the metabolic stability of the phosphate moiety. The previous findings demonstrated the essential role of the phosphate group for Pin1 binding *in vitro*. However, even a promising inhibitor *in vitro* can give disappointing results *in vivo* due to the hydrolysis of phosphate groups by cellular phosphatases. To address this issue, chemists have sought to replace phosphate

groups with bioisostere groups that resist hydrolytic cleavage in biological environments. One approach has been to replace the oxygen atom linking the phosphate group to the peptide with a carbon atom to obtain a phosphonate. Phosphonates are phosphate bioisosteres which are resistant to hydrolytic cleavage in cells. Initially, a major drawback of these compounds was their higher pKa values compared to native phosphate groups which can in turn affect their ionisation state *in vivo* and hence impact their binding properties. However, the introduction of fluorine atoms has been found to significantly reduce the pKa value, leading to CF₂-phosphonates that are isopolar to the phosphate group at neutral pH. Therefore, the incorporation of the CF₂-phosphonate group into Pin1 ligands presents a promising avenue for the development of metabolically stable compounds which maintain high affinity.

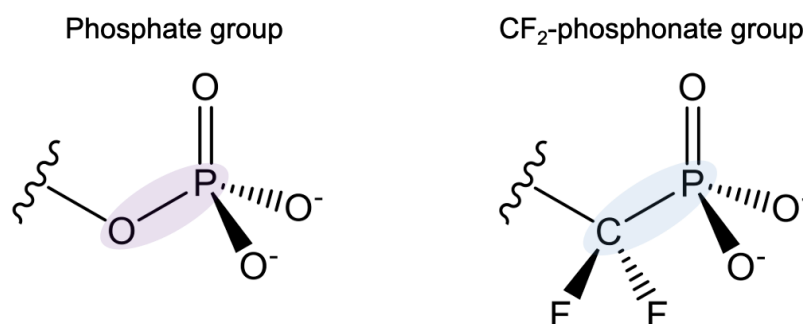
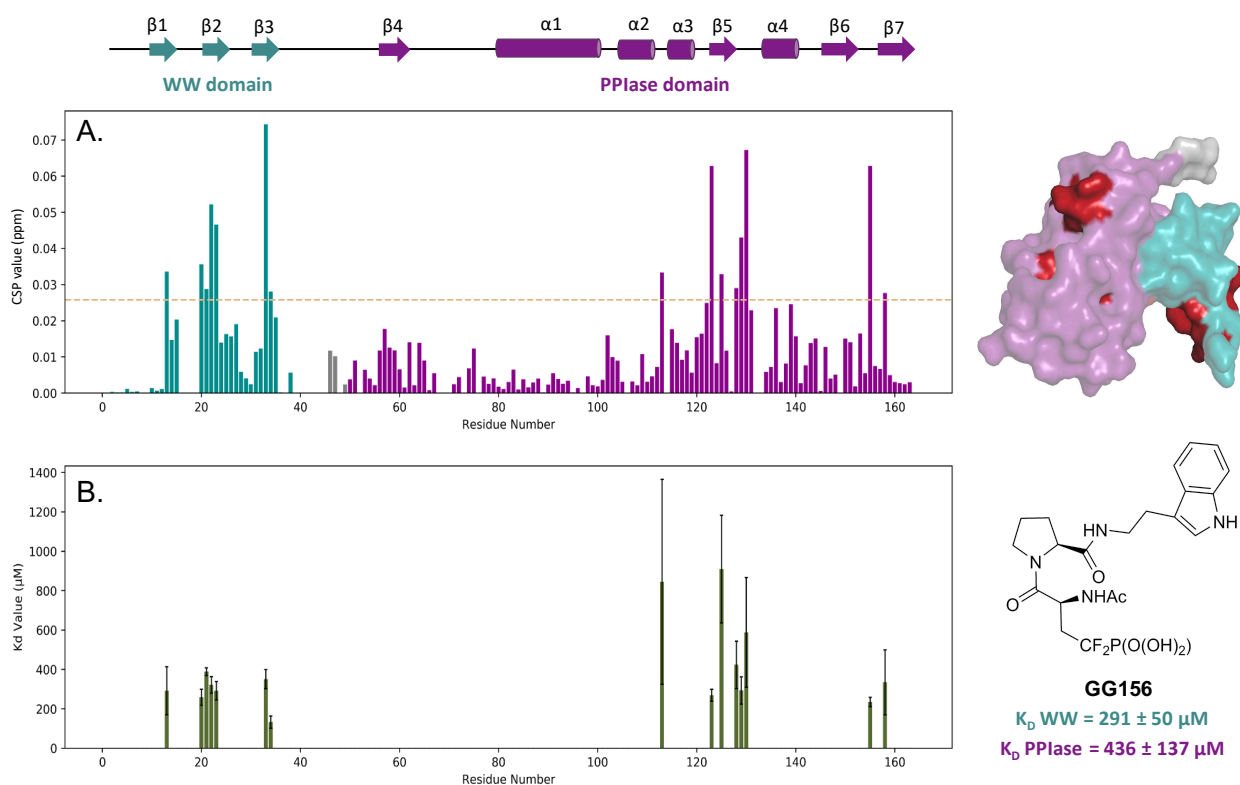


Figure 53: The structure of the phosphate group in comparison with the CF₂-phosphonate group.

A series of ligands were designed to incorporate this group at the P2 position, whereby the pSer residue was replaced by difluoro-phosphonated serines (CF₂-Ser). A tryptamine capped the C-terminus in all ligands, while strategies for *N*-terminal protection using acetyl or Fmoc were envisioned. The residue at position P3 was designed either with a regular proline residue or a Ψ^{CF₃,H}Proline. While four ligands were originally envisaged for this series, only two were synthesised by our collaborators due to time constraints. This led to the production of GG156 (Ac-pCF₂-Ser-Pro-Tryptamine) and GG213 (Fmoc-pCF₂-Ser-Ψ^{CF₃,H}Pro-Tryptamine). The CSPs induced in ¹H-¹⁵N HSQC spectra of Pin1 by binding to GG156 are displayed for a ligand:Pin1 ratio of 6 in figure 54a. Notably, all CSPs observed are relatively weak (<0.1 ppm), offering a first indication that the affinity of this ligand is not very high. In the WW domain, the most significant CSPs were observed for residues usually binding the phosphate group

(Lys13, Gly20, Arg21, Tyr22, Tyr23) and the proline residue (Gln33, Trp34). The perturbations observed in the PPIase domain were mainly centred around the catalytic site (Cys113, Gly155 and Ile158) and the hydrophobic proline binding site (Gly123, Phe125, Gly128, Gln129 and Met130). Interestingly, residues surrounding the phosphate binding site of the PPIase domain (Lys63, Arg68 and Arg69) were virtually unaffected, suggesting that with a *trans* conformation of the pCF₂-Ser-Pro amide bond, simultaneous binding of the proline and CF₂-phosphonate groups was not possible. Apparent K_D values were derived from the measured CSPs of the most perturbed residues (see figure 54b). Upon averaging of the K_D for each domain, values of 291 ± 50 μM and 436 ± 137 μM were obtained for the WW and PPIase domains respectively.



The affinity of GG156 for Pin1 was not particularly high, however stronger binding has previously been observed in the Fmoc-capped peptides than those with an *N*-terminal

acetyl group. Hence, GG213 (Fmoc-CF₂-Ser-Ψ^{CF₃,H}Pro-Tryptamine) had the potential to be a higher affinity ligand for the PPIase domain as it contains the Ser-Ψ^{CF₃,H}Pro moiety. The titration of Pin1 with GG213 was performed, however, upon reaching a 1.5 molar excess of ligand, Pin1 began to precipitate. This was noted due to a sudden loss of peak intensity in the ¹H-¹⁵N-HSQC spectrum taken at this titration point and the presence of visible precipitate in the NMR tube (figure 55). The titration was repeated with a fresh stock of ¹⁵N-labelled Pin1 and the same results were obtained. The precipitation of Pin1 limited the extraction of an apparent K_D value for GG213, however, binding-induced CSPs in the presence of 1 molar equivalent of GG213 were identified (figure 56). CSPs were mainly observed in the PPIase domain, except for a few residues of the WW domain. Interestingly, CSPs of up to 0.1 ppm were obtained with as little as one equivalent of ligand. With most other ligands tested, such values were observed at Pin1 saturation, corresponding to ligand:protein ratios of ~6. This suggests that GG213 could have a very good affinity for the PPIase domain, being near saturation with a ligand:protein ratio close to 1. Accordingly, the precipitation of Pin1 at higher ligand concentrations could be due to aggregation of Pin1:GG213 complexes, indicative of a strong interaction. If additional time and ligand were available, further adjustments to the experimental conditions (e.g., pH, buffer, or temperature GG213:Pin1 in the 0-1 range) could potentially reveal conditions where the Pin1:GG213 complex remains soluble and could aid in the extraction of an accurate K_D value.

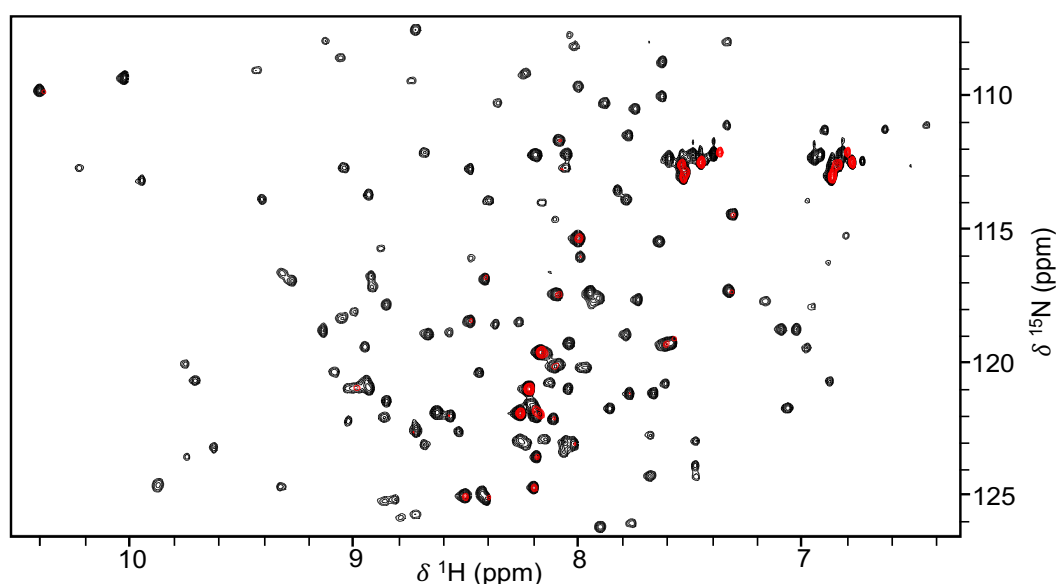


Figure 55: Overlay of ¹H-¹⁵N-HSQC spectra of apo Pin1 (black) and Pin1 + 1.5 molar equivalents of GG213 (orange). Both spectra are displayed at the same contour level, highlighting the weak S/N ratio obtained at GG213:Pin1 molecular ratio of 1.5. Recorded on 150 μM Pin1, 30 mM Tris buffer pH 7 at 298 K using a 500 MHz spectrometer

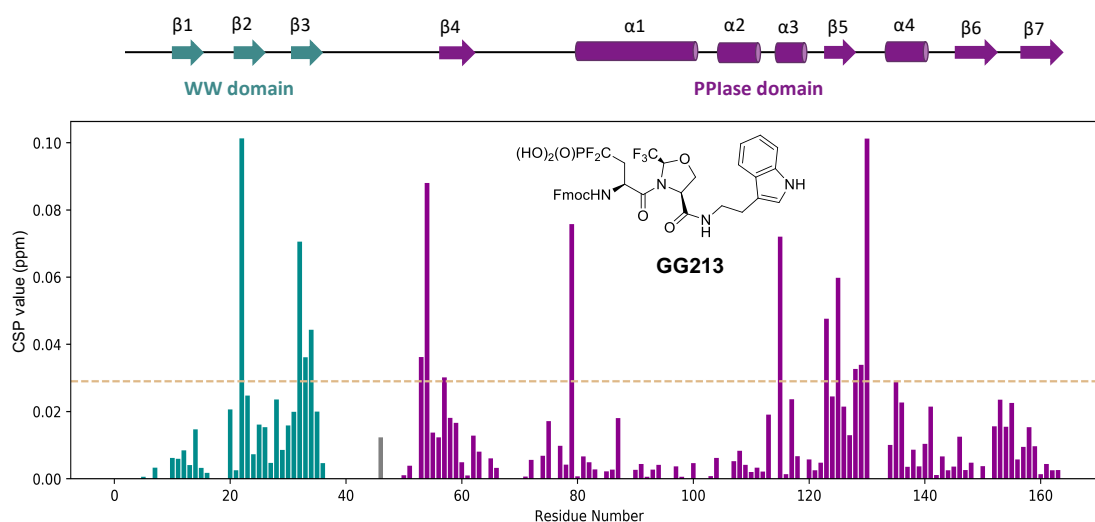
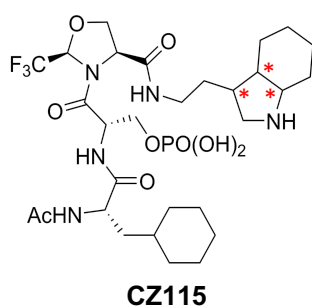


Figure 56: CSPs in FL Pin1 induced by the binding of 1 molar equivalent of GG213. The orange dashed line represents the threshold CSP value ($m + \sigma$)

Exploring additional strategies for higher potency

As phosphate group removal or replacement was thus far unsuccessful, other strategies were explored to obtain more potent compounds incorporating the $\Psi^{\text{CF}_3, \text{H}}\text{Pro}$ group. CZ115 (Ac-Cha-pSer- $\Psi^{\text{CF}_3, \text{H}}\text{Pro}$ -reducedTryptamine) exhibited the highest affinity for Pin1 among previously generated fluorinated ligands and was considered one of the lead compounds of the first generation. Originally, the CZ115 ligand was designed to incorporate a phenylalanine residue at the P1 position and a tryptamine at the P4 position. However, during synthesis, a hydrogenation reaction resulted in the complete reduction of carbon-carbon unsaturations



K_D WW = NB
 K_D PPIase = $28 \pm 6 \mu\text{M}$

Figure 57: Structure of the *cis* conformer of CZ115 highlighting its 3 chiral centres (red stars). NB means no binding.

within the compound. This reduction reaction led to the formation of a cyclohexane moiety at the intended phenylalanine position (P1) in CZ115. Moreover, the hydrogenation process affected the C-terminal tryptamine group, resulting in the creation of three chiral carbons (see figure 57). Consequently, CZ115 exists as a mixture of several diastereoisomers. The inability to isolate each diastereoisomer and to obtain specific K_D values hinders its suitability as a lead compound for further drug development. However, a preliminary ^{19}F NMR study showed that only the main *cis* conformers (80%) of CZ115 were able to bind the PPIase domain with K_D values in the same range.

This inspired the synthesis of VB66 (Ac-Cha-pSer- $\Psi^{CF3,H}$ Pro-(Aminomethyl)piperidine) as a chirally pure mimic of CZ115. The ligand is identical to CZ115 except for a piperidine group in place of the reduced tryptamine. This ensured that during synthesis, the reduction reaction would not lead to the formation of asymmetric carbons, as the cyclohexyl and piperidine rings correspond to saturated cycles. The binding of VB66 to Pin1 was analysed via the CSP method. Upon binding, no CSPs were observed in the WW domain as was expected given that the ligand contains the $\Psi^{CF3,H}$ Pro group and was predominantly populated by the *cis* isomer. Weak CSPs were observed in the PPIase domain of Pin1, even in the presence of 6 molar excess of ligand (see figure 58a). The most perturbed residues were found in three distinct regions of the PPIase domain. Firstly, CSPs were observed for residues His64, Gln66, Ser67, Ser72 of the catalytic loop involved in substrate phosphate group coordination. More significant CSPs were also identified at the proline binding pocket (Gly120, Leu122, Phe125, Gln129, and Met130). Notably, perturbations were also observed for residues of the active site (Cys113, Ala116, Lys117, Ser154 and Gly155). This strongly suggests that VB66 occupies the canonical substrate binding site and catalytic site of the PPIase domain. Apparent K_D values derived from the measured CSPs revealed a K_D value of $386 \pm 77 \mu\text{M}$ for the PPIase domain (see figure 58b). Notably, an accurate K_D value could not be calculated for residue Lys82 due to the low signal-to-noise ratio of this peak. The overall average K_D value for the PPIase domain was significantly higher than that of its reference ligand, CZ115, suggesting that modification of the *C*-terminal reduced tryptamine negatively affected affinity (see table 11).

As a further attempt to produce a more potent tetrapeptide ligand of Pin1, the incorporation of a naphthylalanine (Nal) group at the P4 position was trialled. The incorporation of Nal was inspired by the Fischer group which published two longer Pin1 peptide ligands displaying nM affinity (Wildemann *et al*, 2006). Notably, some of their peptides contained a Nal-Gln-NH₂ *C*-terminus, with the presence of the Nal group appearing to significantly enhance the potency of their ligand series. Consequently, incorporation of this group into our peptide backbone was explored. This led to the synthesis of a series of three Fischer-like ligands: GG168, GG199 and GG201. All three ligands incorporated the pSer residue at position P2 since it has been clarified that this residue is essential for binding to Pin1. The *C*-terminal of this series of ligands incorporated a Nal-Gly-NH₂ moiety. The *N*-terminal extremity was capped either by an acetyl group (GG168) or a Fmoc group (GG199 and GG201). The P3 position was either occupied by a proline residue (GG168 and GG201) or a $\Psi^{CF3,H}$ Proline group (GG199).

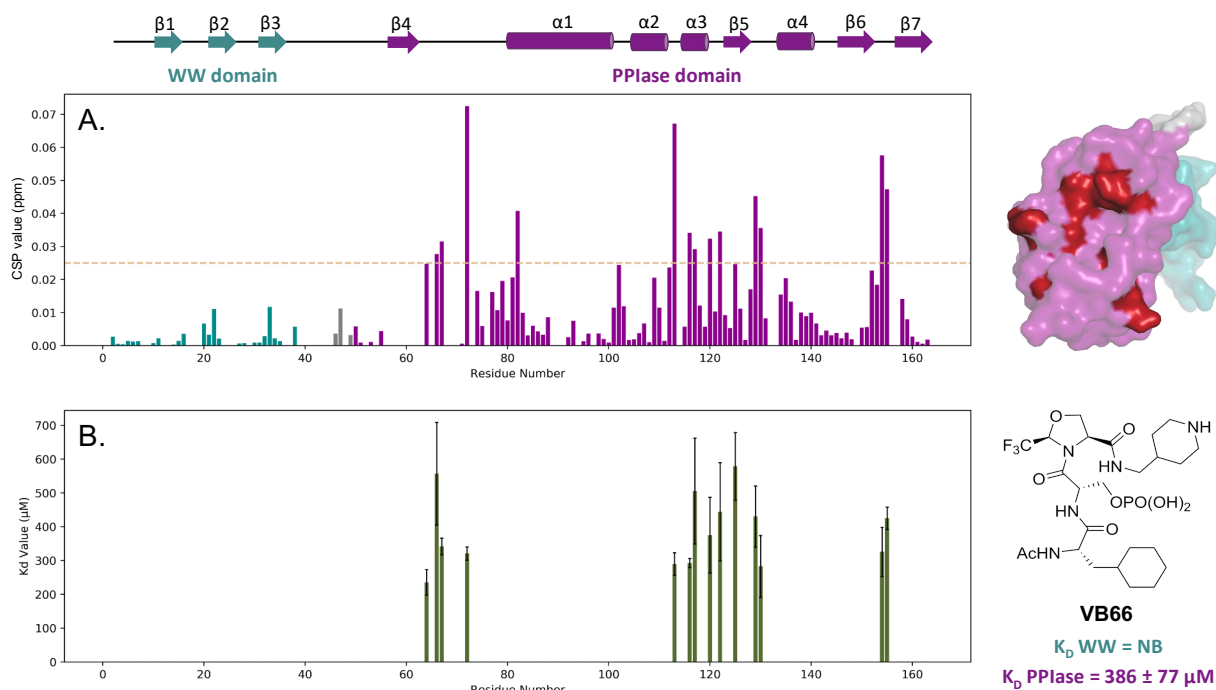


Figure 58: Analysis of the interaction of Pin1 with VB66. **A.** CSPs of FL Pin1 induced by the presence of 6 molar equivalents of VB66 with the surface representation of the most significant CSPs in red (right). Threshold CSP value ($m + \sigma$) shown as orange dashed line **B.** The K_D of residues displaying significant CSPs were used to calculate an average K_D for the PPIase domain (right).

Ligand name	Structure	Kd (μM)	
		WW	PPIase
CZ115	Ac-Cha-pSer- $\Psi^{\text{CF}_3, \text{H}}$ Pro-reduced Tryptamine	NB	20 \pm 9
VB66	Ac-Cha-pSer- $\Psi^{\text{CF}_3, \text{H}}$ Pro-(Aminomethyl)piperidine	NB	386 \pm 72

Table 11: Comparison of K_D values of CZ115 and its chirally pure mimic VB66 for Pin1. Modification of the C-terminal reduced tryptamine lowered affinity of the peptide for the PPIase domain by more than 10-fold.

Firstly, the affinity of ligand GG168 (Ac-pSer-Pro-Nal-Gly-NH₂) for Pin1 was analysed. Perturbations were observed in residues of both domains of Pin1 upon titration of up to 6 molar equivalents of GG168 (see figure 59a). Residues of the phosphate (Lys13, Gly20, Tyr22, Tyr23, Phe25) and proline (Asn30, Ser32, Gln33, Trp34 and Glu35) binding sites of the WW domain displayed the largest perturbations. The perturbations observed in the PPIase domain encompassed residues Ser115 and Gly155 of the catalytic site, residues Gly123 and Phe134 of the proline-binding pocket and residues Phe139-Leu141 of the interdomain interface. Apparent

K_D values were extracted for each residue displaying a significant CSP and their averages suggested K_D values of $221 \pm 28 \mu\text{M}$ and $489 \pm 194 \mu\text{M}$ for the WW and PPIase domains respectively (see figure 59b). The high standard deviation obtained for the PPIase domain is likely the result of different binding events. Notably, the CSP curves for Gly144 and Gly155 did not saturate even after binding to 6 molar equivalents of the ligand, indicating that these residues were involved in non-specific interactions, ultimately preventing the extraction of their respective K_D values. The K_D values of Phe139-Leu141, belonging to the $\alpha 4$ -helix located at the WW/PPIase interface were similar to those of the WW domain binding sites, strongly suggesting an allosteric effect upon WW domain binding rather than direct binding to the PPIase domain. Finally, a K_D value of $30 \pm 12 \mu\text{M}$ was calculated for residue Gly115 of the active site, which is significantly lower than for other residues of the PPIase and WW domains. As this residue is neighbouring the catalytic cysteine, it could be interesting to test the inhibitory activity of GG168 despite its apparent low affinity binding, as potentially the Nal group establishes hydrophobic interactions with Gly115 and occludes the active site.

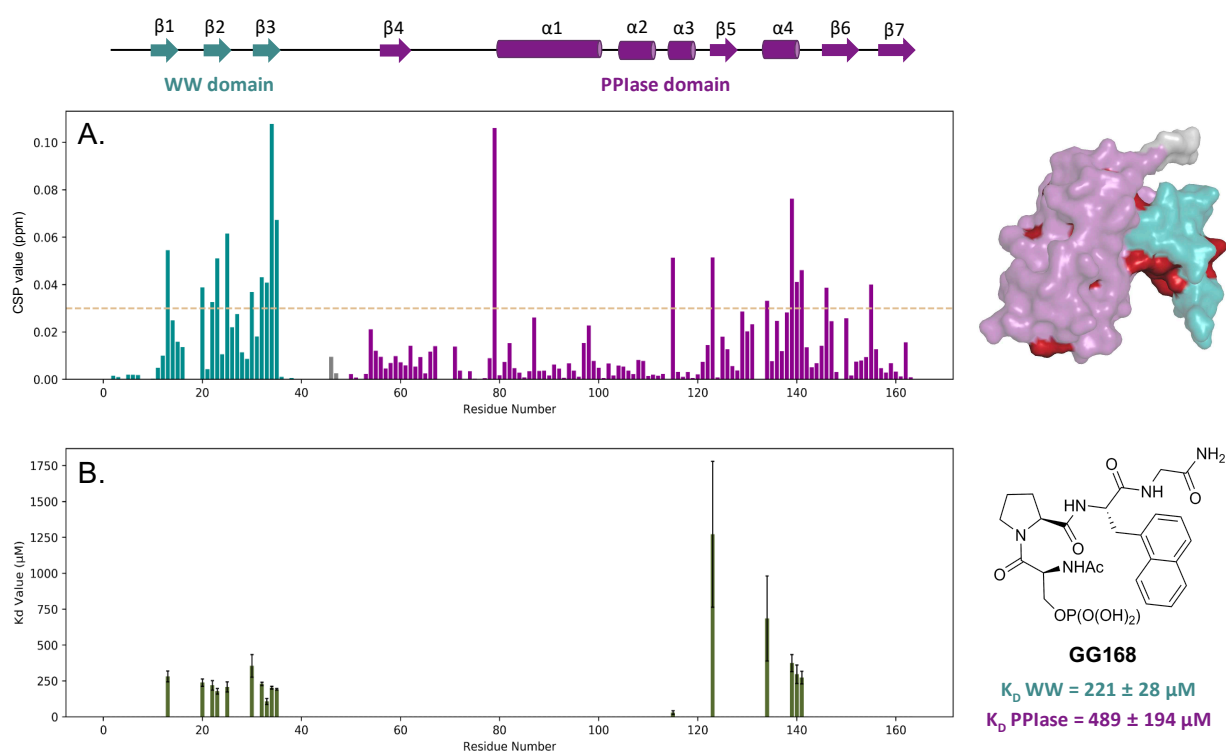


Figure 59: Analysis of the interaction of Pin1 with GG168. A. CSPs of Pin1 induced by binding to 6 molar equivalents of GG168 with the surface representation of the most significant CSPs in red (right). The orange dashed line represents $m + \sigma$. B. K_D values of residues displaying significant CSPs were used to calculate an average K_D for each domain (right). An accurate K_D value could not be calculated for residue Thr79 due to the low signal-to-noise ratio of this peak.

The second non-fluorinated compound of the series, GG201 (Fmoc-pSer-Pro-Nal-Gly-NH₂) incorporated an *N*-terminal Fmoc group as opposed to the acetyl of GGG168. To see if this group enhanced affinity for Pin1, 6 molar equivalents of GG201 were titrated into ¹⁵N Pin1. Binding-induced CSPs were observed in both domains (figure 60a). Notably, the same regions of Pin1 displayed CSPs as for GG168 binding however the CSPs were notably larger upon binding to GG201. Interestingly, this ligand did not display a significant interaction with residue Gly115 as GG168 did. Apparent K_D values for each domain of Pin1 ($26 \pm 5 \mu\text{M}$ and $141 \pm 26 \mu\text{M}$ for the WW and PPIase domains respectively) confirmed that GG201 binds to Pin1 with a higher affinity to Pin1 (figure 60b). Notably, the CSP curves for Thr79 and Gly123 did not saturate even after binding to 6 molar equivalents of the ligand, indicating that their respective CSPs were the result on non-specific interactions. Again, this ligand displays a preference for the WW domain. Hence, *N*-terminal Fmoc protection had a significant impact on the affinity of this compound for Pin1 (see table 12).

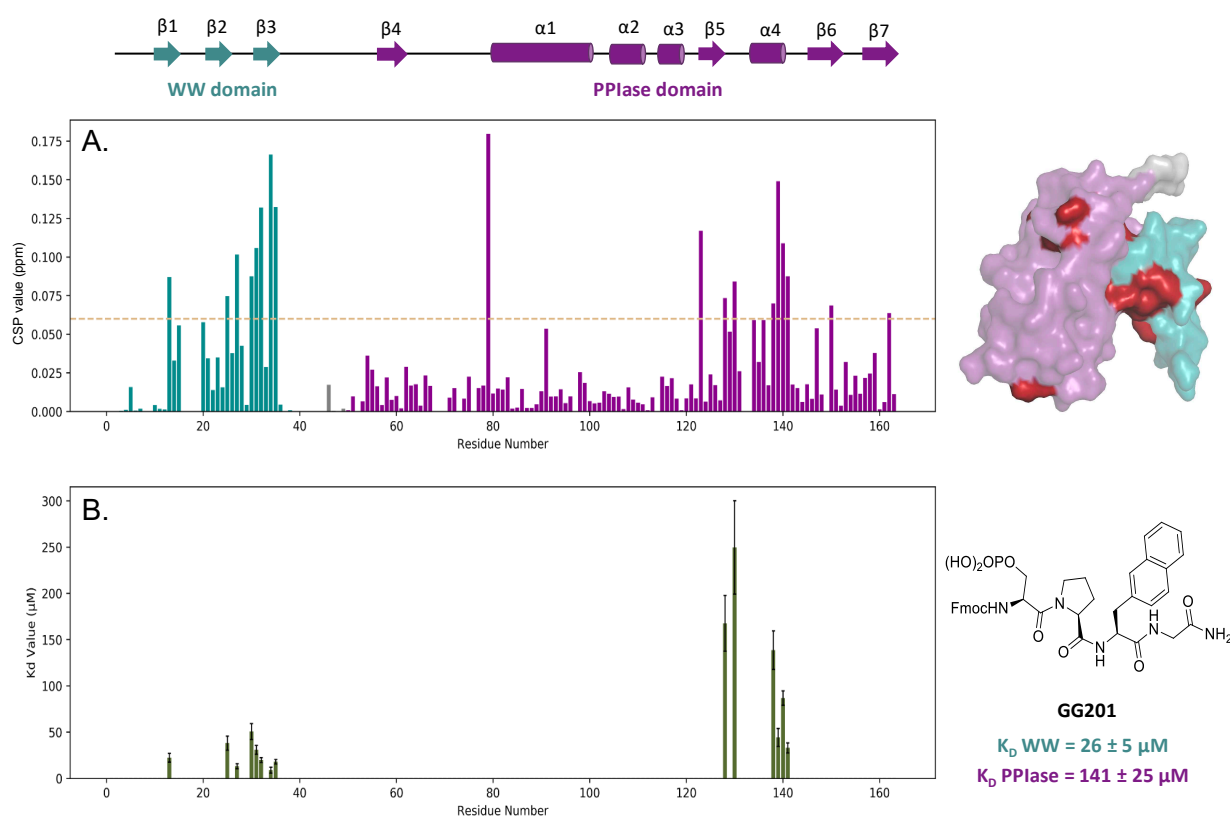


Figure 60: Analysis of the interaction of Pin1 with GG201. **A.** CSPs of FL Pin1 induced by the presence of 6 molar equivalents of GG201 with the surface representation of the most significant CSPs in red (right). The orange dashed line represents $m + \sigma$ **B.** The K_D values of residues displaying significant CSPs were used to calculate an average K_D for each domain (right).

Ligand name	Structure	K _d (μM)	
		WW	PPlase
GG168	Ac-pSer-Pro-Nal-Gly-NH ₂	221± 28	315± 56
GG201	Fmoc-pSer-Pro-Nal-Gly-NH ₂	25±5	135±23

Table 12: Comparison of K_D values of GG168 and GG201 for Pin1. Modification of the *N*-terminal group from acetyl to Fmoc increased the affinity of the peptide for both domains of Pin1, significantly increasing the binding to the WW domain by 10-fold.

Unfortunately, the other *N*-terminal Fmoc-protected compound of this series, GG199 (Fmoc-pSer-Ψ^{CF₃,H}Pro-Nal-Gly-NH₂), was not soluble in aqueous solution, preventing analysis of its binding to Pin1. The production of a fourth Fischer-like ligand incorporating the Ψ^{CF₃,H}Pro group was anticipated (Ac-pSer-Ψ^{CF₃,H}Pro-Nal-Gly-NH₂) however time restraints prevented its synthesis. This ligand is less likely to encounter solubility issues due to the replacement of the hydrophobic Fmoc group with an acetyl group, however the loss of this *N*-terminal aromatic group could be detrimental to affinity based on the results of GG168. Notably, trials of ligands containing a Nal-Gln-NH₂ *C*-terminus as opposed to the Nal-Gly-NH₂ group used in this series should be investigated, as they could prove to be more potent and soluble compounds based on the results of the Fischer group (Wildemann *et al*, 2006).

7.1.3. PROTAC series of Pin1 ligands

The peptide ligands generated from both the first- and second-generation libraries have thus far demonstrated affinities within the 10 to 100 μM range. Although these affinities indicate significant interaction with Pin1, they may not be sufficient for their application as anticancer drugs, as nano- to picomolar affinity is typically desired. Therefore, whilst continuing to refine and enhance the potency of the fluorinated ligand library, an alternative strategy for inhibiting Pin1 using PROTAC technology was also explored. PROTAC (PROteolysis-TArgeting Chimeras) technology exploits the native cellular protein degradation machinery to selectively target and degrade specific proteins of interest (POIs) through the ubiquitin proteasome system (UPS). In the context of pathological overexpression of a protein of interest (POI), the innovative PROTAC technology offers a promising approach (see figure

61). This strategy involves the adaptation of a known POI ligand by incorporating a linker module at site which does not hinder binding (Bondeson *et al*, 2015). The linker-ligand construct is then connected to a second module, which serves as a ligand for a ubiquitin ligase (E3 ligase). When the PROTAC molecule is introduced, the flexible linker enables the simultaneous binding of both ligands to their respective partners, resulting in the formation of a ternary complex consisting of the POI, PROTAC, and E2/E3 ligases. This complex formation brings the POI in close proximity to the E3 ligase, promoting the ubiquitination of the POI. Consequently, the poly-ubiquitinated POI is targeted for degradation by the proteasome. The catalytic behaviour of the PROTAC module allows for efficient ubiquitination of multiple POI molecules using a single drug molecule, offering the potential to optimise drug dosing and minimise potential side effects in patients.

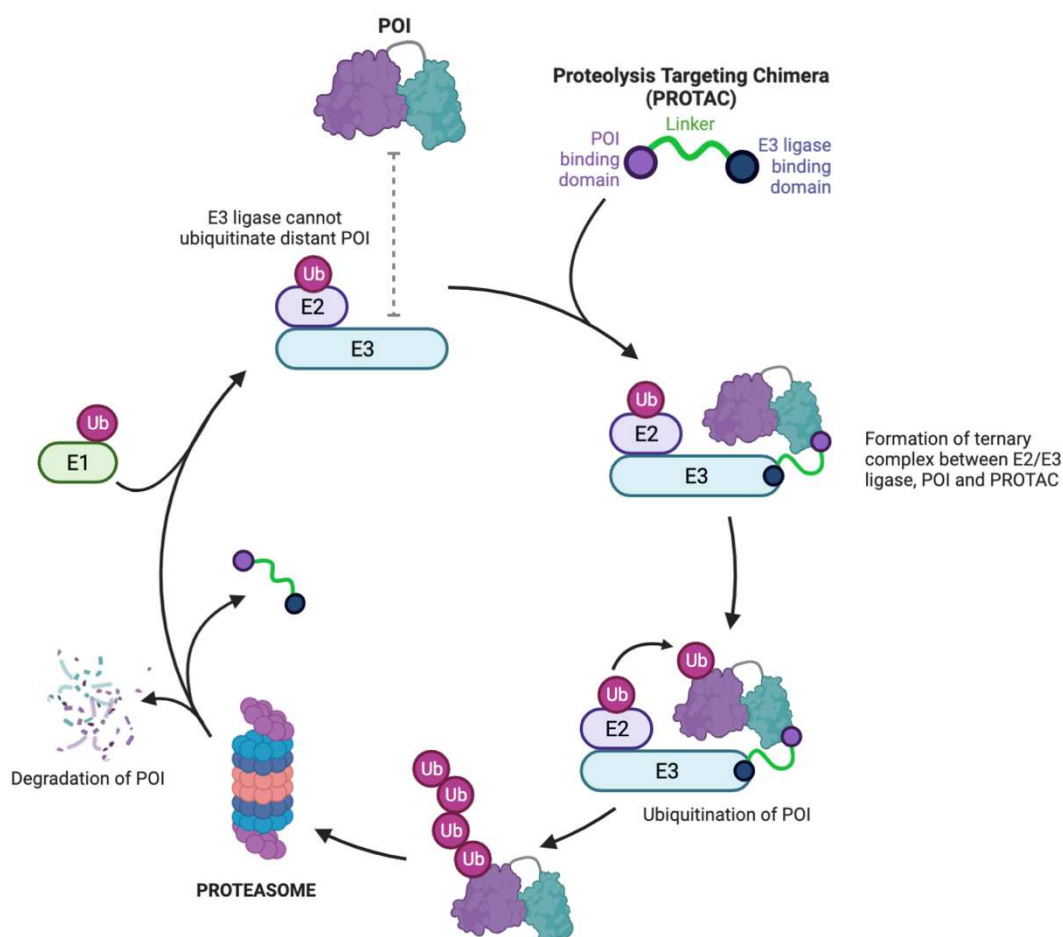


Figure 61: Schematic representation of targeting a POI for degradation using PROTAC technology. Under normal conditions the POI is not targeted for degradation as it is not in close proximity to the E3 ligase complex. Introduction of a PROTAC molecule forms a molecular bridge between the POI and E2/E3 ligase complex, resulting in the formation of a ternary complex. This promotes the poly-ubiquitination of the POI by the E3 ligase. The POI is then degraded by the proteasome and the PROTAC drug is recycled for targeting another POI molecule for degradation.

The promising potential of PROTAC technology is underscored by the fact that there are currently at least six clinical PROTAC candidates undergoing Phase I or II trials as innovative anticancer agents or therapies targeting autoimmune diseases (Mullard, 2021). This active clinical pipeline reflects the growing recognition of PROTACs as a novel therapeutic approach.

By harnessing this mechanism, a strategy for the selective degradation of Pin1 was explored with the aim of reducing Pin1 levels in cancer cells. This approach provides an alternative to complete inhibition, allowing for the restoration of Pin1 levels to a more normal state instead of completely abolishing its activity. Striking a balance between pathogenic overexpression in cancer cells and PROTAC-mediated degradation could evade potential toxic effects associated with the complete abolition of Pin1, such as neurodegeneration. Importantly, nanomolar affinity may not be required for the efficacy of this targeting technique. Moreover, PROTAC ligand design does not necessitate specific targeting of the PPIase domain, as is the case for catalytic inhibition. In the context of the discussed peptide ligands, this is advantageous as it allows for the synthesis of pSer-Pro modules instead of the CF₃-pseudoproline groups, which can complicate the synthetic process.

To rationally design the PROTAC-based Pin1 drug, three components are required: a Pin1 ligand, a ubiquitin proteasome ligand, and a linker (see figure 62). The Pin1 ligand templates, CZ40 (Fmoc-pSer-Pro-Tryptamine) and CZ55 (Ac-pSer-Pro-Tryptamine), were chosen due to being the highest affinity non-fluorinated ligands in the first-generation series (see figure 63). For the E3 ligase recruiter, a thalidomide scaffold was selected, primarily targeting cereblon (CRBN), a component of the CRL4^{CRBN} E3 ubiquitin ligase complex (Ito *et al*, 2010). Notably, CRBN is ubiquitously expressed in human cells, meaning its targeting will enhance the PROTAC-mediated formation of ternary complexes. The linker and its anchor points on each protein play a crucial role in maintaining the binding interactions with Pin1 and CRBN without disruption. Commonly used linkers in PROTAC molecules include PEG_n or aliphatic chains. The synthetic strategy of these ligands was to produce a thalidomide-linker construct via amide coupling. In parallel, the Pin1 ligand was chemically modified by introducing an alkyne hook, which serves as a connection point for the synthesis of the PROTAC molecule. This modification enables linkage of the Pin1 ligand with the thalidomide-linker construct via a triazole ring using click chemistry. First, *N*-terminal and *C*-terminal modifications of the Pin1 ligand were explored to assess their impact on ligand binding.

C-terminal modified Pin1 ligands

As a starting point, the alkyne hook was introduced into the C-terminal extremity of CZ40 and CZ55 to assess its effect on their affinity for Pin1. This led to the production of CZ382 (Fmoc-pSer-Pro-(O-propargyl)-serotonin) and CZ360 (Ac-pSer-Pro-(O-propargyl)-serotonin). The affinity of CZ382 for Pin1 was not analysed due to solubility issues with this compound limiting its application for NMR analysis. The binding of CZ360, however, was analysed via its titration into an ^{15}N -labelled sample of Pin1. Upon binding, weak CSPs were observed primarily in the WW domain of Pin1, even in the presence of 8 molar excess (see figure 64a). The most perturbed residues were Trp34 and Tyr23 and their neighbouring residues (Gly20, Tyr22, Phe25, Ser32, Gln33 and Glu35). Trp34 is known to play a key role in coordinating the proline residue of the pSer-Pro motif and Tyr23 of the $\beta 2$ strand is crucial for stabilising the phosphate group. Even weaker perturbations were observed in the PPIase domain, however only three residues displayed CSP values above the threshold ($m + \sigma$), including residues Gly123, Met130 and Phe139. Apparent K_D values were derived from the measured CSPs of the most perturbed residues (see figure 64b). Upon averaging of the K_D for each domain, values of $439 \pm 62 \mu\text{M}$ and $705 \pm 302 \mu\text{M}$ were obtained for the WW and PPIase domains respectively. Notably, the CSP profile of the PPIase domain differed significantly from what is typically observed upon specific binding to this domain. Even in the case of low-affinity ligands (e.g., VB66), clean binding to the PPIase domain habitually induces significant CSPs in residues associated with phosphate group binding (residues 63-72), the proline-binding pocket (residues 128-134), and the catalytic site (e.g., His59, Cys113, His157). By contrast, in this case, only one residue in the proline-binding site, Met130, displayed a significant CSP upon ligand binding, and notably, this residue also exhibited the largest K_D value ($>2 \text{ mM}$). Hence, it can be hypothesised that the perturbations observed in the PPIase domain could result from an allosteric effect upon ligand binding to the WW domain. This is particularly convincing considering that one of the three affected residues of the PPIase domain resides at the interdomain interface (Phe139). In addition, Phe139 displayed a K_D value of the same order of magnitude as those observed in the WW domain. It is therefore crucial to note the caution that must be taken in the calculated K_D value of $705 \mu\text{M}$ for the PPIase domain. Nevertheless, the affinity of CZ360 for the WW domain was significantly lower than that of its reference ligand, CZ55, suggesting that C-terminal modification was detrimental to affinity (see table 13).

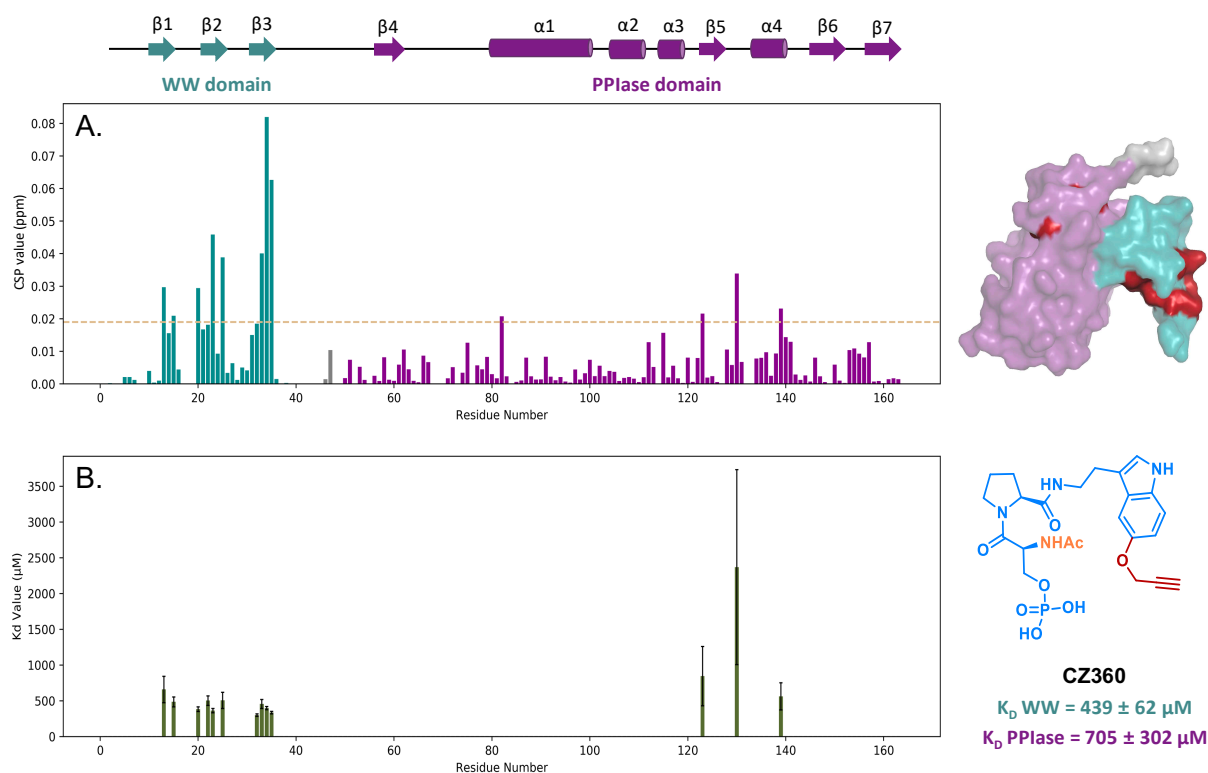


Figure 64: Analysis of the interaction of Pin1 with CZ360. **A.** CSPs of FL Pin1 induced by the presence of 8 molar equivalents of CZ360 with the surface representation of the most significant CSPs in red (right). The orange dashed line represents the threshold CSP value ($m + \sigma$) **B.** The K_D values of residues displaying the most significant CSPs were used to calculate an average K_D for each domain (right).

Ligand name	Structure	K _d (µM)	
		WW	PPIase
CZ55	Ac-pSer-Pro-Tryptamine	32±6	NB
CZ360	Ac-pSer-Pro-(O-propargyl)-serotonin	439±62	705±302

Table 13: Comparison of K_D values of CZ55 and CZ360 for Pin1. C-terminal modification of the tryptamine group by the addition of an alkyne hook was detrimental for binding to the WW domain.

N-terminal modified Pin1 ligands

As C-terminal modification significantly reduced affinity for Pin1, the introduction of the alkyne group at the *N*-terminal extremity was explored. Synthesis of compounds LME27

(FmocPra-pSer-Pro-Tryptamine) and LME51 (AcPra-pSer-Pro-Tryptamine) were again inspired by reference ligands CZ40 and CZ55. The binding of LME27 to Pin1 was analysed, revealing significant binding-induced CSPs in both domains (see figure 65a). These included residues of the WW domain involved in the coordination of the proline group of the pSer-Pro motif (Gln33, Trp34 and Glu35). In addition, residues of the WW interdomain interface were affected (His27-Ser32). Residues of the PPIase domain also showed significant perturbations, including Phe134 of the proline-binding pocket and residues Ser138-Leu141 of the α 4-helix located at the interdomain interface. Apparent K_D values were derived from the measured CSPs of the most perturbed residues (see figure 65b). Upon averaging of the K_D for each domain, values of $44 \pm 6 \mu\text{M}$ and $88 \pm 16 \mu\text{M}$ were obtained for the WW and PPIase domains respectively. However, again, the lack of significant CSPs in the phosphate-binding site of the PPIase domain (residues 63-72) indicate that the phosphate group of LME27 is not binding to the PPIase domain. Moreover, the region of the PPIase domain displaying the most significant CSPs corresponds to the α 4-helix of the interdomain interface. In addition, the K_D values calculated for residues of the PPIase domain are similar to those calculated for the WW domain. Therefore, it can be hypothesised that, similarly as for ligand CZ360, the perturbations observed in the PPIase domain are the result an allosteric effect induced by LME27 binding to the WW domain. Significantly, compared to the CZ40 reference ligand, the *N*-terminal modification in LME27 did not significantly reduce binding affinity for the WW domain (see table 14).

The second *N*-terminal modified pre-PROTAC compound, LME51, was synthesised with an acetyl group instead of Fmoc protection, as in reference ligand CZ55. Weak CSPs were observed for both domains of Pin1 upon titration with 8 molar equivalents of LME51 (figure 66a). Similar to LME27, the WW domain residues displaying the highest chemical shift perturbation included residues of the proline-binding site and interdomain interface. CSPs observed in the PPIase domain were much weaker than those of the WW domain and include only Thr79, Met130 and Phe139. From the most significant CSPs of each domain, apparent K_D values of $285 \pm 32 \mu\text{M}$ and $598 \pm 193 \mu\text{M}$ were obtained for the WW and PPIase domains respectively (figure 66b). Again however, the binding of the ligand to the WW domain is expected to be responsible for the perturbations to residues of the PPIase domain via an allosteric effect. Notably, the K_D value for the WW domain is much higher than for LME27 and the CZ55 reference ligand (table 15). Hence, modification on the *N*-terminal end is seemingly well-tolerated as long as an aromatic moiety (Fmoc) caps the amino group.

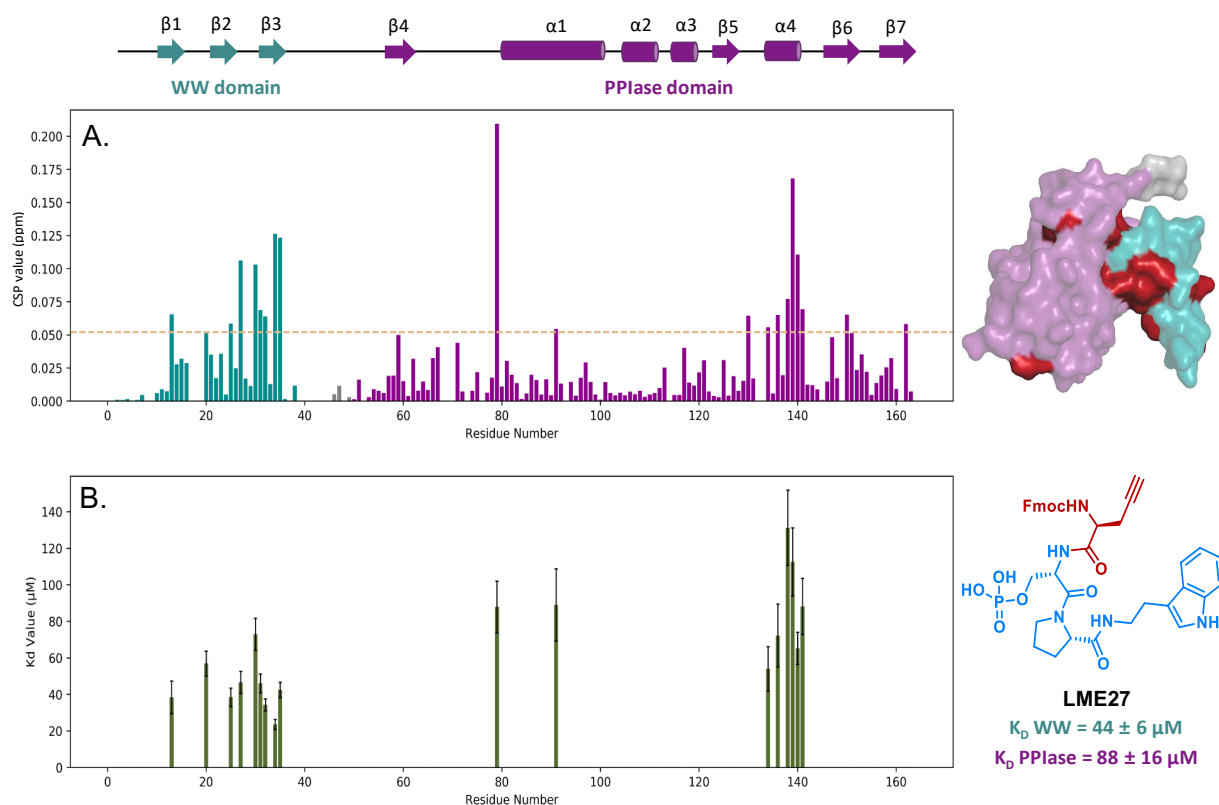


Figure 65: Analysis of the interaction of Pin1 with LME27. **A.** CSPs of FL Pin1 induced by the presence of 4 molar equivalents of LME27 with the surface representation of the most significant CSPs in red (right). The orange dashed line represents the threshold CSP value ($m + \sigma$) **B.** The K_D values of residues displaying the most significant CSPs were used to calculate an average K_D for each domain (right).

Ligand name	Structure	K_D (μM)	
		WW	PPIase
CZ40	Fmoc-pSer-Pro-Tryptamine	26 ± 6	438 ± 80
LME27	FmocPra-pSer-Pro-Tryptamine	44 ± 6	88 ± 16

Table 14: Comparison of K_D values of CZ40 and LME27 for Pin1. *N*-terminal modification of the reference ligand CZ40 with an alkyne hook was overall well-tolerated. Affinity of LME27 for the WW domain of Pin1 was within the same order of magnitude as for CZ40. The K_D value calculated for the PPIase domain should be taken with caution, as the CSPs observed are suspected to be the result of an allosteric effect upon ligand binding to the WW domain.

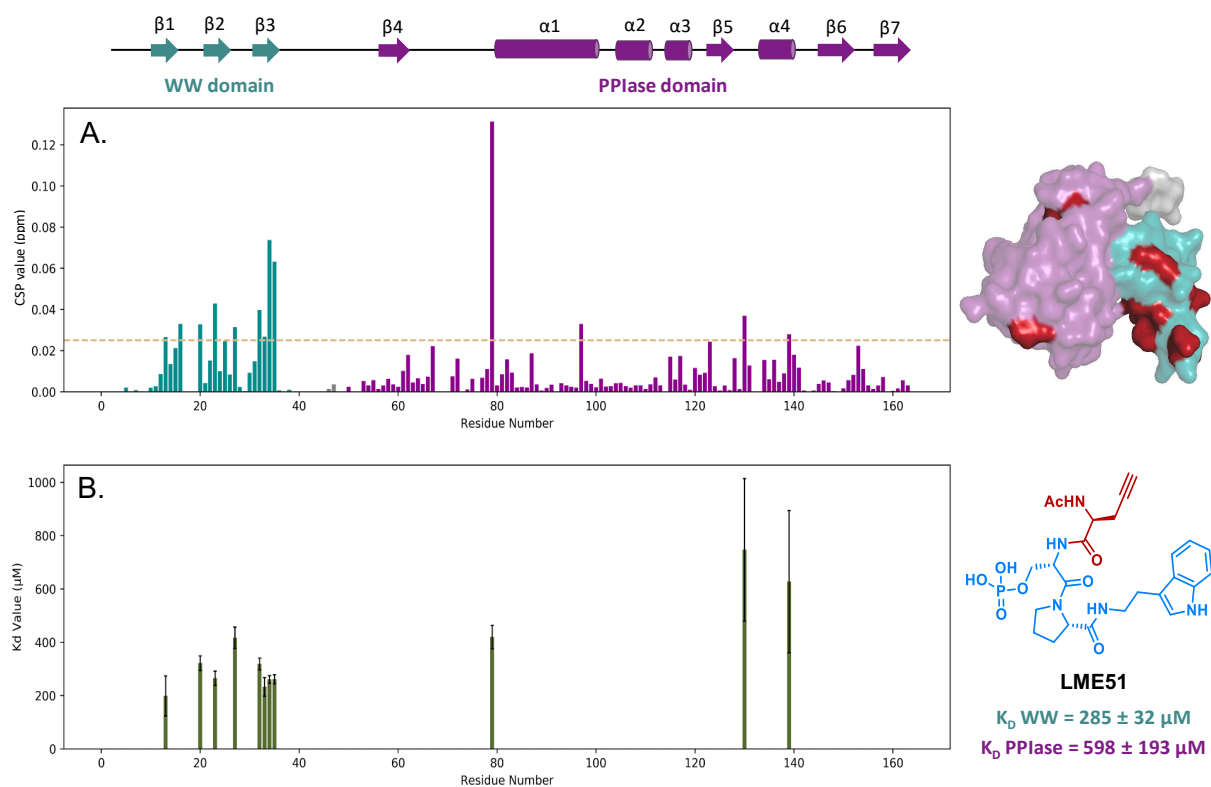


Figure 66: Analysis of the interaction of Pin1 with LME51. **A.** CSPs of FL Pin1 induced by the presence of 4 molar equivalents of LME51 with the surface representation of the most significant CSPs in red (right). The orange dashed line represents the threshold CSP value ($m + \sigma$) **B.** The K_D values of residues displaying the most significant CSPs were used to calculate an average K_D for each domain (right).

Ligand name	Structure	K_D (μM)	
		WW	PPlase
CZ55	Ac-pSer-Pro-Tryptamine	32±6	NB
LME51	AcPra-pSer-Pro-Tryptamine	285±32	598±193
LME27	FmocPra-pSer-Pro-Tryptamine	44±6	88±16

Table 15: Comparison of K_D values of CZ55, LME51 and LME27 for Pin1. *N*-terminal modification of the reference ligand CZ55 with an alkyne hook in LME51 reduced affinity for the WW domain by almost 10-fold. The *N*-terminal modification is better tolerated in the Fmoc-capped LME27 than the Ac-capped LME51.

As LME27 represented the most potent pre-PROTAC compound, it was selected as the starting scaffold for further PROTAC modifications. The affinity of this ligand for Pin1 seemed highly reliant on its *N*-terminal Fmoc protection, however the carbamate bond linking the Fmoc group to the *N*-terminus of the peptide is relatively weak and hence there were concerns that this molecule would lack stability in a cellular environment. Therefore, an analogue of LME27 was synthesised (AGI24) whereby this carbamate bond was replaced by an amide bond (see figure 67). This amide bond is much stronger than the carbamate bond and hence is likely to be more resistant in a cellular environment.

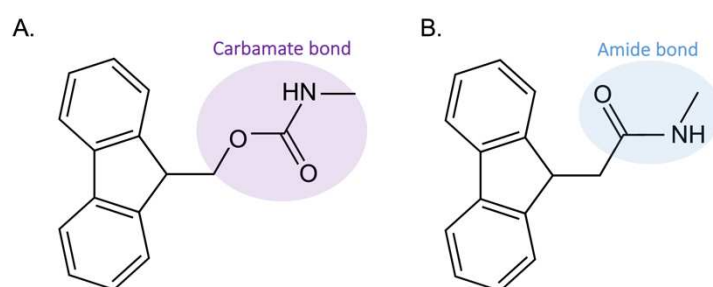


Figure 67: Comparison of the linkage of the Fmoc via a carbamate (A) vs amide (B) bond. LME27 and AGI24 display the same overall structure except that the Fmoc group is linked to the peptide via a carbamate bond (purple) in LME27 whereas it is linked via an amide bond (blue) in AGI24.

Analysis of the binding of AGI24 to ^{15}N -labelled Pin1 revealed significant CSPs in both domains of Pin1, as is observed for its parent compound LME27 (see figure 68a). Within the WW domain, notable CSPs were observed for residues Tyr23, Gln33, Trp34 and Glu35 of the proline-binding pocket. Within the PPIase domain, two residues of the proline binding pocket also displayed significant CSPs (Gly123, Met130) as well residues of the α 4-helix (Ser138, Phe139, Ala140, Leu141). The calculation of apparent K_D values for each domain revealed values of $66 \pm 12 \mu\text{M}$ and $80 \pm 32 \mu\text{M}$ for the WW and PPIase domains (see figure 68b). As with the other PROTAC ligands, the affected residues in the PPIase domain are not indicative of specific binding to the PPIase domain. Accurate K_D values for residues Gly123 and Met130 could not be determined as their CSP curves did not reach saturation even after introduction of 6 molar equivalents of AGI24. This suggests that the K_D values for these residues are likely to be over 1 mM, indicating a non-specific binding event. In contrast, the K_D values observed for

residues Ser138-Leu141 are within the same order of magnitude as those calculated for residues in the WW domain. These residues belong to the interdomain interface, suggesting that their CSPs result from an allosteric effect caused by ligand binding to the WW domain. Overall, the affinity of AGI24 for the WW domain was within the same order of magnitude as for observed LME27 (see table 16). This indicates that the amide modification of the Fmoc linkage was well-tolerated and could be a useful modification for increasing the stability and hence half-life of our final PROTAC compounds.

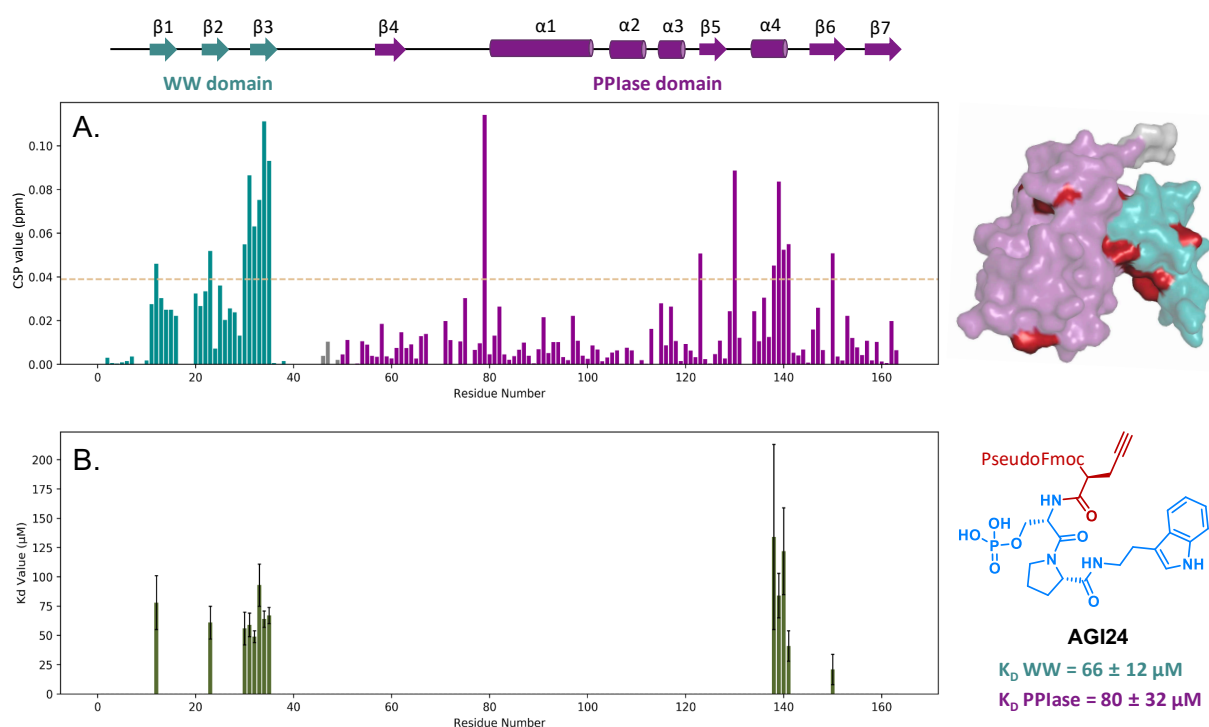


Figure 68: Analysis of the interaction of Pin1 with AGI24. **A.** CSPs of FL Pin1 induced by the presence of 6 molar equivalents of AGI24 with the surface representation of the most significant CSPs in red (right). The orange dashed line represents the threshold CSP value ($m + \sigma$) **B.** The K_D values of residues displaying the most significant CSPs were used to calculate an average K_D for each domain (right).

Ligand name	Structure	K _D (µM)	
		WW	PPlase
LME27	FmocPra-pSer-Pro-Tryptamine	44±6	88±16
AGI24	PseudoFmocPra-pSer-Pro-Tryptamine	66±12	80±32

Table 16: Comparison of K_D values of LME51 and AGI24 for Pin1. The substitution of the carbamate bond linking the *N*-terminal Fmoc group (LME27) for an amide bond (AGI24) did not significantly impact the K_D.

Complete PROTAC ligands

After the successful identification of a Pin1 ligand bearing the alkyne hook and displaying adequate potency, the next step of the synthesis of PROTAC ligands was to join the Pin1 ligand to the thalidomide-linker construct. The first pool of complete Pin1-PROTACs has been prepared by our collaborators at Cergy-Pontoise. In practice, for application in a cellular context, protection of the phosphate group will be necessary to enhance cell permeability. Hence, a complete PROTAC ligand with the protected phosphate group, LME34, was synthesised for cellular studies (see figure 69a). However, in order to confirm that the complete PROTAC ligand still retained binding to Pin1, an additional ligand was synthesised, LME66, which incorporates the unprotected phosphate group (figure 69b). This compound presents the active PROTAC compound in cells (LME27-linker-thalidomide) after esterase-mediated hydrolysis cleaves the thioesters.

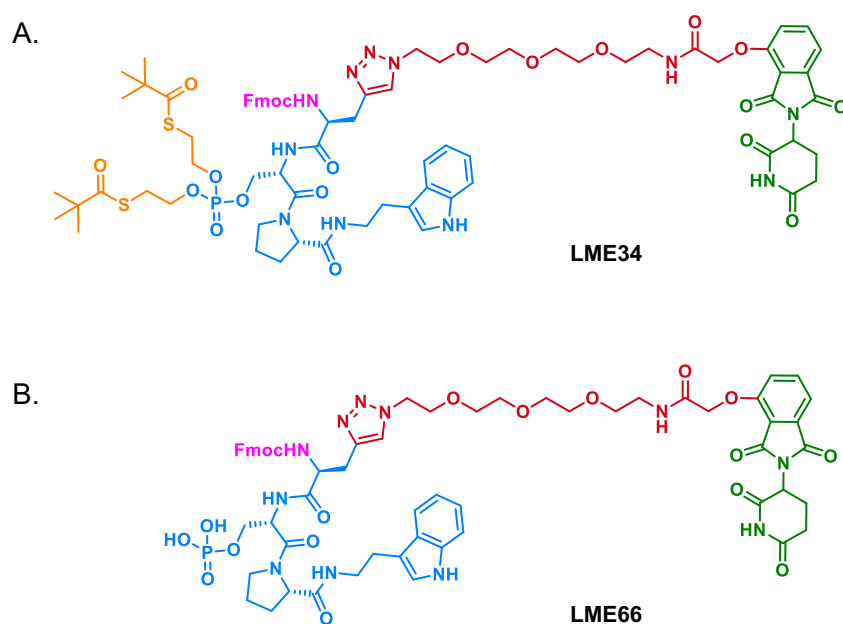


Figure 69: Structure of complete PROTAC ligands comprised of the Pin1 LME27 ligand (blue), a (PEG)₃ linker (red) and thalidomide scaffold (green). Two ligands were produced, with (A. LME34) and without (B. LME66) phosphate group protection.

In order to test if the complete PROTAC molecule retained the same level of affinity for the WW domain of Pin1 as LME27, an NMR titration was performed. Unfortunately, the complete PROTAC compound encountered solubility issues, rendering it only partially soluble in aqueous solution as a 1 mM stock solution. Due to the lower concentration of ligand stock

solution, a final molar equivalent of only 1.5 could be reached due to severe dilution of the Pin1 sample after this point. The CSPs induced by binding to 1.5 equivalents of LME66 were comparable to those experienced upon binding to LME27 (figure 70). This includes residues of the WW domain substrate binding site (Tyr23, Trp34 and Glu35), residues of the WW interdomain interface (His27, Asn30 and Ala31) and residues of the PPIase domain interface (Ser138-Leu141). The affected residues of the PPIase domain are again indicative of an allosteric modulation from the binding event at the WW domain, not a specific interaction with the PPIase domain. Unfortunately, the CSP curves did not quite reach saturation at a ligand:protein ratio of 1.5, so the calculated K_D values for each residue displayed a large margin of error. However, a consensus average of less than 100 μM could be obtained. Overall, these preliminary K_D determinations indicate that the compound LME66 binds to the WW domain with an affinity similar to that of LME27. However, a more accurate K_D determination is required to confirm this.

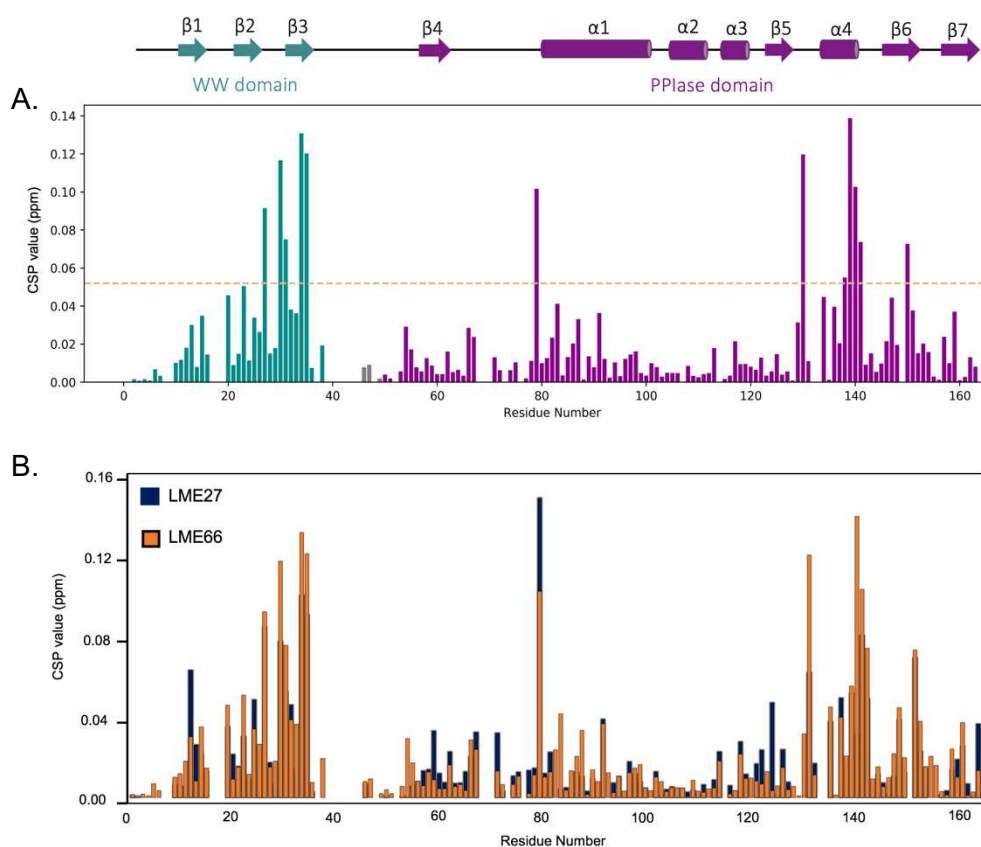


Figure 70: Analysis of the interaction of Pin1 with LME66. A. CSPs of FL Pin1 induced by the presence of 1.5 molar equivalents of LME66 with the surface representation of the most significant CSPs in red (right). The orange dashed line represents the threshold CSP value ($m + \sigma$) B. Comparison of CSP values obtained upon binding of Pin1 at a 1:1.5 molar ratio of LME27 (dark blue) or LME66 (orange). Overall, the same regions were affected by the binding of both ligands.

7.2 Optimising a Pin1 activity assay

After characterising the binding of the first- and second-generation peptide libraries to Pin1, key groups responsible for affinity have been identified, resulting in the generation of several ligands displaying K_D values below 100 μM . However, it is crucial to note that affinity alone does not fully describe the potency of these compounds, as their inhibitory activity against Pin1 remains to be determined. Previous attempts to assess the inhibitory activity of the first-generation fluorinated ligands employed a commercial fluorescent-based assay, which unfortunately yielded inconsistent results. This indirect assay is widely used in the literature and utilises a coupled approach that takes advantage of chymotrypsin's conformational specificity for *trans* Xaa-Pro-Phe motifs (Kofron *et al*, 1991). The substrate used in this assay incorporates a C-terminal p-nitroanilide group (Xaa-Pro-Phe-pNA). Upon cleavage of the Phe-pNA bond by chymotrypsin, the p-NA group is released, resulting in a detectable fluorescence signal. Since Xaa-Pro bonds naturally exist in peptides in a predominantly *trans* conformation (with a ratio of approximately 90:10 *trans:cis*), it is necessary to first stabilise the *cis* conformation so that the immediate cleavage of the *trans* population by chymotrypsin does not mask the initial rate of prolyl isomerisation (Kofron *et al*, 1991). Preparing the substrate in a solution of LiCl/TFE has the ability to shift the *cis* population from 10% to up to ~70% dependent on the substrate sequence. Notably, when this assay was applied to monitor the inhibitory activities of CZ55 and CZ52 of the first-generation of fluorinated ligands, strange results were obtained. At lower concentrations of CZ55 (< 50 μM) and with a protein concentration of 270 nM, there was an apparent enhancement of Pin1 isomerase activity instead of inhibition. Only at higher concentrations (> 50 μM) did the ligands exhibit Pin1 inhibition. CZ52 displayed similar effects on Pin1 activity, despite being designed to target the WW domain of Pin1. Due to this peculiar behaviour, it was not possible to extract IC_{50} values for the ligands in question. These unexpected results raised concerns about the overall reliability of the chymotrypsin-coupled enzymatic assay for the specific application of Pin1 kinetic parameter determination. One potential explanation for these unexpected results was the susceptibility of Pin1 or its peptide ligands to cleavage during the assay.

To investigate this hypothesis, 1D proton NMR spectra of the ligands were recorded both before and after incubation with the reaction mixture. Surprisingly, these spectra remained unaltered, suggesting that the ligands were indeed resistant to cleavage by chymotrypsin. However, upon

examining the ^1H - ^{15}N HSQC spectrum of Pin1 in the presence of 0.12 equivalents of chymotrypsin, the minimum proportion used in the assay, significant spectral differences were observed (see figure 71). Notably, the global reduction in spectral dispersion in the proton dimension strongly indicated the denaturation of Pin1. This observation suggests that the recombinant Pin1 was cleaved by chymotrypsin, which could explain the surprising outcomes obtained with the commercial fluorometric assay. In addition, although the substrate/enzyme ratio recommended by the test supplier was as low as 2, continuous activity was observed for over 8 hours and Pin1 activity did not vary linearly with enzyme concentration. For all these reasons, the use this assay has been discontinued in the lab.

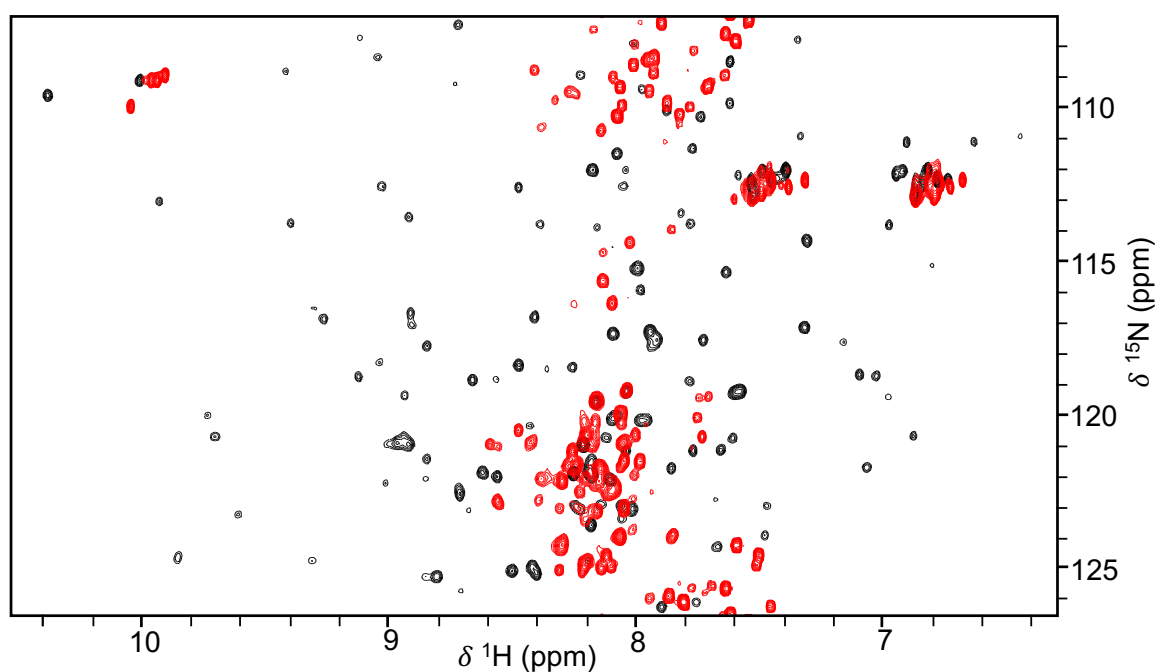


Figure 71: Overlay of ^1H - ^{15}N HSQC spectrum of Pin1 (black) and Pin1 in the presence of 18 μM chymotrypsin (red). Recorded on 150 μM sample of ^{15}N -Pin1 in 30 mM Tris buffer pH 7, 298 K using a 500 MHz spectrometer.

7.2.1. Protease-free UV assay

Due to the unreliability of the fluorometric chymotrypsin-coupled assay, a different approach was pursued here for monitoring Pin1-catalysed prolyl-isomerisation using a protease-free assay (Janowski *et al*, 1997). This assay relies on the minor difference in UV absorption coefficients of the *cis* and *trans* conformers of nitroanilide-containing peptides. The

protease-free assay has been widely used in the literature to assess the activity of various PPIases, including Pin1 (Janowski *et al*, 1997; Zhang *et al*, 2009). Similar to the chymotrypsin-based assay, the substrate is prepared in a solution LiCl/TFE which is supposed to promote the adoption of the *cis* conformation, thereby perturbing the *cis:trans* ratio from its equilibrium. After a solvent jump into aqueous solution, absorption is monitored to observe *cis* to *trans* isomerisation as the populations return to their equilibrium distribution. In the presence of Pin1, the return to equilibrium should be faster. This protease-free assay offers a more direct and reliable approach for studying Pin1 kinetics without the confounding effects of enzymatic cleavage. Three substrate sequences were trialled based on those used in the literature (figure 72) (Janowski *et al*, 1997).

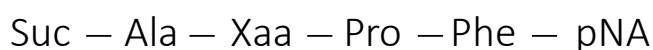


Figure 72: Sequence of substrates used in the protease-free Pin1 activity assay.

Substrates share the same common sequence whereby Suc refers to succinate and pNA refers to p-nitroanilide. The variant group (Xaa) is replaced by either a Glu, Ala or pSer residue (referred to as the Glu-substrate, Ala-substrate and pSer-substrate respectively).

Glu-Substrate

The Glu-substrate, despite lacking a pSer/pThr group, has been previously employed as a substrate for Pin1 in this protease-free assay and hence was used for this study (Zhang *et al*, 2009). First, to optimise the experimental conditions, it was necessary to identify the optimal wavelength for measuring the *cis* to *trans* isomerisation, so to say the wavelength at which there is the largest difference between *cis* and *trans* absorption properties of the substrates (figure 73). Notably, the total absorbance of the ‘*cis*’ (in 0.47 M LiCl/TFE) and ‘*trans*’ (in HEPES buffer pH 7.8) Glu-substrate was initially monitored for wavelengths 230-390 nm at ambient temperature using a UV spectrometer available in the lab (figure 73a). However, this spectrometer lacked a cooling module. Consequently, kinetic experiments were ultimately conducted at 10°C using a CD spectrometer equipped with a Peltier temperature controller. No CD signal was observed for these two samples but interestingly, the absorbance measurements obtained simultaneously during CD data acquisition showed significant differences. The resulting difference spectrum between the ‘*cis*’ and ‘*trans*’ forms of the Glu-substrate exhibited a maximum ΔA at 300 nm, well below the value obtained with the UV spectrometer (see figure

73b). Additionally, it was observed that the overall UV absorbance of the Glu-substrate in LiCl/TFE was higher than in HEPES buffer ($\Delta A > 0$), contrary to prior research findings. Given these inconsistencies, kinetic experiments were attempted at two distinct wavelengths: 300 nm and 330 nm (with 330 nm being documented in the literature as the wavelength corresponding to the maximum absorbance difference between *cis* and *trans*-Glu substrates). Following a solvent jump to HEPES buffer (with a dead time of ~ 25 s), the time course of *cis*-to-*trans* isomerisation of the Glu-substrate was closely monitored at both wavelengths. Surprisingly, no change in absorption was observed over a 20-minute period for either wavelength. This led to the hypothesis that the *cis* form may not have been stabilised in the solution of LiCl/TFE and the different UV absorption properties of the Glu-substrate in LiCl/TFE and HEPES buffer was simply due to the solvent.

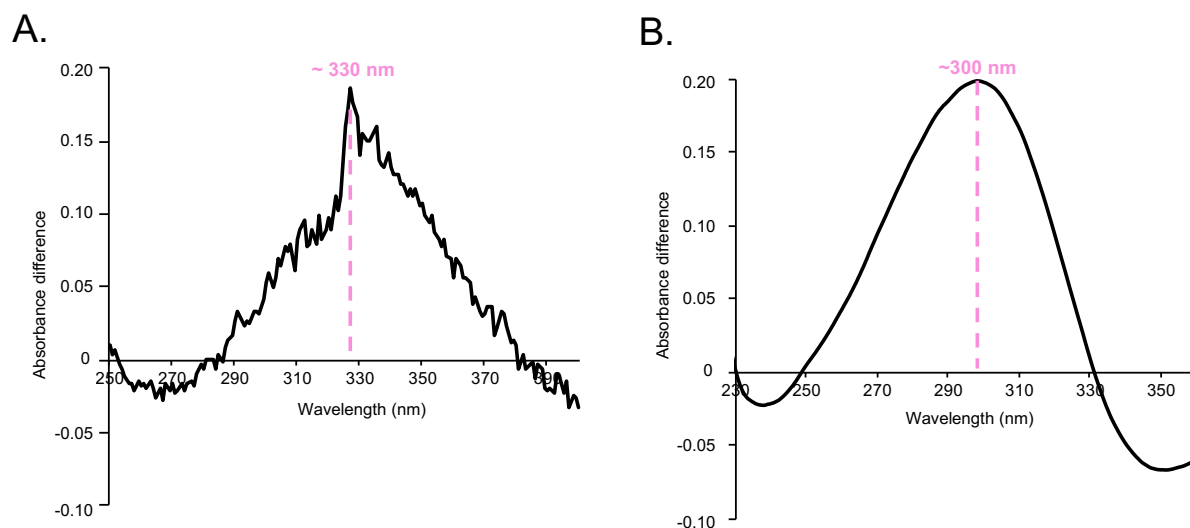


Figure 73: Difference spectrum of the Glu-Substrate (75 μ M) calculated from spectra recorded in 0.47 M LiCl/TFE and HEPES pH 7.8 recorded on a UV spectrometer at ambient temperature (A) and using a CD spectrometer at 10 °C (B). The maximal absorption difference between the *cis* and *trans* isomers was variable depending on the used instruments (estimated at a wavelength of 330 nm on the UV spectrometer and 300 nm on the CD spectrometer).

To investigate further, NMR experiments were conducted to determine the *cis:trans* ratio of the Glu-substrate in a solution of 0.47 M LiCl/TFE. To characterise the populations of *cis* amide and *trans* amide around the Glu-Pro bond, it was necessary to first assign the protons of the different conformers using NMR. NMR (Rotating frame Overhauser Effect Spectroscopy (ROESY), Total Correlation Spectroscopy (TOCSY) and 2D ^1H - ^{13}C HSQC experiments were

performed on 1 mM of Glu-substrate. After assignment of peaks corresponding to the *cis* and *trans* conformers of the Glu-substrate and integration of an isolated resonance from each conformer, the *cis:trans* amide ratio was determined to be 30:70. This indicated that the *cis* conformer comprised only 30% of the Glu-substrate population in the LiCl/TFE solution. Consequently, the dynamic range of population variations should be very low under these conditions (change of *trans* conformer population from 70% to 90% at the thermodynamic equilibrium in aqueous solvents). This may explain why UV spectra were inappropriate for describing this kinetics.

Ala-Substrate

As the *cis* conformer of the Glu-substrate could not be sufficiently stabilised in the presence of LiCl/TFE, a second substrate sequence, which had successfully been applied to this PPIase assay in the literature, was trialled (Janowski *et al*, 1997). First, NMR characterisation of the *cis* and *trans* isomers of the Ala-substrate was performed in a solution of 0.47 M LiCl/TFE. It was confirmed that the population indeed favoured the *cis* conformer, with a 60:40 ratio of *cis:trans* Ala-Pro amide populations. Thus, it was concluded that the 0.47 M LiCl/TFE solution adequately stabilised the *cis* population of the Ala-substrate but not of the Glu-substrate. Since it was possible to significantly stabilise the *cis* conformer of this substrate, the Ala-substrate was used in the protease-free assay. First, the optimal wavelength for measuring *cis* to *trans* isomerisation was calculated. Due to the solvent-induced discrepancies observed for the UV absorbance by the Glu-substrate, the absorbance difference between the *cis* and *trans* substrates was obtained in HEPES buffer, pH 7, 30s and 600s after a solvent jump, revealing the largest difference in absorption at 320 nm (figure 74). The evolution of this absorption difference was monitored on the Ala-substrate over the course of 10 minutes after the solvent jump (deadtime ~25 s). It was observed that after 350 s the reaction had already finished (figure 75). However, upon incubation of the reaction mixture with different concentrations of Pin1, the observed *cis-trans* isomerisation, as indicated by the absorption curves, remained unaltered. It was hence hypothesised that the Ala-substrate might not serve as a substrate for Pin1. This notion is rationalised by considering that the Pin1 consensus sequence requires the presence of a phosphorylated threonine or serine residue prior to the prolyl peptide bond. Moreover, the ligand binding investigations elucidated in section 8.1 underscore the essential role of this phosphorylated residue for effective binding to Pin1. To investigate if the Ala-substrate could indeed bind to Pin1, an NMR titration was performed. However, no significant backbone amide

CSPs were observed upon addition of 6 molar equivalents of the Ala-substrate to a ^{15}N -labeled sample of Pin1, confirming that there was no interaction between them.

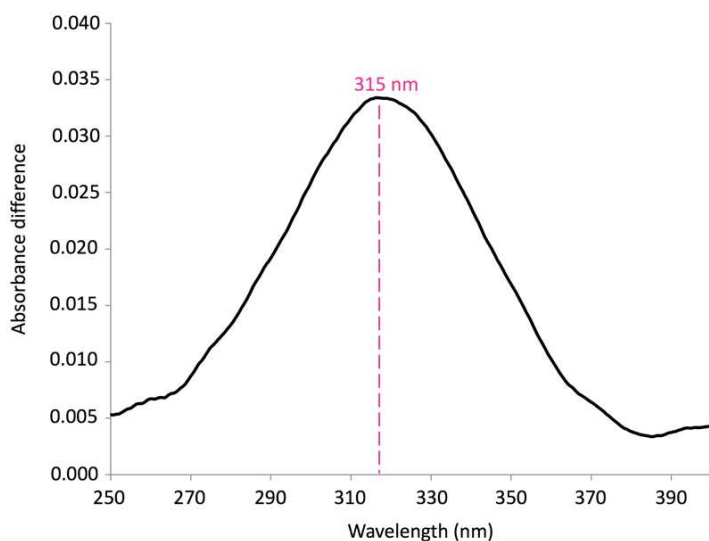
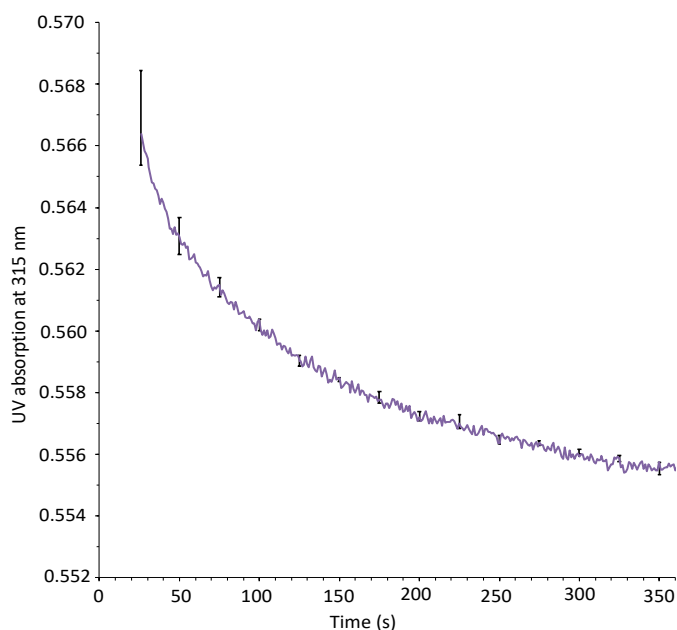


Figure 74: Difference spectrum of the Ala-Substrate (48 μM) obtained in HEPES pH 7.8 calculated from spectra at 30s and 600s after a solvent jump from LiCl/TFE to HEPES. The maximal absorption difference between the *cis* and *trans* isomers was observed at a wavelength of 315 nm.

Figure 75: Time course of reversible prolyl isomerisation of the Ala-Substrate (48 μM) was studied after a rapid transition from a solution containing 0.47 M LiCl/TFE to 35 mM Hepes at pH 7.8. The isomerisation process is monitored by a decrease in absorption of the *trans* isomer in comparison with the *cis*. Error bars were calculated based on 3 replicates.



pSer-Substrate

The lack of success of the protease-free assay was attributed to the lack of a suitable substrate, with the Ala-substrate not serving as a substrate of Pin1 and the *cis* population of the Glu-substrate unable to be stabilised in a solution of LiCl/TFE. Hence, a third pNa-containing peptide was designed which was identical to the Ala- and Glu-substrate except it contained the pSer-Pro motif necessary for Pin1 recognition (figure 72). The first test performed was to

stabilise the *cis* form using a solution of 0.47 M LiCl/TFE. Unfortunately, the peptide was insoluble in a solution of LiCl/TFE. Further research revealed that solutions of LiCl/ACN and LiCl/THF could effectively stabilise the *cis* form of prolyl peptide bonds (Kofron *et al*, 1991; Kunz *et al*, 2012). However, solubility tests confirmed that the pSer-Substrate was exclusively soluble in DMSO and acetic acid, both of which were unsuitable for the protease-free assay. Further modifications to this sequence could be anticipated in order to design a pNA-bearing peptide which is both a Pin1 substrate and amenable to this assay.

7.2.2. NMR EXchange SpectroscopY (EXSY) assay

As the UV-based activity assay was hindered by the lack of an appropriate Pin1 substrate and difficulties in stabilising the *cis* isomer, the optimisation of an NMR EXchange SpectroscopY (EXSY)-based assay was explored. This method has the advantage of recording kinetic measurements at equilibrium, hence it is not needed to bias the *cis:trans* ratio of substrates as in the fluorometric and UV-based assays. Additionally, it can be applied to an already established Pin1 substrate.

Principle of NMR EXSY experiments

The use of 2D NMR experiments for kinetic measurements was first described by Ernst *et al* (Jeener *et al*, 1979). ^1H - ^1H EXSY experiments consist of a series of 2D ^1H - ^1H NOESY experiments recorded at varying mixing times. The basic scheme of these experiments can be divided into four time periods (figure 76). The experiment consists of three 90° pulses separated by delays t_1 and t_m . First, during the preparation period, a 90° pulse is applied to the sample to rotate the magnetisation from its equilibrium position (along B_0) to the transverse plane. During the evolution period of length t_1 , the nuclei acquire a precession frequency. A second 90° RF pulse is applied, moving the spins back to the z axis. The mixing time (t_m) is the time during which chemical exchange is monitored (through space coupling by Overhauser effect also occurs). After a final 90° pulse, the spins are rotated to the transverse plane once again and the FID is obtained during the detection period. The double Fourier transform of the FID results in the acquisition of a 2D NMR spectrum as a function of two frequency variables: ω_1 and ω_2 , corresponding to the t_1 and t_2 evolution periods, respectively. In the simple case of chemical exchange between two states, A and B, characteristic protons of each state precess at frequencies ω_a and ω_b respectively giving rise to distinct diagonal peaks in EXSY spectra

(figure 77a). During the mixing time (t_m) of the EXSY experiment, the system may or may not undergo chemical exchange between states A and B. If there is no exchange during this period, the resulting NMR spectrum displays only diagonal peaks. These diagonal peaks indicate that the molecule remains in the same state (e.g., state A) both at the beginning and the end of the mixing time. On the other hand, if chemical exchange occurs between states A and B during the mixing time, the spectrum exhibits cross-peaks. These cross-peaks signify that the molecule initiated in one state (e.g., state A) but switched to the other state (e.g., state B) at some point during the mixing time. The integration values of diagonal and exchange cross-peaks can be plotted as a function of mixing time, giving rise to characteristic build-up curves (figure 77b). A mathematical model of the exchange phenomenon facilitates the extraction of kinetic constants through fitting these experimental curves.

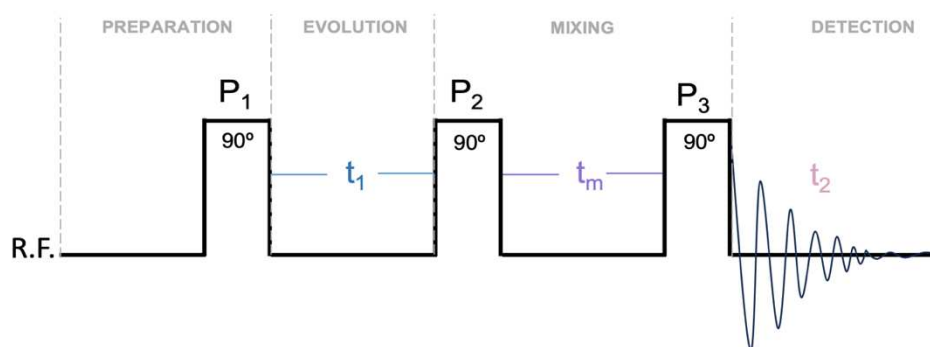


Figure 76: Basic pulse sequence corresponding to the 2D EXSY experiments

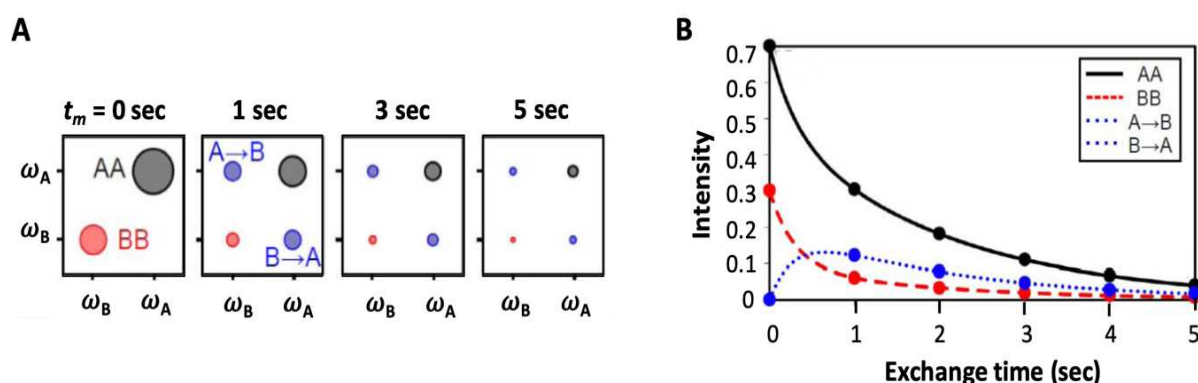


Figure 77: Analysis of data obtained from typical EXSY experiments. **A.** Diagonal peaks of characteristic protons of state A (grey) and B (red) their corresponding cross-peaks (blue). **B.** Characteristic curves of integration value of diagonal and cross-peaks as a function of mixing time. The signal decays as mixing time increases due to relaxation.

Testing pCdc25C as a substrate for EXSY-based assay

To implement the EXSY-based activity assay for Pin1, the first step was to identify a suitable substrate yielding distinct and well-defined exchange peaks for integration. As a well-established substrate for Pin1, the pCdc25C peptide was selected for this assay. The pCdc25C peptide is a peptide derivative of the pThr48-Pro49 site of the Pin1 substrate Cdc25C from *Xenopus* (EQPLpTPVTDL) and is a confirmed substrate for the PPIase domain and ligand of the WW domain (Namanja *et al*, 2011; Wilson *et al*, 2013). NMR EXSY experiments were first conducted on pCdc25C both in the presence and absence of Pin1. In the absence of Pin1, *cis/trans* isomerisation of the pCdc25C pSer-Pro bond occurs on a timescale of several seconds to minutes and hence is not detectable with the typical mixing times used in EXSY experiments (10-500 ms range). Consequently, no exchange peaks related to *cis/trans* isomerisation were observed in the absence of Pin1. However, upon incubation with Pin1 at a 1:40 ratio of Pin1:Cdc25C, the reaction was catalysed to the millisecond timescale, resulting in the appearance of several discernible cross-peaks in the amide region of the spectra (figure 78). To confirm that these cross-peaks indeed represented Pin1-catalysed *cis-trans* prolyl isomerisation, an assignment of the 1D ¹H pCdc25C NMR spectrum was performed. NOESY, TOCSY and 2D ¹H¹³C HSQC assignment experiments were performed on a 1 mM sample of pCdc25C in the presence of 60 μM unlabelled Pin1 and Tris buffer. The analysis of the resulting spectra allowed the assignment of the amide region of the pCdc25C spectrum (figure 79). This assignment confirmed that the peaks undergoing Pin1-induced exchange during the EXSY experiments corresponded to the *cis* and *trans* conformers of residues pThr5 and Val7 of the pCdc25C peptide. These residues flank the proline peptide bond (Pro6) that undergoes isomerisation by Pin1. The proline resonances also experienced significant chemical shift differences between the *cis* and the *trans* forms but its resonances are located in a very crowded region of the spectra. Thus, it can be confidently stated that the observed exchange peaks in the experiment are indeed the result of the *cis-trans* isomerisation of the pThr5-Pro6 bond of pCdc25C catalysed by Pin1. While these resonances are also present in the absence of Pin1, they do not display any discernible cross-peaks. However, it should be noted that the resonances of the minor state (corresponding to the *cis* conformer) are difficult to identify in the absence of cross-peaks as the signal is very weak.

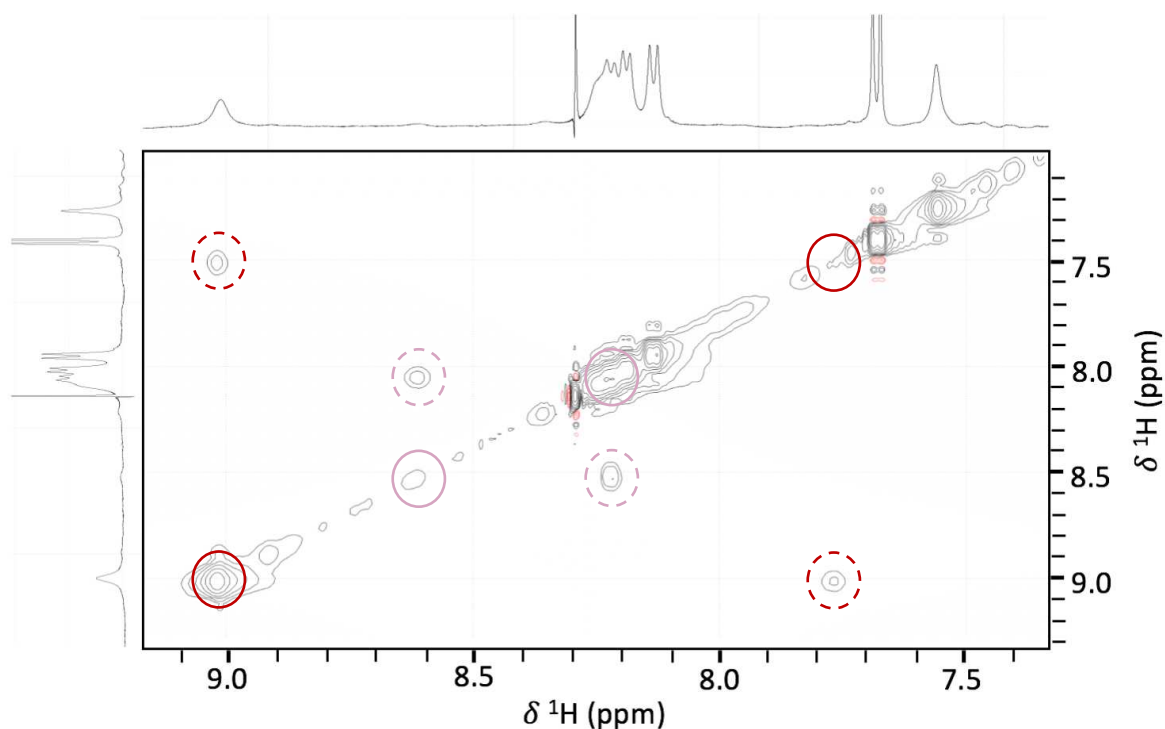


Figure 78: The amide region of the 2D the $^1\text{H}^1\text{H}$ EXSY spectrum of pCdc25C ($t_m = 100$ ms) in the presence of Pin1. Peaks highlighted by red and pink circles indicate peaks exhibiting exchange, cross-peaks are highlighted by dashed circles. Experiments recorded on a 2.5 mM sample of pCdc25C with 60 μM Pin1 at 298K on a 500 MHz spectrometer equipped with cryoprobe.

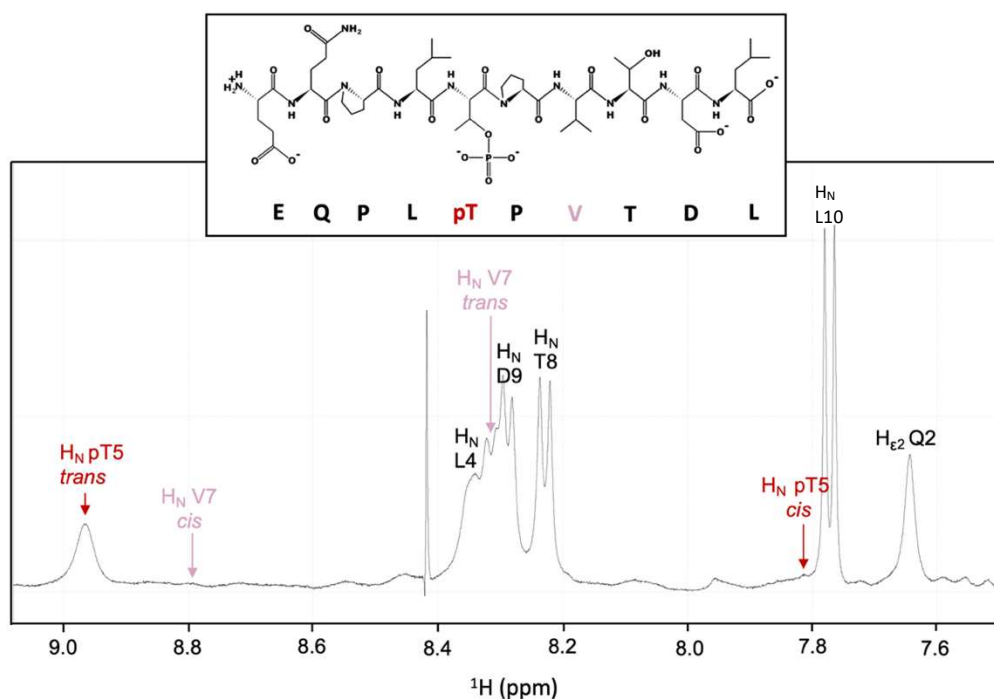


Figure 79: Assignment of the amide region of the 1D ^1H NMR spectrum of pCdc25C in Tris buffer, pH7 and in the presence of 60 μM Pin1. The colour code corresponds to the one used in figure 78.

The exchange peaks corresponding to pT5-H^N were found to be well-isolated, making them suitable for integration and analysis. Originally, the analysis of the integration values for diagonal and corresponding exchange peaks was carried out using the ChemEx program (<https://github.com/gbouvignies/ChemEx>), in collaboration with Dr. Guillaume Bouvignies of the Laboratoire des BioMolécules (LBM), ENS. ChemEx is a powerful software which, based on a given reaction equation and experimental data, can derive a complete set of kinetic constants, including relaxation parameters. However, analysis of the data led to poor fits of the experimental integrations which were not accurate enough for the extraction of precise k_{ex} values (which correspond to the sum of the exchange rate constants). This outcome may be attributed to the analysis being conducted on amide protons prone to water exchange. The integration of the *cis* minor form could also be influenced by the intense surrounding resonance corresponding to H^N-L10. It appeared that the convergence of the fit procedure in ChemEx was very sensitive to any systematic errors in the integrations. It was therefore concluded that it would be more appropriate to analyse isolated and non-exchangeable protons for the extraction of accurate k_{ex} values from the experimental data. Unfortunately, for pCdc25C, there were no discernible cross-peaks corresponding to methyl protons (of the Leu or Thr residues). This absence of observable exchange peaks may imply that the chemical shift values for the two forms are in such close proximity that the cross-peaks appear too near to the diagonal, preventing their detection as distinct peaks. Furthermore, the considerable spectral crowding in this spectral region likely adds to the challenge of identifying discernible cross-peaks. A second well-established Pin1 substrate sequence, known as pintide (WFYpSPR), was also trialled to see if cross-peaks corresponding to methyl protons were present upon incubation with Pin1. However, similarly to pCdc25C, the only well-isolated cross-peaks appearing in the presence of Pin1 corresponded to amide protons.

Therefore a different method for analysis was considered for exchange peaks corresponding to H^N pThr5 of pCdc25C, inspired by the work of the Lippens team, who applied a similar method to compare the rate of *cis/trans* isomerisation catalysed by Pin1 from *Arabidopsis thaliana* (Landrieu *et al*, 2000). Neglecting the spin relaxation at low mixing times, cross-peak intensities in EXSY experiments are proportional to the number of molecules undergoing exchange between the *cis* and *trans* conformers. Increasing the mixing time from 0.5 to a few tens of ms was accompanied by a build-up of the volume of the pT5-H^N cross peaks. By calculating the integrated sum of the cross-peak volumes normalised by the intensity of the corresponding

diagonal peaks, a measure of the population undergoing exchange can be estimated for each mixing time. Note, the normalisation value is taken from the sum of the integrated volumes of the diagonal peaks from the experiment with the shortest mixing time (0.5 ms) and is therefore a function of sample concentration, the effect of relaxation being ruled out. The normalised integration values are then plotted as a function of the mixing time (figure 80). Preliminary EXSY experiments were performed with mixing times ranging from 1 ms to 1 s. However, beyond a mixing time of 100 ms, a significant loss of signal was observed, due to spin relaxation phenomena. To accurately assess the exchange rate, the focus was placed on the initial slope whereby a linear increase in isomerisation is observed with increasing mixing time. To focus on this region, further experiments were conducted at shorter mixing times, ranging from 0.5 to 70 ms.

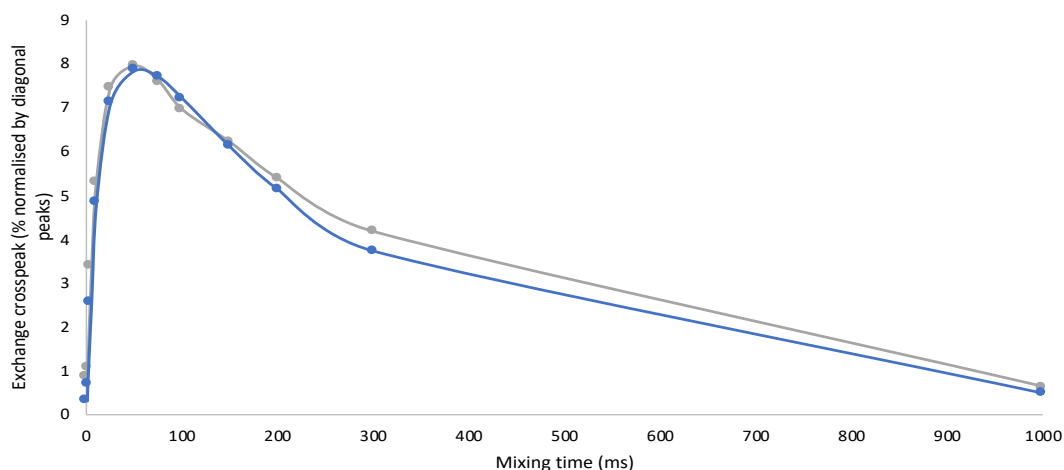


Figure 80: Time dependence of the isomerisation activity of 60 μ M Pin1 on 2.5 mM pCdc25C. The integrated volume of the cross-peak between pT5- H^N protons normalised by the intensity of the corresponding diagonal peaks was plotted as a function of the mixing time. The experiment was repeated to ensure reproducibility (grey and blue). The initial slope is comparable between both repeats. The slower growth and the signal loss after \sim 50 ms were due to relaxation.

Validating the EXSY Pin1 Activity Assay

To confirm that this EXSY-based assay can be used as a reliable measure of Pin1-catalysed prolyl isomerisation, a series of EXSY experiments were performed keeping pCdc25C concentration constant but varying the concentration of Pin1. These experiments were conducted to ensure a linear concentration dependence of Pin1 activity. EXSY experiments

were recorded on a 2.5 mM sample of pCdc25C in the presence of 5, 15, 30, 45 and 60 μM Pin1. Analysis was performed as described above by normalising the summed volumes of the cross-peaks corresponding to pT5- H^{N} by the integrated volume of the diagonal *cis* and *trans* peaks. The normalised values for each concentration of Pin1 were plotted against mixing time (figure 81a). Since relaxation impacts the measurements as soon as linearity is lost, the initial slope provides the most reliable measure of isomerisation. After plotting the value calculated for the initial slope against Pin1 concentration a linear dependence on Pin1 concentration was observed (figure 81b). Although very crude, this approach provides direct evidence of the dose dependent activity of Pin1. An absolute value of the Pin1 activity is clearly not measurable from these slopes, but, prior to the introduction of any inhibitors, they give a reference value for a 100% Pin1 activity in a given condition (enzyme concentration, temperature, pH etc.). Furthermore, at the expense of accuracy but much faster, this approach makes it possible to use the intensity of a single cross peak (obtained in an EXSY experiment at a given short mixing time, *ca.* <10 ms) as a measurement of the 100% Pin1 activity.

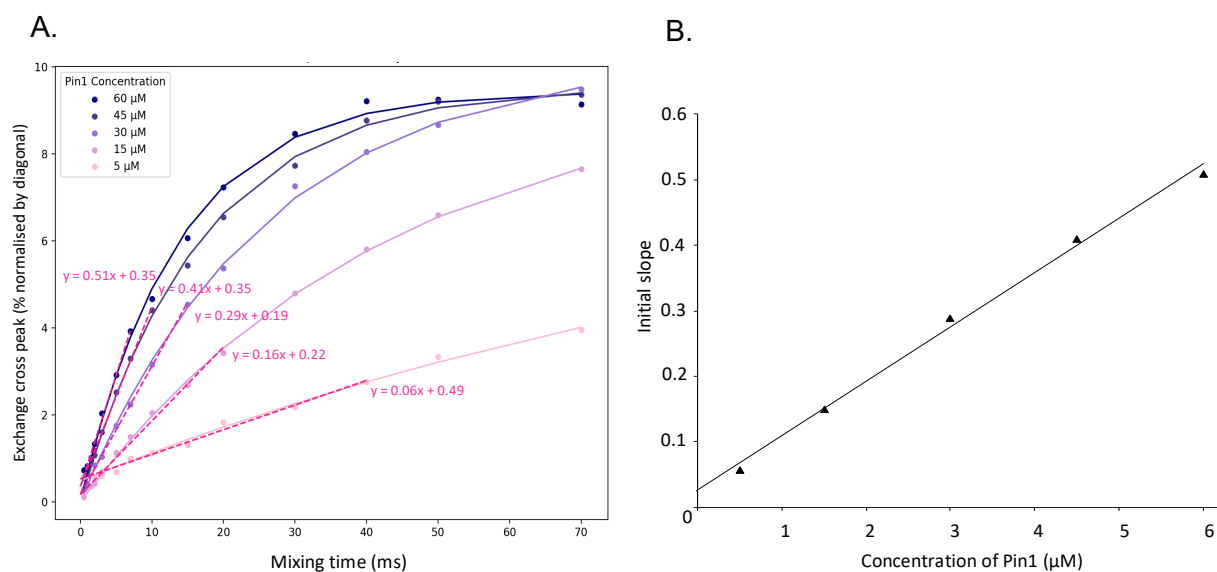


Figure 81: Concentration dependence of the isomerisation activity of Pin1. **A.** The integrated volume of the cross-peak between the *cis* and *trans* pT5- H^{N} protons of pCdc25C in ^1H - ^1H EXSY experiments was normalised by the intensity of the corresponding diagonal peaks and plotted as a function of the mixing time. Experiments recorded on 2.5 mM pCdc25C in the presence of increasing concentrations of Pin1 (5 to 60 μM) using a 500 MHz spectrometer. **B.** The isomerisation rate, estimated by the initial slope (magenta dashed lines), increases linearly with Pin1 concentration.

Inhibition Studies of CZ115: extracting an IC₅₀ value

The IC₅₀ represents the concentration of an inhibitor required to reduce the rate of an enzymatic reaction by 50%. It can be used to directly compare the inhibitory potential of the first- and second-generation ligands against Pin1 provided that they are all tested under the same conditions. The inhibitory effect of ligand CZ115, which demonstrated the highest affinity for the PPIase domain among the first- and second-generation fluorinated ligands, was investigated first. EXSY experiments were conducted on a solution containing 2.5 mM pCdc25C and 60 μM Pin1, while the concentration of CZ115 was varied (0.125 mM, 0.25 mM, 0.75 mM, and 1.25 mM). These concentrations corresponded to 0.05, 0.1, 0.3, and 0.5 molar equivalents of the inhibitor with respect to the substrate. The pT5-H^N diagonal and cross-peaks were integrated, and the cross-peaks were normalised by the sum of the diagonal peaks as described previously. Subsequently, the data was plotted against the mixing time (see figure 82a). The initial slope of the resulting curves was analysed, representing a measure of the exchange rate of Pin1-catalysed prolyl isomerisation. Comparison of the respective slopes allowed calculation of the percentage of remaining Pin1 activity in the presence of each concentration of CZ115, considering that Pin1 activity is 100% in the absence of CZ115 (figure 82b). These values were then plotted as a logarithmic function of CZ115 concentration, facilitating the extraction of the IC₅₀ value of 204 μM (see figure 82c).

These results indicate that CZ115 can reduce Pin1 activity on the pCdc25C substrate by 50% at a CZ115:pCdc25C ratio of ~1:10. As a competitive inhibitor, this could be a direct consequence of CZ115's much greater affinity for the PPIase domain ($K_D \approx 20 \mu\text{M}$) than the pCdc25C substrate ($K_D > 1\text{mM}$) (Wang *et al*, 2015). This observation underscores the intricate mechanism of Pin1, which, facilitated by the WW domain, enables catalysis of *cis-trans* isomerisation even for substrates that are not directly recognised by the PPIase domain. However, it should be noted that the IC₅₀ of a competitive inhibitor for a given substrate is not solely dependent on the difference in affinity for the PPIase domain between the inhibitor and substrate. The affinity of pCdc25C for the WW domain ($K_D \approx 10 \mu\text{M}$) cannot be ignored, as the binding process of pCdc25C to the PPIase domain could be a direct consequence of its interaction with the WW domain.

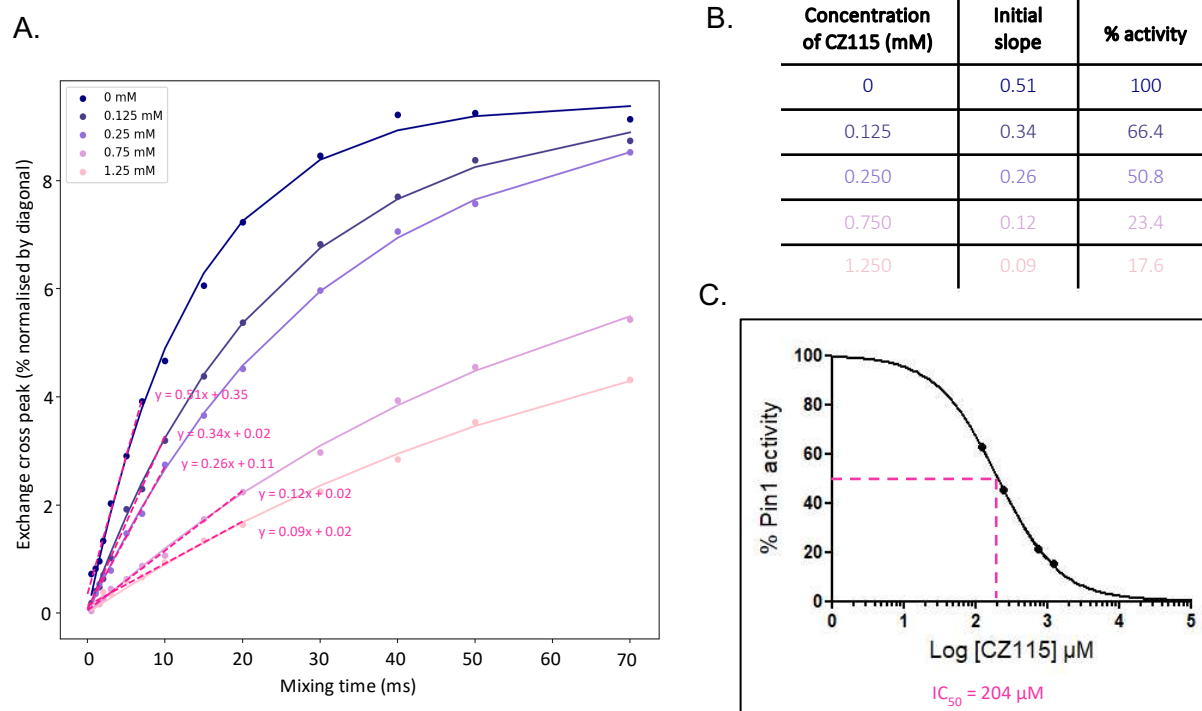


Figure 82: Dependence of the isomerisation activity of Pin1 on the concentration of CZ115. The normalised volumes of the cross-peaks between pT5-H^N protons of pCdc25C in ¹H-¹H EXSY experiments were plotted as a function of the mixing time. In the presence of increasing concentrations of CZ115 (0.125 to 1.25 mM), the isomerisation rate, estimated by the initial slope (magenta dashed lines), decreases. **B.** The initial slope was used to calculate the % of remaining activity of Pin1 for each concentration of CZ115. **C.** Semilog plot of the Pin1 activity versus CZ115 inhibitor concentration using the EXSY-based assay. Extraction of an IC₅₀ value for CZ115 was possible after fitting the plot by nonlinear regression (GraphPad Prism software) with constant slope equations.

Inhibition Studies of CZ40: extracting an IC₅₀ value

The inhibitory effect of a WW domain ligand was assessed next to investigate if a WW domain-specific ligand of similar potency could exert, in an indirect manner, a similar inhibitory effect as a PPIase domain-specific ligand. Ligand CZ40 of the first-generation ligands was selected due to availability of the ligand, and it displayed a very similar K_D value for the WW domain as CZ115 displayed for the PPIase domain (see table 17). Notably, CZ40 also displays weak binding to the PPIase domain but this is suspected to be a non-specific binding event due to the K_D value being 20-times higher than for the WW domain. This is supported by the fact that only four residues of the PPIase domain display significant CSPs upon binding to 8 molar

equivalents of Pin1 (Gly123, Met130, Gln131, and Gly155). Additionally, none of the residues within the PPIase interdomain interface exhibit significant CSPs. This, in conjunction with the considerably higher K_D values observed for the affected residues within the PPIase domain, provides evidence that the impact on the PPIase domain is not a consequence of an allosteric effect stemming from CZ40's binding to the WW domain. Additionally, the absence of perturbations within the phosphate-binding site of the PPIase domain, where perturbations are typically observed upon specific binding (even in cases of low-affinity binding, as demonstrated with VB66), suggests that there is no distinct binding interaction occurring with this domain.

Ligand name	Structure	Kd (μ M)	
		WW	PPIase
CZ115	Ac-Cha-pSer- $\Psi^{CF3,H}$ Pro-reduced Tryptamine	NB	20 \pm 9
CZ40	Fmoc-pSer-Pro-Tryptamine	26 \pm 6	438 \pm 80

Table 17: Comparison of K_D values of CZ115 and CZ40 for Pin1.

EXSY experiments were conducted under the same conditions, on a solution containing 2.5 mM pCdc25C and 60 μ M Pin1, while the concentration of CZ40 was varied (0.125 mM, 0.25 mM, 0.75 mM, and 1.25 mM). The spectra were analysed in the same way as described previously. Subsequently, the data was plotted against the mixing time (see figure 83a). Comparison of the initial slopes allowed calculation of the percentage of remaining Pin1 activity in the presence of each concentration of CZ40, considering that Pin1 activity is 100% in the absence of CZ40 (figure 83b). These values were then plotted as a logarithmic function of CZ40 concentration, facilitating the extraction of the IC_{50} value of 287 μ M (see figure 83c). These results suggest that CZ40 can reduce Pin1 activity on the pCdc25C substrate by 50% at a CZ40:pCdc25C ratio of \sim 1:10. The IC_{50} value extracted is very similar to that of CZ115, indicating that masking the WW domain effective for limiting catalytic activity on pCdc25C. The affinity of pCdc25C for the WW domain has been calculated in several reports, yielding different K_D values (see table 18). Generally, these K_D values fall within the range of approximately 10 μ M to 100 μ M, with discrepancies persisting even when experimental conditions were similar, as seen between the Namannja and Wang teams. Due to these

discrepancies, our lab previously determined the K_D of the pCdc25C peptide to be $\sim 10 \mu\text{M}$ under conditions of 30 mM Tris, 50 mM NaCl, 2 mM DTT, pH 7. These conditions most closely resemble those of the EXSY experiments (50 mM Tris, 50 mM NaCl, 2 mM DTT, pH 7), consequently, this value can be considered the most reliable for these experiments. This K_D value is within the same order of magnitude as the K_D of CZ40 for the WW domain of Pin1 ($\sim 20 \mu\text{M}$). The fact that both the ligand and substrate bind to the same domain, with a similar affinity, along with the fact that no allosteric effect of CZ40 binding in the PPIase domain has been observed, means that this again represents a case of competitive inhibition.

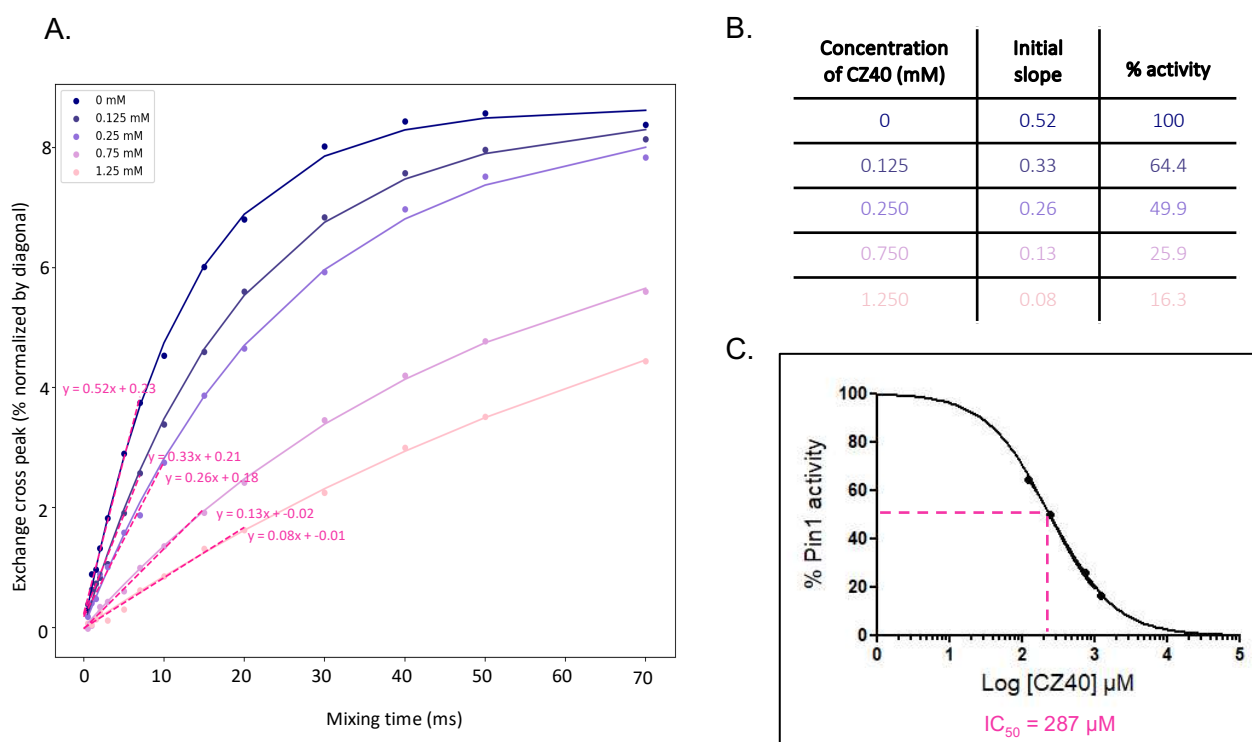


Figure 83: Dependence of the isomerisation activity of Pin1 on the concentration of CZ40. The normalised volumes of the cross-peaks between pT5- H_N protons of pCdc25C in ^1H - ^1H EXSY experiments were plotted as a function of the mixing time. In the presence of increasing concentrations of CZ40 (0.125 to 1.25 mM), the isomerisation rate, estimated by the initial slope (magenta dashed lines), decreases. **B.** The initial slope was used to calculate the % of remaining activity of Pin1 for each concentration of CZ115. **C.** Semilog plot of the Pin1 activity versus CZ40 inhibitor concentration using the EXSY-based assay. Extraction of an IC_{50} value for CZ40 was possible after fitting the plot by nonlinear regression (GraphPad Prism software) with constant slope equations.

Report	K _D (μM)	Pin1 construct	Buffer conditions	Temp (K)
Wintjens et al., 2001	100	WW domain	50 mM Tris, 100 mM NaCl, pH 6.4	285
Jacobs et al., 2003	117	FL Pin1	100 mM Imidazole, 100mM NaCl, 5 mM DTT, 0.03% NaN ₃ , pH 6.6	298
Namanja et al., 2007	120	FL Pin1	30 mM imidazole-D ₄ 30 mM NaCl, 5 mM DTT, 0.03% NaN ₃ , pH 6.6	295
Wang et al., 2015	9	FL Pin1	30 mM imidazole-D ₄ , 30 mM NaCl, 5 mM DTT, 0.03% NaN ₃ , pH 6.6	295
Zhu et al., 2019	6.2	WW domain	50 mM Tris, 5 mM DTT, 0.3% NaN ₃ , pH 6.8	298
	6.6	FL Pin1		

Table 18: Dissociation constants calculated for the pCdc25C peptide for the WW domain (isolated or within FL Pin1) in different reports.

7.3 Discussion and perspectives

7.3.1. Ligand binding studies: Unraveling binding modes and key affinity-determining groups

In studying the binding interactions of the second-generation series of ligands with Pin1, several key findings emerged. The initial aim was to build upon the first-generation ligands produced by the lab, to explore additional strategies for improved potency and bioavailability. Notably, the presence of the phosphate group presents a major drawback of Pin1 peptide ligands for drug application due its negative effect on cell permeability. Here, it was shown that the removal or replacement of the phospho-serine residue was detrimental to affinity, even in the presence of *N*- and *C*-terminal aromatic. Noticeably, *N*-terminal Fmoc group protection consistently improved binding, aligning with Pin1's preference for aromatic motifs flanking pSer-Pro (Wildemann et al., 2006). The incorporation of the aromatic naftylalanine group (Nal) also appeared to enhance binding, with GG201 (Fmoc-pSer-Pro-Nal-Gly-NH₂) displaying the highest affinity binding of the second-generation ligands. However, increasing the aromatic content of the ligands also risked decreasing their solubility in aqueous solution, as was observed for GG199 (Fmoc-pSer-Ψ^{CF₃,H}Pro-Nal-Gly-NH₂). In particular, it would be interesting to explore the Fischer-like ligands further (incorporating the Nal group), including the production of peptides containing the pSer-Ψ^{CF₃,H}Pro moiety which, as we have

shown, specifically target the PPIase-domain. While this was attempted with GG199, enhancing solubility, perhaps by replacing Gly or by extending the peptide sequence with a more polar residue like Gln, could facilitate their analysis. As observed on GG213 (Fmoc-pCF₂-Ser-Ψ^{CF₃,H}Pro-Tryptamine), the use of difluoro-phosphonated serines also appeared very valuable to mimic the phosphate group, although the aromatic Fmoc *N*-capping group was required to retain affinity. Then, *C*-extensions of GG199 and GG213 that could improve water solubility or prevent aggregation of the Pin1 complexes appear to be a promising perspective for the design of potent Pin1 inhibitors.

In light of our persistent challenge in achieving sub-micromolar affinity, we explored an alternative strategy for Pin1 inhibition through targeted degradation. Surprisingly, lower-affinity ligands can be advantageous in PROTAC technology. They facilitate the dissociation of the PROTAC chimera from Pin1 after ubiquitination, allowing it to target other Pin1 molecules. Consequently, for this mode of inhibition, lower affinity can translate to lower drug doses. The incorporation of an alkyne hook at the *N*-terminus of CZ40, designed for conjugation to a linker-thalidomide module, did not significantly reduce affinity (LME27). A general trend for the PROTAC ligands was that, similarly to their parent compounds, CZ40 and CZ55, they all displayed specific binding to the WW domain. Furthermore, binding to the WW domain seemingly induced an allosteric effect in the PPIase domain, as evident from characteristic CSP patterns. This was characterised by minimal perturbations to residues within the PPIase phosphate binding pocket (residues 64-72) and the presence of a cluster of CSPs at the α4-helix of the ID interface. Additionally, calculated K_D values for these residues, comparable to those in the WW domain binding site, suggested they resulted from the same binding event. The discovery of lead pre-PROTAC compound LME27 inspired the synthesis of complete PROTAC molecules LME66 and LME34. Unfortunately, LME66's limited solubility in aqueous solutions prevented reaching binding saturation, hindering extraction of an accurate K_D value. Nevertheless, noticeable CSPs even at low LME66 concentrations suggest its binding affinity is comparable to its parent compound LME27. To facilitate accurate K_D value determination, a cross-titration in the presence of DMSO could be considered. This would involve first producing a 3mM ligand stock solution in DMSO, before introducing to aqueous solution, hopefully rendering it soluble. Considering the DMSO effect on Pin1 resonances that we observed, the concentration of DMSO should be kept constant throughout the titration to mitigate this effect. A cross-titration setup would allow this, whereby two NMR

tubes would be prepared: one containing Pin1 with DMSO and buffer and another identical tube with an additional 6 molar equivalents of LME66. By mixing the two tubes at different volumes, intermediary equivalents can be obtained (see Materials and Methods section). Nevertheless, the potential interference of DMSO on the interaction should be considered.

7.3.2. Activity assay optimisation: Bridging the gap between affinity and inhibition

The affinity of a ligand for Pin1 does not necessarily imply inhibitory potential. Therefore, a pivotal aspect of this project was to refine a robust Pin1 activity assay for assessing the inhibitory capabilities of the ligand series. First, we provided evidence that chymotrypsin, a key component of the protease-coupled assay, had the ability to cleave Pin1, with Pin1 being completely denatured in the presence of 0.3 equivalents of chymotrypsin. This offers an explanation for the strange results obtained previously, which suggested that ligands CZ115 and CZ52 act as activators at lower concentrations of ligand. Notably, the dramatic effect of chymotrypsin on the structural integrity of Pin1 sheds some doubt on values obtained with this assay in the literature.

In terms of the UV assay, the absence of a suitable substrate ultimately hindered the assessment of ligand inhibitory potential. As suspected, the Ala-substrate, which lacks the Pin1 recognition module, did not serve as a substrate for Pin1. Furthermore, the LiCl/TFE solution could not adequately stabilise the *cis* conformation of the Glu-substrate, highlighting the solvent's limitations for maintaining the *cis* conformation. This underscores the need to assess *cis* and *trans* populations before performing UV or Fluorometric assays, which also rely on LiCl/TFE for stabilisation of the *cis* conformation. To continue optimising this assay, modifying substrate design is a viable option. While the pSer-substrate (Suc-Ala-pSer-Pro-Phe-pNa) is presumed to be a superior substrate for Pin1, it encountered solubility issues in LiCl/TFE. TFE does not readily solubilise polar compounds, hence the incorporation of the phosphate group was likely detrimental for solubility. To improve solubility, introducing hydrophobic groups, such as additional aromatic amino acids like Phe, Trp, or Tryptamine, into the substrate design may be worthwhile.

Despite the lack of success of the UV- and protease-coupled fluorometric assays, the optimisation of the EXSY-based assay described here proved to provide a reliable measure of

Pin1 activity in the presence of different inhibitors. While ultimately, the assay could not provide an absolute value of Pin1 activity, nor facilitate the extraction of k_{ex} values, it provides a reliable measure of Pin1 activity in the presence of inhibitors in a given condition. This makes it a suitable method for extracting comparable IC_{50} values for both first- and second-generation ligands. Notably, this assay offers distinct advantages over the aforementioned biochemical assays. Firstly, it maintains a simplified system, comprising only Pin1, substrate, and inhibitor, eliminating additional components like TFE or proteases. Secondly, it grants flexibility in substrate choice, allowing us to work with established Pin1 substrates, without reliance on pNA-containing compounds. Furthermore, it facilitates the recording of kinetic measurements at equilibrium, eliminating the need to bias the *cis:trans* ratio of substrates. Importantly, unlike UV and fluorometric assays that only monitor *cis-to-trans* isomerisation, the EXSY assay monitors both the forward and reverse reaction, providing a comprehensive view of Pin1's enzymatic activity.

However, the assay does present some disadvantages. Firstly, it demands a substantial amount of substrate (2.5 mM) to achieve a satisfactory signal-to-noise ratio, consuming approximately 1.5 mg of peptide per experiment. Presently, the commercially acquired pCdc25C peptide costs €35 per milligram, rendering a full set of EXSY experiments for each inhibitor relatively expensive. To mitigate this cost, future endeavours could involve the routine in-house synthesis of the pCdc25C peptide through collaboration with our partners at Cergy-Pontoise. In addition, the assay is time-consuming, with a single EXSY experiment taking around 21 hours. One method to significantly reduce the time of this assay, albeit at the expense of accuracy, would be to run the experiment with only two mixing times: 0.5 ms for the extraction of the diagonal peak intensities (which act as a normalisation factor) and a mixing time <10 ms for the extraction of cross-peak intensities. In this region (<10 ms), the normalised cross-peak intensities show a linear relationship with Pin1 concentration (see figure 81b). Therefore, at a fixed mixing time in this range, in the absence of inhibitor, the normalised intensity of cross-peaks serves as a reference value representing 100 % Pin1 activity. Similarly, the normalised volume of cross-peaks at this same fixed mixing time, in the presence of increasing concentrations of inhibitor, could be used to calculate the percentage of remaining Pin1 activity. After scrutinising the data derived from the evaluation of two inhibitors, we determined that a mixing time of 7 ms consistently delivered a dependable assessment of Pin1 activity across different inhibitor concentrations. The activity values obtained at this mixing

time closely matched those calculated from the initial slope of the complete EXSY experiments.

To further validate this assay and provide a robust benchmark, it would be beneficial to include a positive control using a well-characterised Pin1 inhibitor, such as juglone. Determining the IC_{50} value for this inhibitor using the EXSY-based assay described here serves multiple purposes. Firstly, it enhances the credibility of our assay by demonstrating its ability to accurately assess the inhibitory potential of known compounds. Secondly, it allows for a direct and meaningful comparison between our ligand series and a widely acknowledged Pin1 inhibitor, thereby placing our compounds within the context of established inhibitors. It is important to note that the IC_{50} values presented in this thesis for ligands CZ115 and CZ40 should be interpreted with consideration for the unique experimental conditions employed, making direct comparisons with literature values challenging. IC_{50} values are notably influenced by both enzyme and substrate concentrations used in the assay. Therefore, the IC_{50} values obtained here can be directly compared only when the same conditions are applied. Consequently, for our ligand series, we intend to conduct further testing with the EXSY assay to identify the most effective inhibitors.

7.3.3. Targeting the WW domain to inhibit activity of the PPIase domain

The IC_{50} values calculated for CZ115 and CZ40 from the first-generation peptide ligands are directly comparable considering they were tested under the same conditions. Their respective IC_{50} values ($IC_{50}(CZ115) = 210 \mu M$, $IC_{50}(CZ40) = 287 \mu M$) are within the same order of magnitude suggesting they exert a very similar inhibitory effect on Pin1 catalysis of pCdc25C. A key difference between the ligands is that CZ115 specifically and exclusively binds to the PPIase domain, while CZ40 targets the WW domain. While it is reasonable to anticipate inhibitory potential from a PPIase-domain ligand, given its higher affinity for the PPIase domain ($K_D = 20 \mu M$) compared to pCdc25C ($K_D > 1mM$), the fact that CZ40 exhibits a similar potency is more surprising. These results suggest that targeting the WW domain can inhibit catalytic activity in the PPIase domain just as effectively as occlusion of the catalytic site. As no observable allosteric effect on the PPIase domain was noted upon CZ40 binding to the WW domain, we can assume that the observed inhibition can be solely attributed to the competition of CZ40 and pCdc25C for the WW domain. This could indeed provide evidence

that, for this substrate, binding to the WW domain is a pre-requisite for its isomerisation by the PPIase domain. The pCdc25C substrate notably displays very low affinity for the PPIase domain (estimated $K_D > 1$ mM), while displaying significantly higher affinity for the WW domain ($K_D \approx 10$ μ M). Considering that pCdc25C is indeed a substrate of the PPIase domain, as confirmed by our EXSY assays, this suggests a direct role of the WW domain in targeting pCdc25C to the PPIase domain. According to the sequential binding model (see chapter 3.4), binding to the WW domain is proposed to produce an entropic gain by increasing the local concentration of substrate so that it can bind to, and be isomerised by, the PPIase domain. Hence the two ligands described here could be targeting different steps of this binding and catalytic process, with CZ40 targeting the first step (binding to the WW domain) and CZ115 targeting the second step (catalysis by the PPIase domain) (see figure 84).

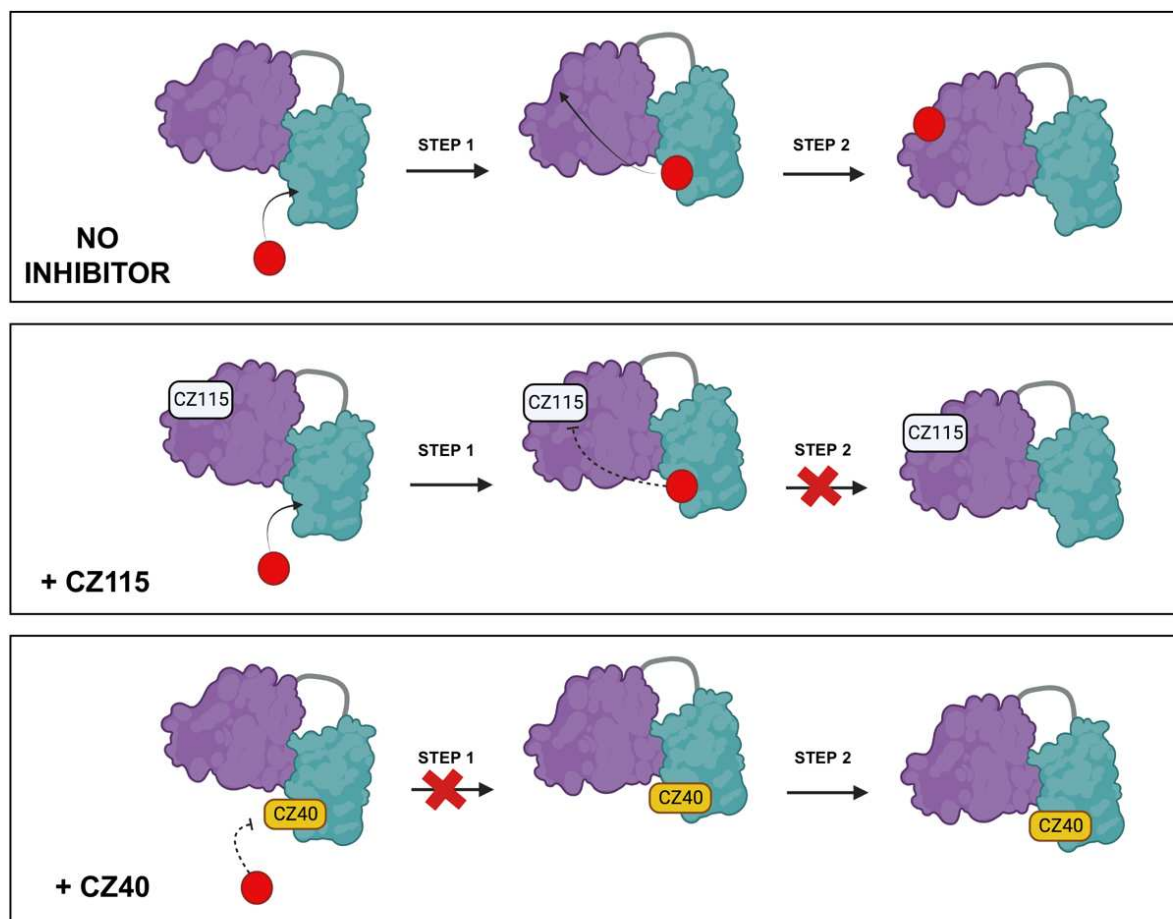


Figure 84: A potential two-step binding model to explain the comparable inhibitory effect of CZ115 and CZ40. Upper panel: In the absence of inhibitor, pCdc25C (red circle) binds to the WW domain (step 1) which increases the local concentration of pCdc25C so that it can then bind to the PPIase domain (step 2). **Middle panel:** The binding of CZ115 to the PPIase domain blocks step 2 of this binding and targeting process. **Lower panel:** The binding of CZ40 to the WW domain blocks step 1 of this targeting process.

As both inhibitors target two steps of the same substrate targeting process, it can explain why they produce such similar inhibitory effects. These results underscore the complexities surrounding Pin1 substrate binding and catalysis. In order to further test this hypothesis of a two-step binding, the EXSY experiments could be repeated with the isolated PPIase domain in place of FL Pin1. Using the isolated PPIase domain in the assay, we would expect that CZ40 displays a significantly lower inhibitory effect than CZ115.

What further complicates this model, is the fact that the isolated PPIase domain has been shown to exhibit catalytic activity for pCdc25C. Remarkably, some studies have even suggested that the isolated PPIase domain displays a higher isomerase activity than FL Pin1 for pCdc25C (Wang *et al*, 2015; Wilson *et al*, 2013). This is in accordance with other observations of the WW domain exhibiting an inhibitory allosteric effect on the PPIase domain. Interestingly, the binding of pCdc25C to the WW domain is associated with an allosteric effect, decreasing interdomain contacts and favouring the adoption of the extended state of Pin1 (Jacobs *et al*, 2003; Wang *et al*, 2015). This loss of interdomain contacts is correlated with enhanced PPIase activity for this substrate (Wilson *et al*, 2013). Therefore, another potential hypothesis to explain the ability of CZ40 to inhibit catalysis of pCdc25C could be that it prevents this allosteric effect by competing with the substrate for the WW domain binding site. Consequently, in the absence of pCdc25C binding, the WW domain maintains its inhibitory effect on the PPIase domain. If this model stands, it could indeed suggest that CZ40's impact on PPIase activity may vary depending on the substrate. For example, for the Pintide, which seems to have no effect on interdomain contacts when binding to the WW domain, and which also independently shows binding to the PPIase domain. It would hence be intriguing to explore whether CZ40 can demonstrate similar inhibitory potential on other Pin1 substrates.

It is essential to note that these models are simply qualitative hypotheses of how we can interpret the ability of a WW domain specific ligand to inhibit isomerase activity of the PPIase domain as efficiently as an inhibitor which directly occludes the catalytic site. Clearly, the mechanism of Pin1 substrate binding and catalysis is complex and further studies would be required to provide a more robust model.

Chapter 8 Structural characterisation of the interaction between Pin1 and the IDR of Cdc25C

To gain more insight into the mechanism underlying Pin1's substrate interactions, an investigation into its binding with the IDR of Cdc25C was conducted. As a reminder, the IDR of Cdc25C serves as the host to a series of potential phosphorylation sites, specifically six S/T-P motifs, which, when phosphorylated, are anticipated to constitute potential binding sites for Pin1. Importantly, prior research has already confirmed the phosphorylation of five of these motifs within the mitotic form of Cdc25C, which is a well-characterised binding partner of Pin1. Thus far, the absence of structural studies limits the full understanding of this interaction. Due to the presumed intrinsic disorder of this region within Cdc25C, NMR spectroscopy will be employed to gain insights at the residue level. To enhance the relevance of this study, the hyperphosphorylated state of the Cdc25C IDR will be explored, utilising phosphomimetic mutants and *in vitro* phosphorylated forms.

8.1 Overview of structural studies of IDPs using NMR spectroscopy

Intrinsically Disordered Proteins (IDPs) and Intrinsically Disordered Regions (IDRs) have emerged as intriguing subjects of study within the realm of structural biology. These biomolecules defy the conventional structure-function paradigm, as they characteristically lack a well-defined three-dimensional structure in their native state (Kosol *et al*, 2013). Instead, IDPs exhibit substantial flexibility and conformational heterogeneity; it is more fitting to describe such proteins as constantly sampling a range of conformational ensembles. This inherent adaptability equips them to dynamically interact with a myriad of binding partners and actively engage in diverse cellular processes. Sequence analysis of complete proteomes indicated that at least 40% of human proteins are likely to contain long disordered regions (> 30 amino acids) (Oates *et al*, 2012). The importance of IDPs and IDRs is accentuated by their prevalence in forming regulatory regions of proteins and frequently participating in protein-protein interactions and complex cell signalling pathways.

When it comes to studying their structural characteristics, methods like X-Ray Crystallography are limited to identifying the presence of disorder, often observed as missing electron density

or the inability to form crystals. This arises from the fundamental requirement of well-ordered crystalline structures, which struggle to accommodate the dynamic and variable conformations of disordered or flexible protein regions. By contrast, the intrinsic flexibility of IDPs works to the advantage of NMR spectroscopy. This increased flexibility enhances their relaxation properties, which, consequently, increases spectral signal-to-noise ratio in comparison to a folded protein of equivalent size, resulting in sharp, narrow peaks in NMR spectra, even of large proteins. As a result, NMR has been shown to facilitate the structural *and* dynamic characterisation of IDPs. The dynamic nature of IDPs is crucial for their function, as conformational flexibility is closely associated with their ability to recognise targets and establish dynamic complexes. While techniques like electron spin resonance (ESR) and fluorescence correlation spectroscopy (FCS) can provide some molecular insight into flexibility and conformational dynamics, NMR spectroscopy offers residue-level information about dynamics throughout the protein. Notably, ^{15}N relaxation measurements, including R_1 and R_2 , and steady-state ^1H - ^{15}N heteronuclear NOEs (hNOEs) are particularly powerful tools for characterising the backbone dynamics of IDPs in the picosecond-to-nanosecond timescale.

However, NMR investigations of disordered proteins still pose several challenges. As for globular proteins, micromolar to millimolar sample concentrations are required as well as a high level of purity. Additionally, due to the inherent susceptibility of IDPs to proteolytic degradation, the integrity of samples needs to be monitored regularly throughout the course of data acquisition. An important characteristic of the NMR analysis of IDPs is the weak spectral dispersion observed in the proton dimension, a phenomenon particularly pronounced for amide protons. This results from the uniform chemical environment experienced by amide protons, mainly influenced by their interaction and exchange with water. By consequence, this can lead to severe spectral overlap, requiring higher spectral resolution afforded by the use of higher field spectrometers. Spectral resolution of IDPs can also be increased by employing specific NMR methods such as carbon-detected experiments, to benefit from the dispersion of ^{13}C chemical shifts, and high-dimensional (4D and 5D) experiments which are possible thanks to the favourable transverse relaxation properties of IDPs (Felli & Pierattelli, 2014). Combined with fast acquisition schemes and/or non-uniform sampling, high-dimension NMR experiments have allowed for the assignment of backbone and side chain resonances of large IDPs (> 300 amino acids) exhibiting severe overlap in typical 2D and 3D spectra (Nováček *et al.*, 2014). Another feature of the NMR investigation of IDPs is the exchange rate of amide

protons with water, which in in the intermediate to fast exchange under physiological conditions, potentially leading to extensive line broadening or complete signal loss (Gil *et al*, 2013). To address this, spectroscopists often opt for low pH and/or low temperature conditions to slow down water exchange, albeit at the risk of moving further from physiological conditions.

The last but very important challenge of the NMR investigation of IDPs is the structural characterisation of their interactions. Indeed, the interaction of an IDP with a protein partner can lead to folding of a local sequence motif while most parts of the IDP usually remain disordered. By contrast, some IDPs remain fully disordered when interacting with their protein partners, forming fuzzy complexes (Mittag *et al*, 2010; Tompa & Fuxreiter, 2008). The structural investigation of protein complexes by NMR remains challenging for large systems (> 30 kDa) and this includes the region of a disordered protein that folds upon binding to a large domain. This part of the IDP has comparatively unfavourable relaxation properties, behaving as part of a large domain with slow tumbling due to the drag of the disordered region. Fast relaxation leads to weak and broad signals that are difficult to identify among the many intense and sharp signals from the regions that retain disorder. Furthermore, conformational dynamics on μ s-ms time scale are often observed in complexes involving IDPs, which usually results in dramatic line broadening. To overcome these issues, NMR methods based on chemical exchange (CPMG and CEST experiments) were developed to characterise the bound form of “invisible” IDRs that fold upon binding and/or show conformational dynamics in the unfavourable intermediate exchange regime (Charlier *et al*, 2017).

Another important consideration is how phosphorylation can modulate the structural and dynamic properties of IDPs. While IDPs typically maintain their disordered nature upon phosphorylation, it is noteworthy that phosphorylation, especially of serine and threonine residues, can influence the local secondary structure propensity within these proteins (Newcombe *et al*, 2022). Markedly, phosphorylation introduces additional volume and two negative charges. This alteration in charge distribution can lead to changes in hydrogen bonding patterns and electrostatic interactions within the backbone, thereby influencing the local structural characteristics of IDPs. Moreover, multi-site phosphorylation events have, rarely, been shown to induce large-scale folding transitions in disordered proteins. A

remarkable example of this phenomenon is described in a study conducted by the Forman-Kay lab, where multisite phosphorylation led to the folding of a previously disordered protein (Bah *et al*, 2015).

8.2 NMR investigation of unphosphorylated WT Cdc25C⁴⁴⁻²²⁸

8.2.1. Backbone assignment of unphosphorylated WT Cdc25C⁴⁴⁻²²⁸

Studying the interaction between Pin1 and Cdc25C⁴⁴⁻²²⁸ by NMR first required the assignment of the backbone resonances of each protein in the free form. The backbone resonances of Pin1 at pH 7, 298 K were previously assigned (see chapter 7.1.1). However, as no NMR characterisation of the IDR of human Cdc25C had previously been performed, it was necessary to begin this investigation by fully assigning the backbone resonances of unphosphorylated Cdc25C⁴⁴⁻²²⁸. Initially, preliminary 2D and 3D NMR experiments (¹H-¹⁵N HSQC, HNCOC, HNCA, HNCACB, HN(CA)CO, CBCA(CO)NH) were recorded on a ¹⁵N/¹³C-labelled sample of WT Cdc25C⁴⁴⁻²²⁸ using a 500 MHz spectrometer equipped with cryoprobe. The resulting HSQC spectrum of Cdc25C⁴⁴⁻²²⁸ was characteristic of an IDP, with a typical weak dispersion of signals in the proton dimension (see figure 85). The strong signal overlapping observed in the 3D spectra of Cdc25C⁴⁴⁻²²⁸ recorded at 500 MHz (in particular in the HNCA and HNCACB spectra) barely allowed for the discrimination between resonances and prevented the full assignment of the protein.

Therefore, to perform the full backbone assignment of Cdc25C⁴⁴⁻²²⁸, access to a high magnetic field spectrometer was essential in order to improve signal resolution. For this purpose, we used the 950 MHz spectrometer of the ICSN (Gif-sur-Yvette), available through the Infranalytics research federation of the CNRS. Experiments were recorded on a uniformly labelled sample of ¹⁵N/¹³C Cdc25C⁴⁴⁻²²⁸ (500 μ M) in 40 mM Bis-Tris pH 6.50, 100 mM NaCl, and 2 mM DTT. Assignment experiments were performed under low pH and temperature conditions (pH 6.50, 283 K) to limit water exchange of amide protons. To assign H^N, N^H, CO, C α , and C β resonances, seven ¹H/¹⁵N/¹³C triple resonance 3D experiments were recorded: HNCOC, HNCA, HNCACB, HN(CA)CO, CBCA(CO)NH, HN(COCA)N and HN(CA)N. The combined analysis of these 3D experiments allowed the identification of 178 spin systems with CO, C α , and C β chemical shift values for residues *i* and *i-1* of each amide NH group.

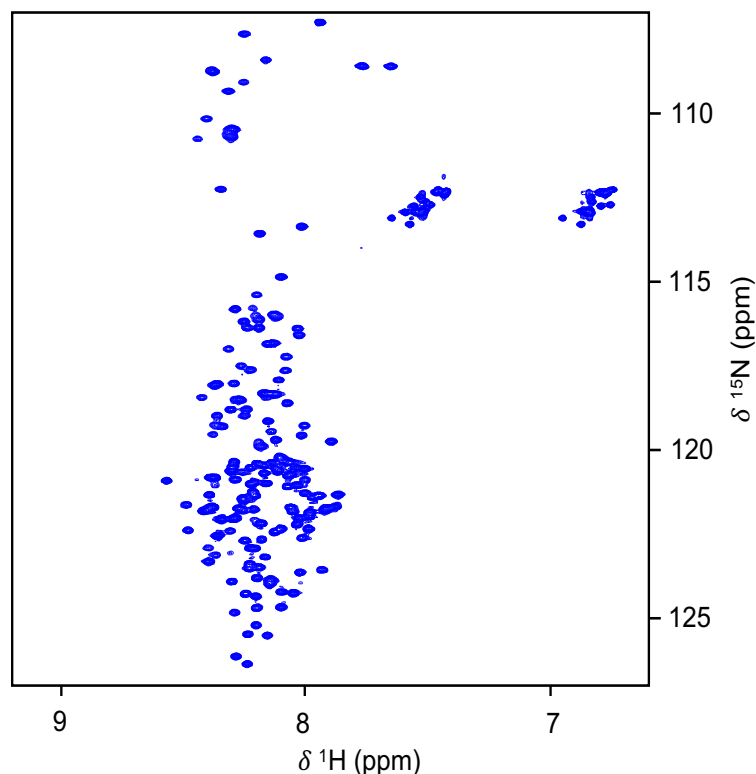


Figure 85: ^1H - ^{15}N -HSQC folded spectrum of apo Cdc25C⁴⁴⁻²²⁸ in Bis-Tris buffer (pH 7) recorded at pH 6.5, 298K using a 500 MHz spectrometer.

Subsequently, a sequential assignment was performed by comparing the corresponding chemical shift values of each spin system. The assignment was aided by identifying characteristic $\text{C}\alpha$ and $\text{C}\beta$ chemical shift values of particular amino acids (e.g., $\delta \text{C}\alpha_{\text{Gly}} \approx 45$ ppm; $\delta \text{C}\beta_{\text{Thr}} \approx 68$ ppm; $\delta \text{C}\beta_{\text{Ala}} \approx 19$ ppm). Additionally, to confirm the accuracy of the assignment, HN(COCA)N and HN(CA)N experiments enabled a comparison of the backbone N^{H} chemical shift values among residues i , $i-1$, and $i+1$. Regarding proline residues that are devoid of amide protons, their CO, $\text{C}\alpha$, and $\text{C}\beta$ chemical shifts were obtained from residues $i+1$. The sequential assignment process of residues T83-L87, focussing on CO, $\text{C}\alpha$, $\text{C}\beta$, and N^{H} nuclei, is illustrated in Figures 86 to 89 respectively. Overall, the analysis of 3D spectra allowed for the complete backbone assignment of the Cdc25C⁴⁴⁻²²⁸ construct that includes 4 cloning residues at the *N*-terminus (figure 90). In particular, all H^{N} , N^{H} , CO, $\text{C}\alpha$, and $\text{C}\beta$ resonances of the 174 non-proline residues, and all CO, $\text{C}\alpha$, and $\text{C}\beta$ resonances of the 11 proline residues could be assigned (see annex 1). Despite the presence of 11 proline residues in the sequence of Cdc25C⁴⁴⁻²²⁸, only two additional peaks with very low intensity were observed in

the HSQC spectrum, corresponding to minor forms of Thr172, suggesting that all proline residues adopt a predominant conformation in solution.

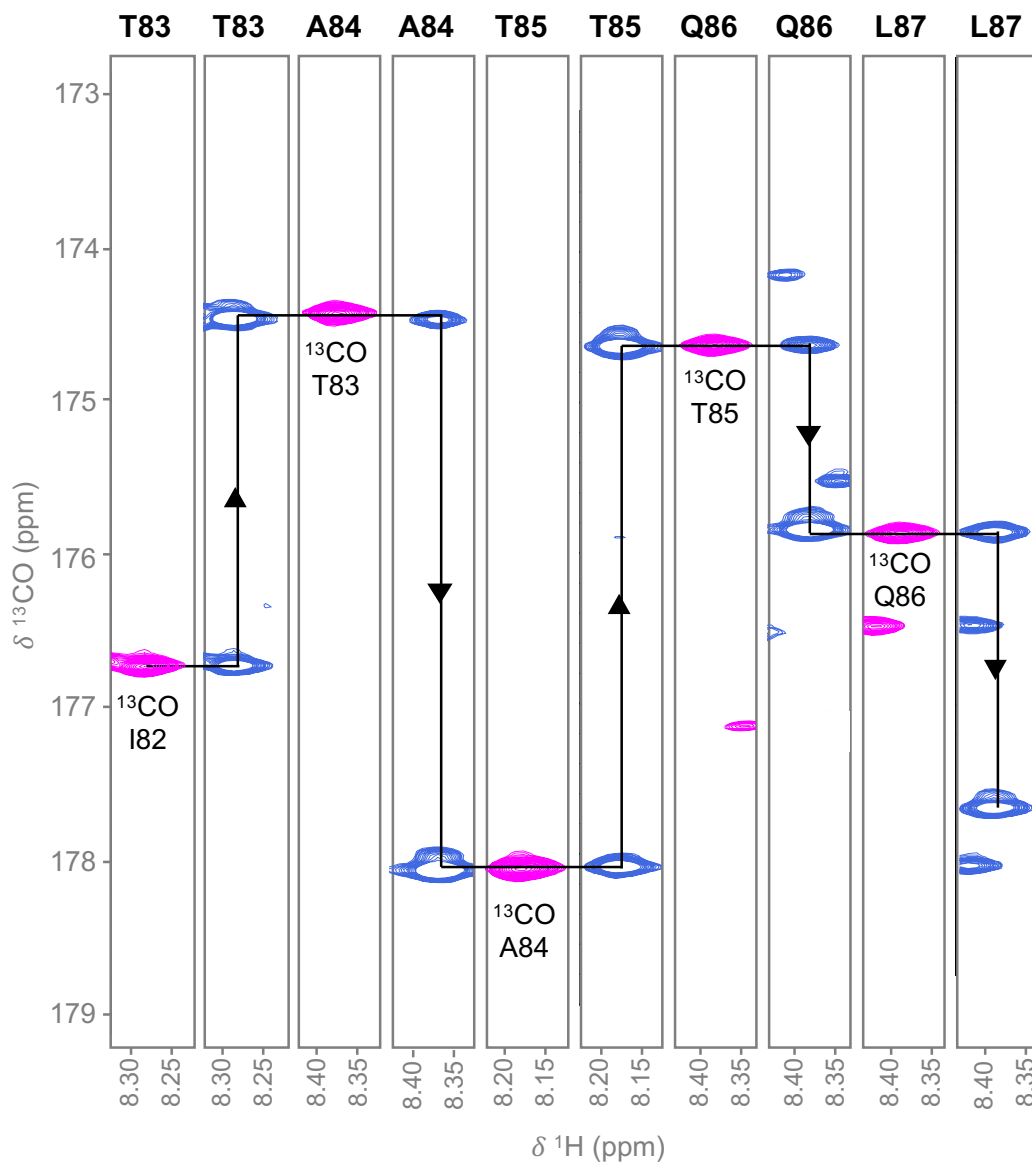


Figure 86: NMR assignment of ^{13}CO nuclei of residues T83-L87 of Cdc25C⁴⁴⁻²²⁸. Comparison of HNCOC (pink) and HN(CA)CO (blue) experiments allowed the sequential assignment of CO resonances as indicated by the arrows.

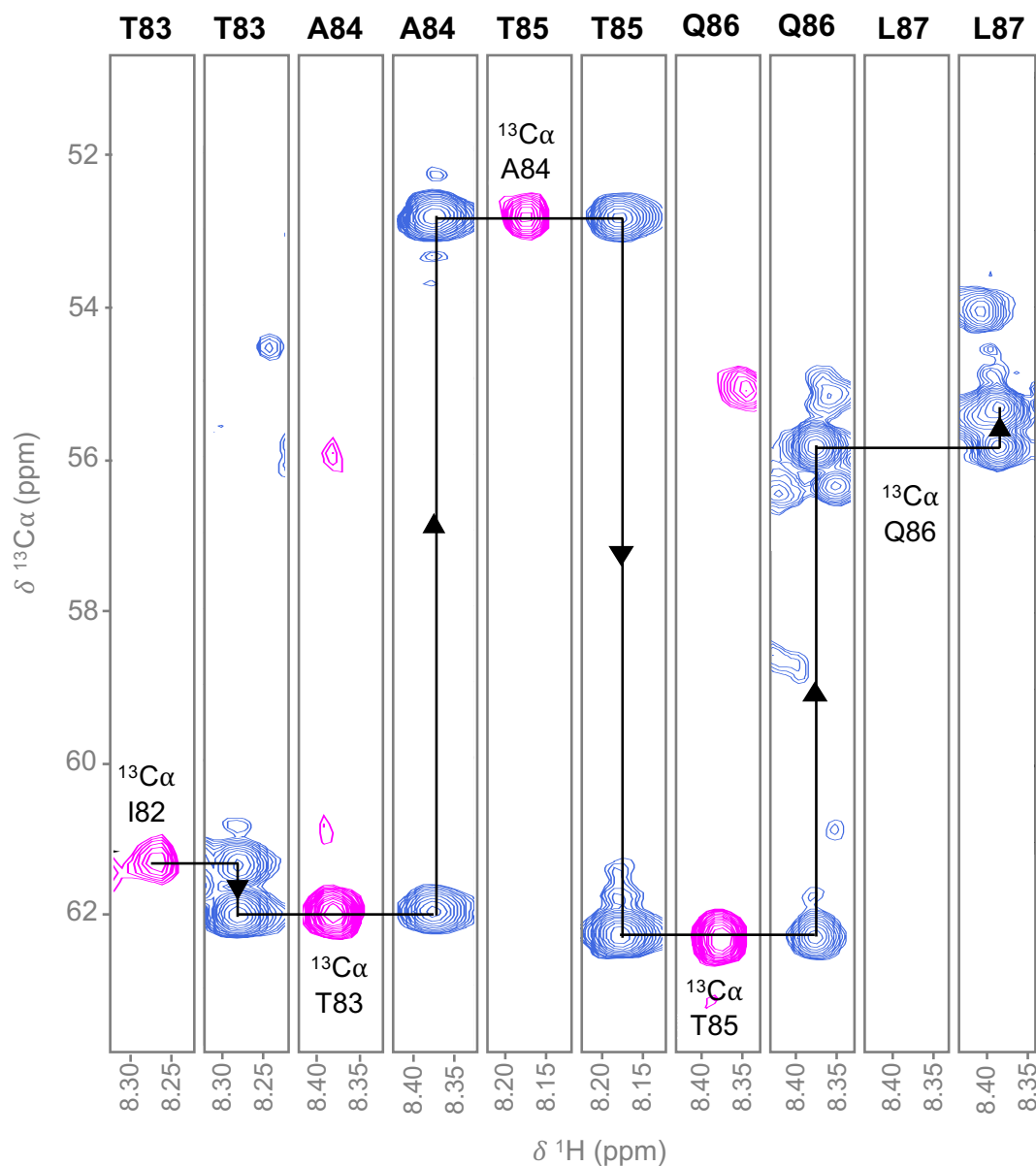


Figure 87: NMR assignment of $^{13}\text{C}\alpha$ nuclei of residues T83-L87 of Cdc25C⁴⁴⁻²²⁸. Comparison of the negative contour of HN(CO)CACB (pink) and HNCA (blue) experiments allowed the sequential assignment of $\text{C}\alpha$ resonances as indicated by the arrows.

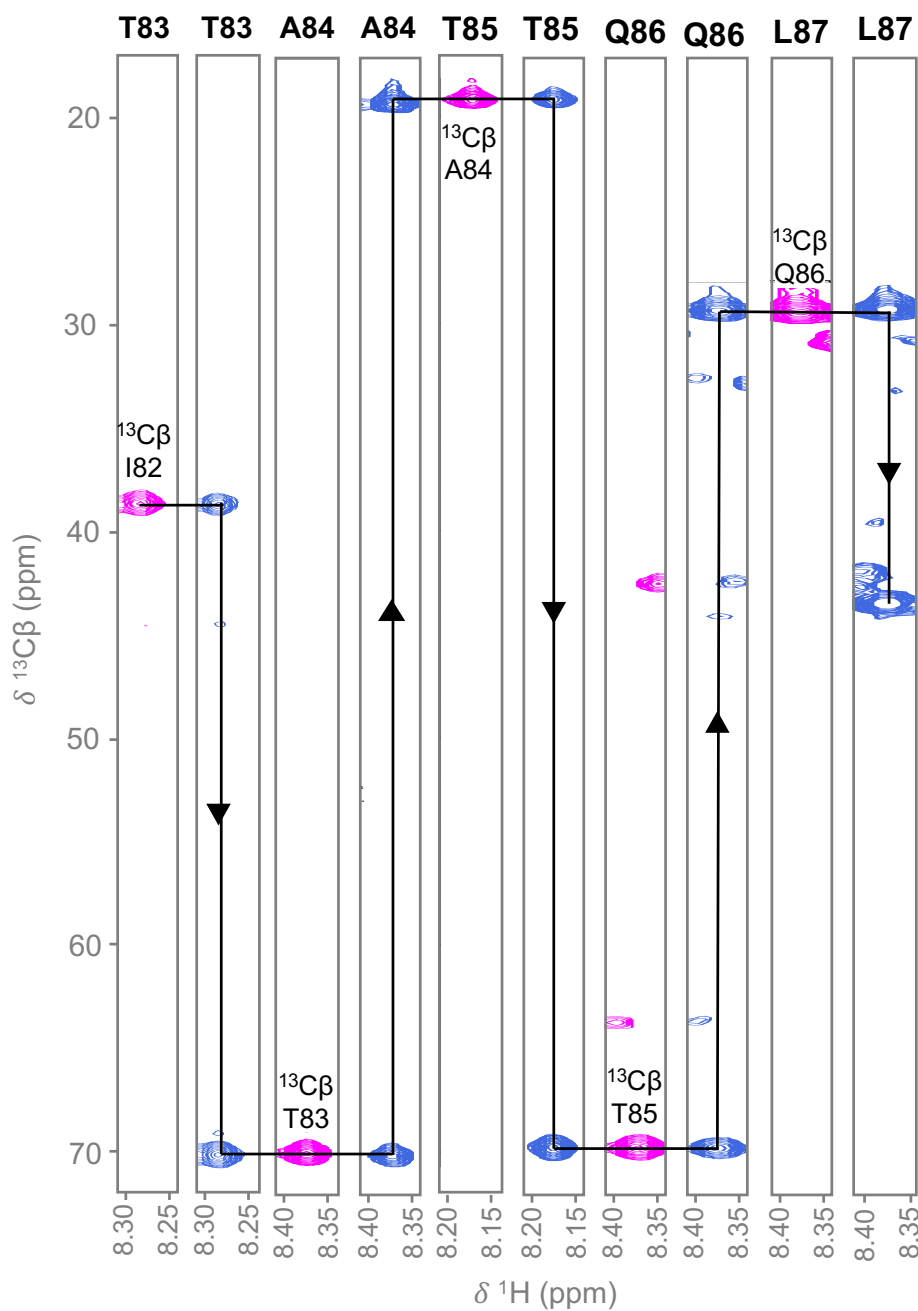


Figure 88: NMR assignment of $^{13}\text{C}\beta$ nuclei of residues T83-L87 of Cdc25C⁴⁴⁻²²⁸. Comparison of the positive contour of HN(CO)CACB (pink) and HNCACB (blue) experiments allowed the sequential assignment of $\text{C}\beta$ resonances as indicated by the arrows.

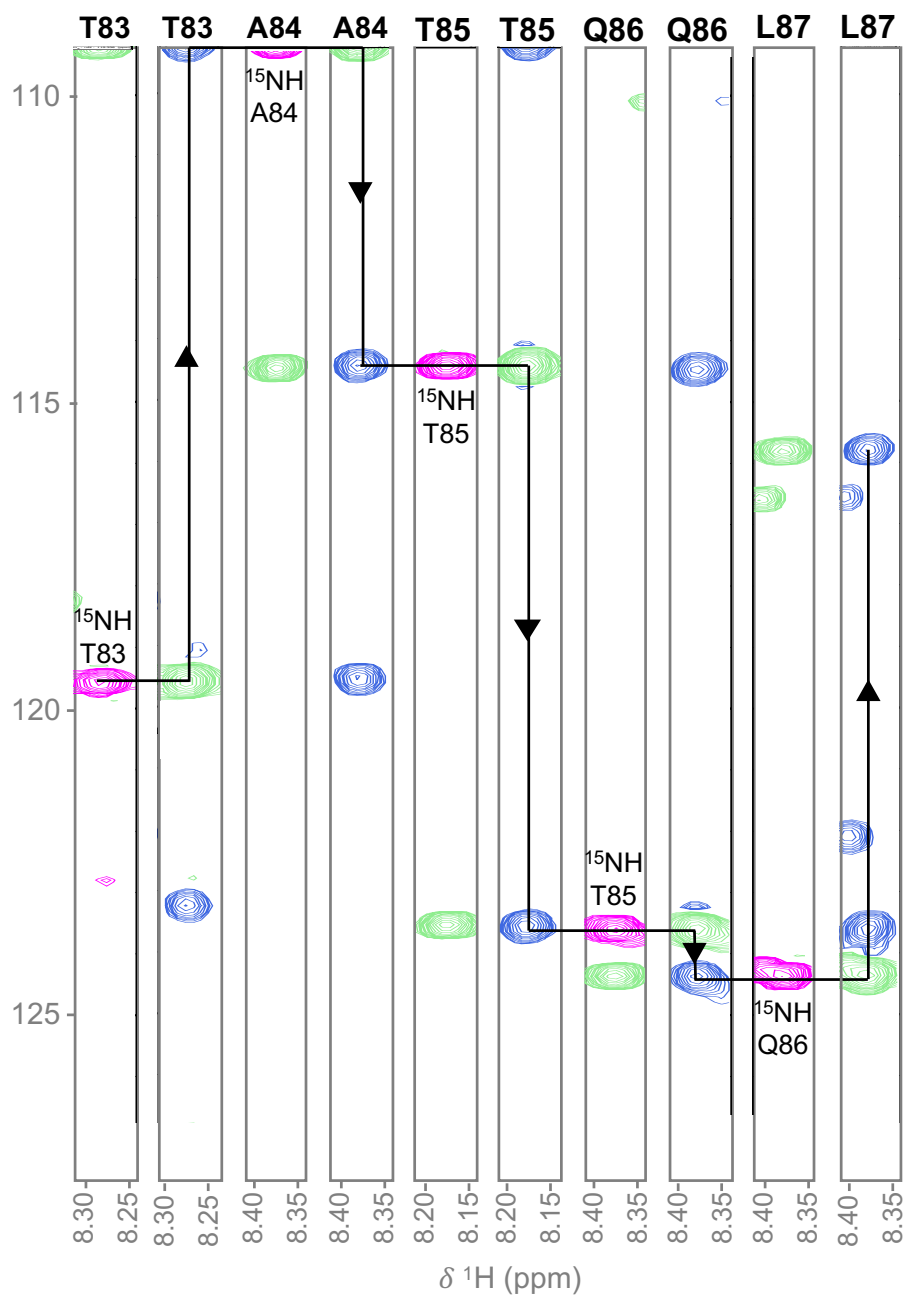


Figure 89: NMR assignment of ^{15}N amide nuclei of residues T83-L87 of Cdc25C⁴⁴⁻²²⁸. Comparison of HN(COCA)NH (pink) and HN(CA)NH (blue) experiments allowed confirmation of the assignment with sequential amide nitrogen chemical shift values (indicated by arrows). Negative contours are displayed in green.

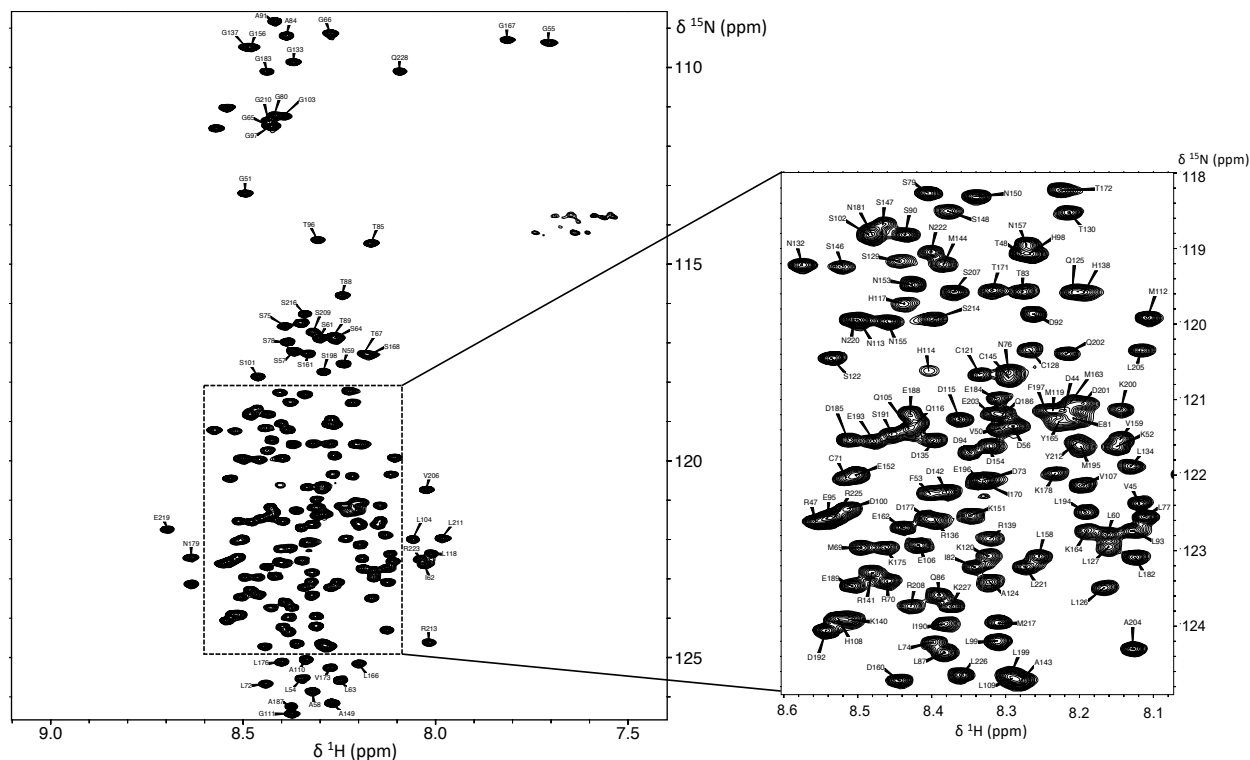


Figure 90: Assigned of ^1H - ^{15}N BEST-TROSY spectrum of Cdc25C^{44-228} at pH 6.5, 283 K recorded on a 950 MHz spectrometer. Unlabelled resonances correspond to remaining residues of the PreScission cleavage site after GST-tag removal.

Since interaction studies with Pin1 will be performed under different conditions (pH 7, 298K), temperature and pH titrations were performed on ^{15}N -labelled samples of Cdc25C^{44-228} . First, a series of ^1H - ^{15}N -BEST-TROSY experiments were recorded at 283 K, 286.5 K, 290 K and 293 K using the 950 MHz spectrometer of the Infranalytics network (see figure 91). Due to time constraints, the final titration point, at 298 K, was recorded using the 500 MHz spectrometer available in the laboratory. Tracking temperature-induced CSPs allowed the assignment of the ^1H - ^{15}N -HSQC spectrum of Cdc25C^{44-228} at pH 6.5, 298K. While most resonances were easily tracked throughout the temperature titration, those in crowded regions of the spectra could not be confidently assigned at 298 K. Additionally, the increase in temperature resulted in the line broadening of several peaks, likely due to faster exchange of amide protons with water. This left the following residues unassigned in spectra of Cdc25C^{44-228} at pH 6.5, 298K recorded at 500 MHz: D44, D50, D73, T83, Q105, Q116, H117, M119, C121, S129, H138, S146, S147, S148, M163, Y165, I170, T171, Q186, S191, E196, F197,

D201, E203 and R208. Hence the assignment of Cdc25C⁴⁴⁻²²⁸ at pH 6.5, 298K represents an 85 % sequence coverage of non-proline residues.

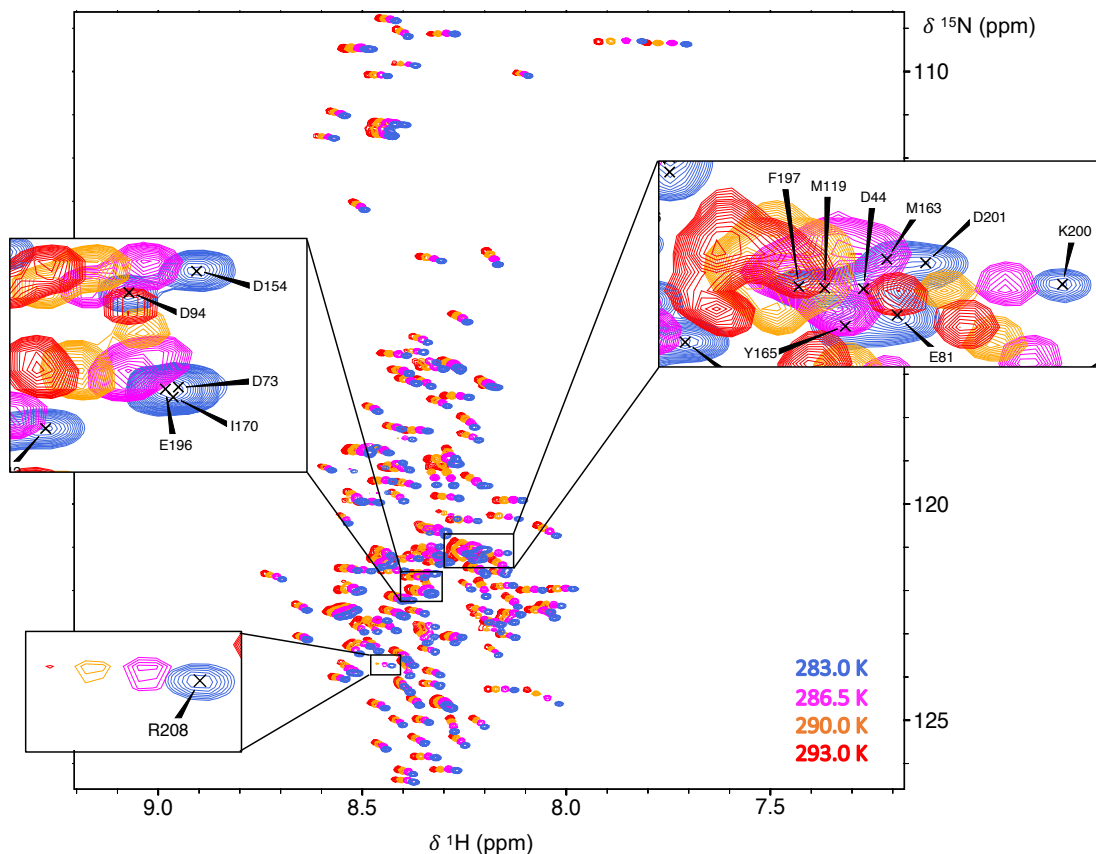


Figure 91: Overlay of ^1H - ^{15}N BEST-TROSY spectra of WT Cdc25C⁴⁴⁻²²⁸ recorded throughout a temperature titration at pH 6.5, recorded using a 950 MHz spectrometer equipped with cryoprobe. All peaks displayed temperature-induced chemical shift perturbations. The movement of most peaks is linear and easily traceable. Some peaks (e.g., R208) exhibit line-broadening upon increase in temperature. For clustered peaks (e.g., the E196, I170 and D73 cluster), it was not possible to confidently track their movement in response to temperature change.

Subsequently, a pH titration was performed on a 100 μM sample of Cdc25C⁴⁴⁻²²⁸ at 298 K by recording ^1H - ^{15}N SOFAST-HMQC experiments at pH 6.5, 6.75 and 7.0 using a 500 MHz spectrometer (see figure 92). While most resonances did not shift upon altering the pH, the spectrum recorded at pH 7 resulted in the extensive line broadening of 14 peaks: G66, N76, S79, S90, S101, D115, S122, G133, A143, A149, N150, T172, L199, S216. This was likely due to increased amide water exchange at pH 7. Overall, the assignment of 136 out of 174 non-proline amide resonances of Cdc25C⁴⁴⁻²²⁸ was possible at pH 7, 298 K, using the combined temperature and pH titrations. This represents an overall sequence coverage of 78 %.

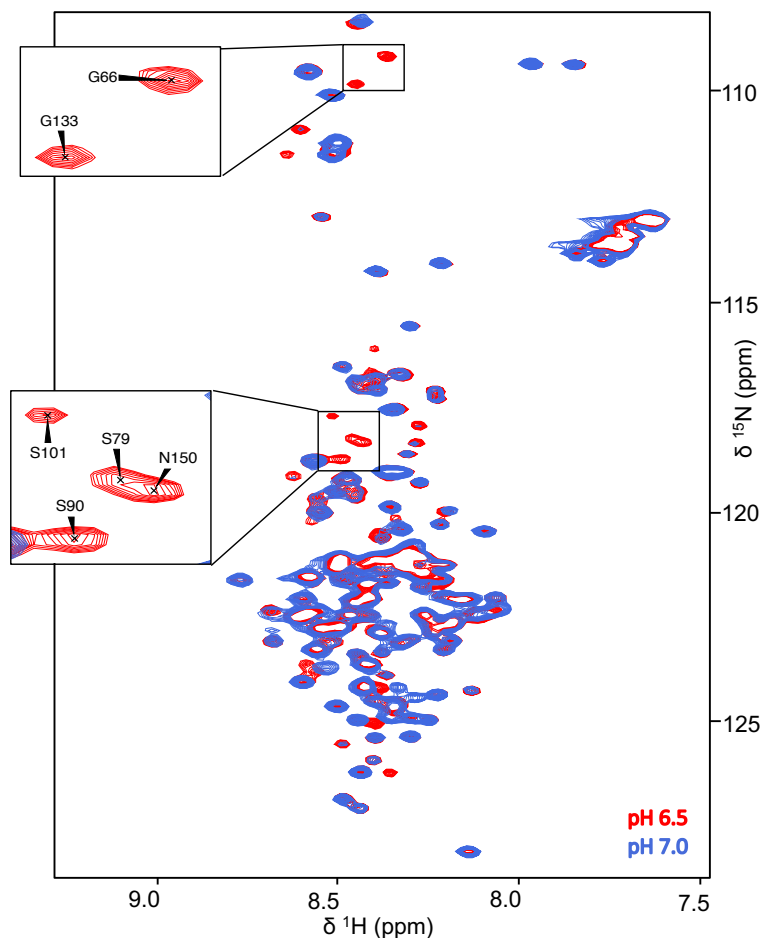


Figure 92: Overlay of ^1H - ^{15}N SO-FAST spectra of WT Cdc25C⁴⁴⁻²²⁸ recorded at titration at pH 6.5 (red) and pH 7 (blue) using a 500 MHz spectrometer equipped with cryoprobe. Most peaks did not exhibit any change upon pH change. Some peaks however, (e.g., G66, S79, S90, S101, G133 and N150) exhibit extensive line-broadening upon increasing the pH, leaving these peaks unassigned in spectra taken at pH 7.

8.2.2. Structure propensity and dynamics of unphosphorylated WT

Cdc25C⁴⁴⁻²²⁸

In order to characterise the secondary structure and dynamics of the Cdc25C⁴⁴⁻²²⁸ protein, several analyses were conducted. As a primary investigation, $\Delta\text{C}\alpha\text{-C}\beta$ values and Secondary Structure Propensity (SSP) scores could be derived from the backbone C α and C β chemical shifts of Cdc25C⁴⁴⁻²²⁸ at pH 6.5, 283 K.

Secondary chemical shift values ($\Delta\delta$) can be calculated by comparing observed C α and C β chemical shift values to the corresponding values expected in a random coil structure ($\Delta\delta =$

$\delta_{\text{obs}} - \delta_{\text{coil}}$) (Wishart & Sykes, 1994). Their calculation relies on the use of a precisely referenced random coil standard. The secondary chemical shift of C α nuclei is strongly associated with the secondary structure they are part of. To report secondary chemical shifts effectively, calculating $\Delta\text{C}\alpha\text{-C}\beta$ values for each resonance is a common approach which helps to mitigate referencing errors (Marsh *et al*, 2006). For the Cdc25C⁴⁴⁻²²⁸ sequence, $\Delta\text{C}\alpha\text{-C}\beta$ values were calculated for each residue, whereby positive values indicate the propensity to form α -helical structures and negative values suggest a higher likelihood of β -strands or extended structures (figure 93, upper panel). In particular, consecutive residues displaying values higher than +2 or lower than -2 indicate the presence of stable α -helices or β -strands respectively, while less positive/negative values indicate the propensity to form transient structural motifs. The calculated $\Delta\text{C}\alpha\text{-C}\beta$ values for Cdc25C⁴⁴⁻²²⁸ confirm its disordered nature, particularly evident due to the lack of regions displaying consecutive positive or negative values. However, it can be noted that a region spanning residues 195-206 (¹⁹⁵MEFSLKDQEALV²⁰⁶) display successive positive values. While these values do not exceed +1.3, and hence do not indicate the presence of a stable α -helix, it could represent a short region with helical propensity.

As a more resilient measure of the secondary structure propensity in IDPs, the SSP approach developed by the Forman-Kay's lab was used (Marsh *et al*, 2006). For each residue in a protein, an SSP score is determined by comparing the observed chemical shift values to the average chemical shift values expected for fully formed secondary structures (α -helical or β -strands), considering their standard deviations. The scores are then weighted based on their reliability, and an average is taken over a few neighbouring residues to minimise the influence of unreliable data. High positive SSP scores (+1) indicate the formation of stable α -helices, while very negative values (-1) suggest the presence of fully-formed β -strands. In the case of IDPs, a score of 0.5 for a given position would indicate that about 50% of conformers in the disordered state ensemble are helical at that position. The calculation of SSP scores for each residue of Cdc25C⁴⁴⁻²²⁸ is represented in figure 93 (middle panel). The SSP values of Cdc25C⁴⁴⁻²²⁸ are low, not exceeding ± 0.5 , confirming the overall disordered nature of the protein. However, there does appear to be a higher propensity to form α -helices than β -strands, indicated by the majority of SSP values being positive. More specifically, the regions displaying the highest α -helical propensity span residues 56-65 (⁵⁶DSANLSILSG⁶⁵), 73-80 (⁷³DLSNLSSG⁸⁰) and 192-208 (¹⁹²DELMEFSLKDQEALVSR²⁰⁸). Significantly, all of these sequences feature *N*-terminal aspartate residues, aligning with prior research that has indicated

that the presence of this residue at the *N*-terminus increases the propensity to promote α -helix formation (Forood *et al*, 1992). Note, none of these sequences contain S/T-P motifs. The region encompassing residues 192-208 exhibits the longest and most populated α -helical conformation, estimated to be populated in 29% of conformational states based on SSP values. Remarkably, it also corresponds to the region displaying successive positive $\Delta C\alpha$ - $C\beta$ values. This region is capped by a *C*-terminal arginine residue, which, in conjunction with the *N*-terminal aspartate residue, aligns with the Presta and Rose helix hypothesis (Presta & Rose, 1988). This hypothesis postulates that the sidechains of *N*-terminal residues serve as hydrogen bond acceptors for residues flanking the *C*-terminus of the helix.

As a preliminary NMR investigation into the dynamics of Cdc25C⁴⁴⁻²²⁸, a ¹H-(¹⁵N) heteronuclear Overhauser effect (hNOE) experiment was performed at 500 MHz. Heteronuclear NOE experiments provide information on the dynamics of a protein over the ps-ns timescale by measuring the relative magnetisation transfer between two nuclei, typically from the proton to a heteronucleus, under conditions where the nuclei are in thermal equilibrium. The hNOE experiment relies on comparing the intensities of two resonance spectra: one obtained with selective saturation of proton magnetisation (the ‘NOE’ peak) and the other without selective saturation (the ‘reference’ peak). In the case of a folded protein, regions that are well-ordered and rigid exhibit higher NOE values, as the motions of the nuclei in these regions are restricted. Conversely, regions that are flexible and disordered exhibit lower NOE values, as the motions of the nuclei in these regions are more rapid and unrestricted. Importantly, hNOE values highly depend on the spectrometer magnetic field and show higher heterogeneity at lower magnetic fields, which makes it easier to distinguish flexible regions from well-ordered ones at 500 MHz. The measured ¹H-(¹⁵N) hNOE values of Cdc25C⁴⁴⁻²²⁸ obtained at 500 MHz were predominantly negative, providing additional experimental evidence for the overall disorder of the protein (see figure 93, lower panel). However, the region corresponding to residues 195-205 displayed slightly positive values, indicating slightly restricted mobility in the ps-ns timescale. Interestingly, this region corresponds to the sequence motif ¹⁹²DELMEFSLKDQEALVSR²⁰⁸ which SSP scores suggest helical propensity. While the SSP approach also predicted potential helical propensities for the *N*-terminal regions 56-65 and 73-80, these regions exhibit weak and negative hNOE indicative of high flexibility.

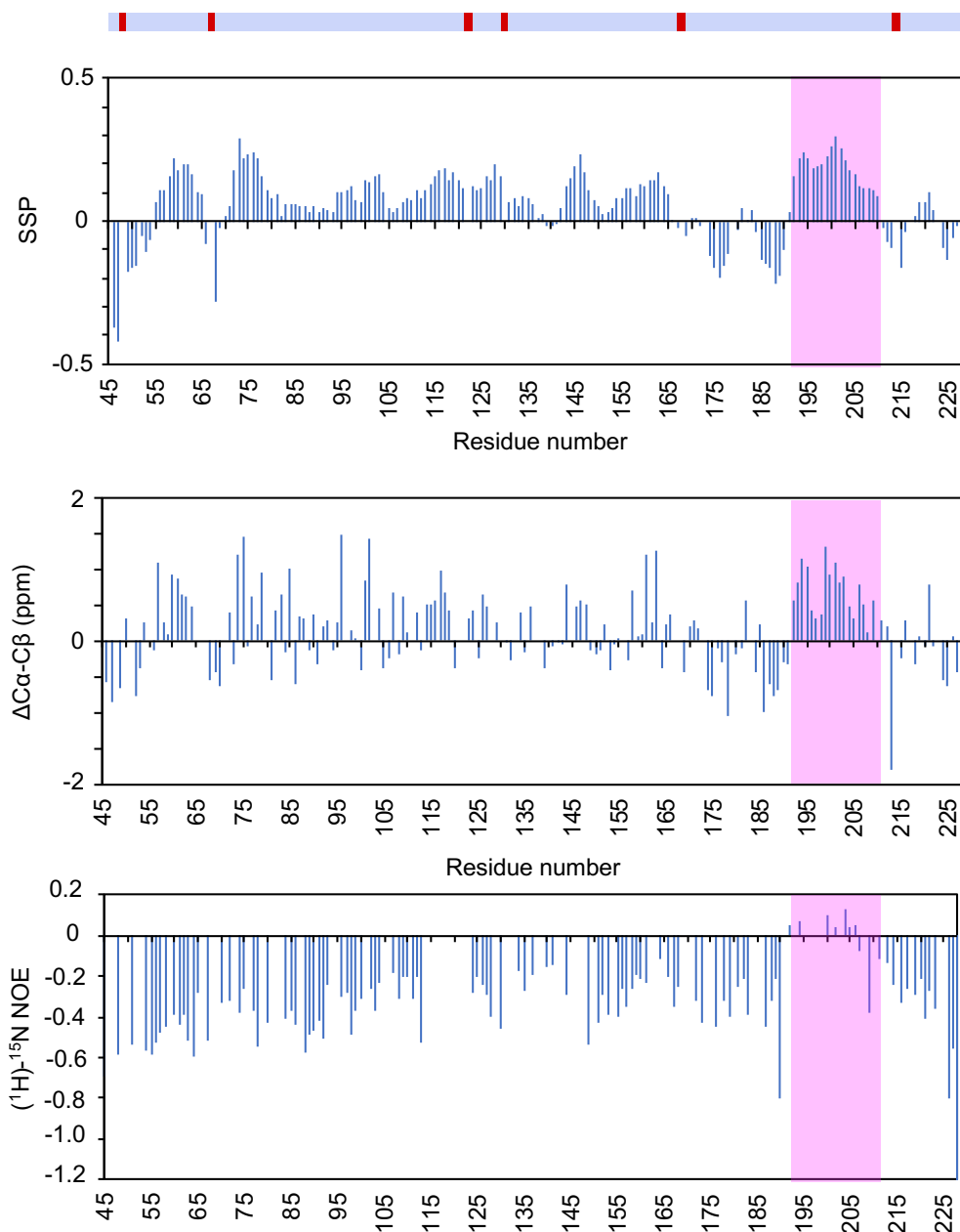


Figure 93: Secondary chemical shift ($\Delta C\alpha-C\beta$) predictions (upper panel), SSP scores (middle panel) and $(^1H)-^{15}N$ NOE values measured at 500 MHz for WT CDC25C⁴⁴⁻²²⁸. The bar above the charts represents the sequence of CDC25C⁴⁴⁻²²⁸ with the six S/T-P sites highlighted in red. The pink box highlights a region of the sequence which seemingly displays a marked structural propensity. Because of the steric specificities of prolines, there is no SSP prediction nor secondary chemical shift for residues directly before prolines. In addition, there are no $\Delta C\alpha-C\beta$ predictions for glycine residues (due to the lack of a C β atom) or cysteine residues (as the chemical shift databases of the program do not take into account the oxydation state, which largely impacts their C β chemical shift values). Missing values $(^1H)-^{15}N$ NOE values correspond to unassigned residues in crowded regions of the spectra at 500 MHz.

8.2.3. The interaction between Pin1 and unphosphorylated WT Cdc25C⁴⁴⁻²²⁸

As a preliminary step to characterise the role of Cdc25C⁴⁴⁻²²⁸ phosphorylation in the interaction with Pin1, we first probed the interaction of the unphosphorylated WT protein. For this purpose, chemical shift perturbation (CSP) experiments were performed by recording ¹H-¹⁵N HSQC spectra on a ¹⁵N-labelled sample of 150 μM Pin1 in the presence of up to 2 molar equivalents of unlabelled Cdc25C⁴⁴⁻²²⁸ (figure 94). The titration revealed two noteworthy observations: first, only weak CSPs were observed for specific amide cross-peaks, and second, a considerable reduction in peak intensity was observed for numerous resonances upon the addition of more than 1 molar equivalent of Cdc25C⁴⁴⁻²²⁸. These observations suggest a binding process in the fast to intermediate exchange regime on the chemical shift timescale.

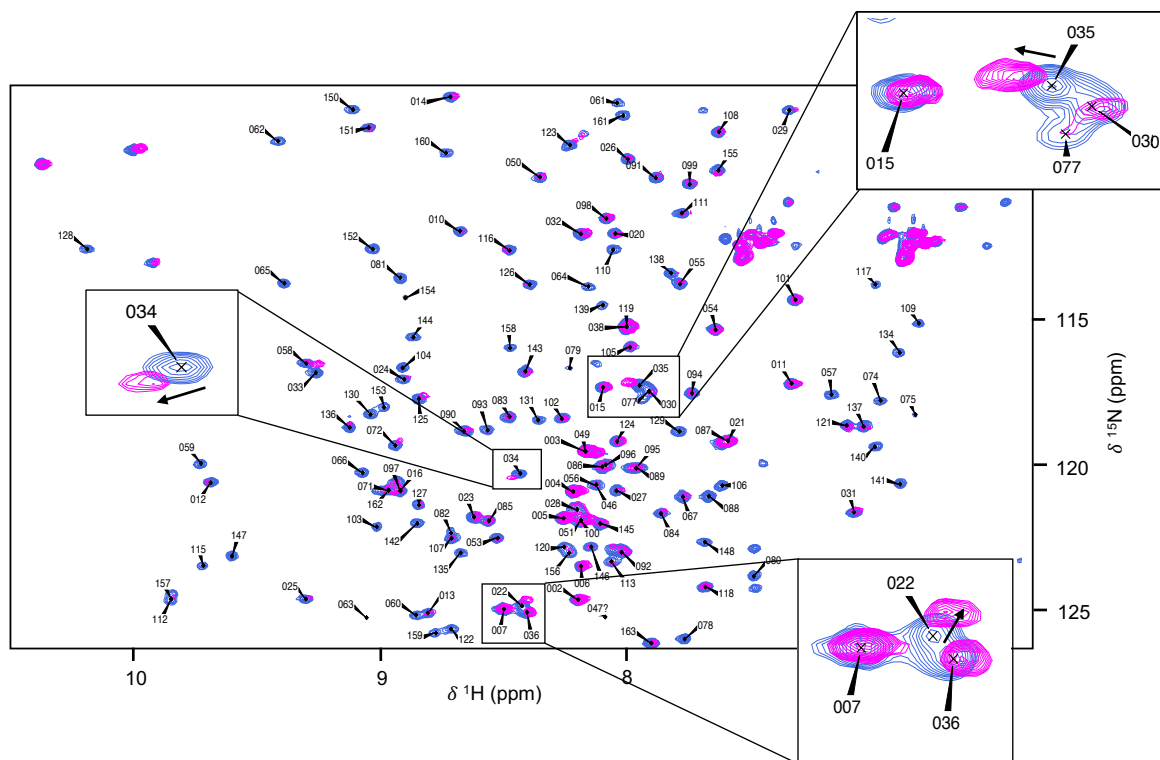


Figure 94: Overlay of ¹H-¹⁵N HSQC spectra of apo-Pin1 (blue) and Pin1 in the presence of 2 molar equivalents of WT Cdc25C⁴⁴⁻²²⁸ (magenta). Several weak perturbations can be observed upon binding of Pin1 to Cdc25C⁴⁴⁻²²⁸ (highlighted by arrows in the zooms of the spectra). Notably, most peaks also show a significant decrease in intensity upon binding to 2 molar equivalents of Cdc25C⁴⁴⁻²²⁸. Spectra were recorded on a 500 MHz spectrometer.

As shown in figure 11, weak CSPs were observed in both the WW and PPIase domains of Pin1 upon binding to 2 molar equivalents of Cdc25C⁴⁴⁻²²⁸ (see figure 95). The residues displaying the highest CSPs in the WW domain corresponded to residues Val22, Gln33, Trp34, and Glu35 of the substrate binding site. The highest CSPs in the PPIase domain involved residues of the hydrophobic proline binding pocket (Gly123, Phe125, Gly128, Gln129, Met130, and Glu136). Additional CSPs were observed in the PPIase domain for residues of the catalytic loop (Ser72, Lys82), residue Leu141 of the interdomain interface, and Gly155 which neighbours the catalytic site. Notably, significant perturbations were not observed for residues of the phosphate binding sites of both domains, as could be expected upon binding of the unphosphorylated WT Cdc25C⁴⁴⁻²²⁸ protein to Pin1. As an exception, Val22, the residue neighbouring Tyr23 of the WW domain, which is involved in phosphate-group coordination, displays a subtle CSP. However, the absence of further perturbations in the residues of loop I (residues 17-20) within the WW domain does not support a specific interaction with the phosphate binding site. The severe line broadening of amide resonances after the addition of 1-2 molar equivalents of Cdc25C⁴⁴⁻²²⁸ prevented the estimation of K_D values from chemical shifts as CSP curves did not reach saturation under these conditions. Nonetheless, the linear trend of CSP curves from 0 to 2 molar equivalents of Cdc25C⁴⁴⁻²²⁸ suggested a very low affinity, with a K_D value in the millimolar range.

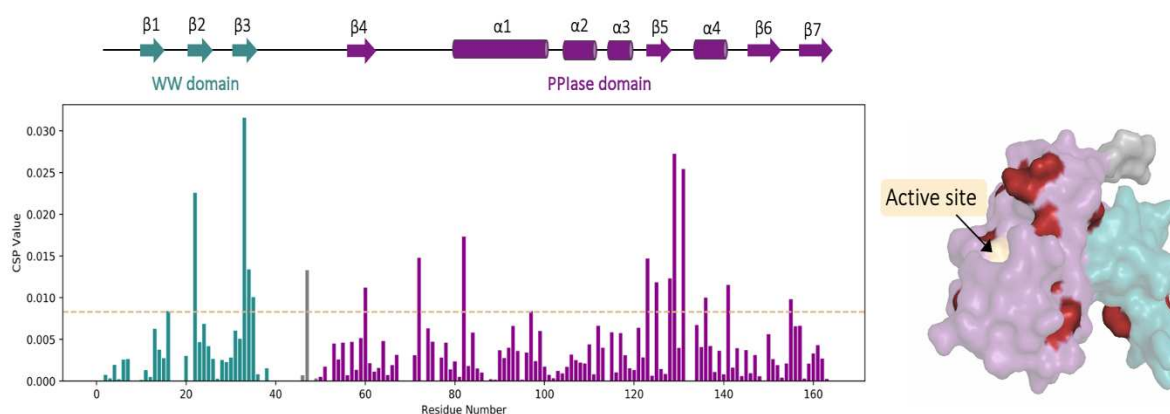


Figure 95: Analysis of amide CSPs observed in Pin1 ¹H-¹⁵N HSQC spectra upon binding to 2 molar equivalents of Cdc25C⁴⁴⁻²²⁸. The CSP chart (left) reveals residues in the WW and PPIase domain which show (weak) CSP values above the threshold CSP value ($m + \sigma$) represented by the orange dashed line. These residues are mapped onto the structure of Pin1 (right) in red.

To quantify the observed signal loss following the addition of Cdc25C⁴⁴⁻²²⁸, the relative intensities of each amide cross-peak (I) were compared to their reference peaks from apo-Pin1

(I_{\max}) throughout the titration (figure 96). The analysis of signal loss following the addition of 1 and 2 molar equivalents of Cdc25C⁴⁴⁻²²⁸ indicated a general decrease in intensity across all peaks, and in particular, the lack of a specific region exhibiting distinct behaviour. The overall line broadening of amide resonances may be explained by the formation of a complex between Pin1 and Cdc25C⁴⁴⁻²²⁸ that would be too large to be effectively detected by NMR without protein deuteration. Indeed, the binding of the WW and PPIase domains to two distinct Cdc25C⁴⁴⁻²²⁸ molecules (stoichiometry 1:2) would result in the formation of a 60 kDa complex. Furthermore, since the Cdc25C⁴⁴⁻²²⁸ protein contains 6 potential Pin1 binding sites, the binding of these regions to distinct Pin1 molecules (when Pin1 is in excess) would lead to the formation of even larger complexes. The existence of internal dynamics within the complex is also likely to explain the overall line broadening of Pin1 resonances, in particular if the WW and PPIase domains do not recognise specific Cdc25C binding sites and the exchange between them is in the μ s-ms timescale.

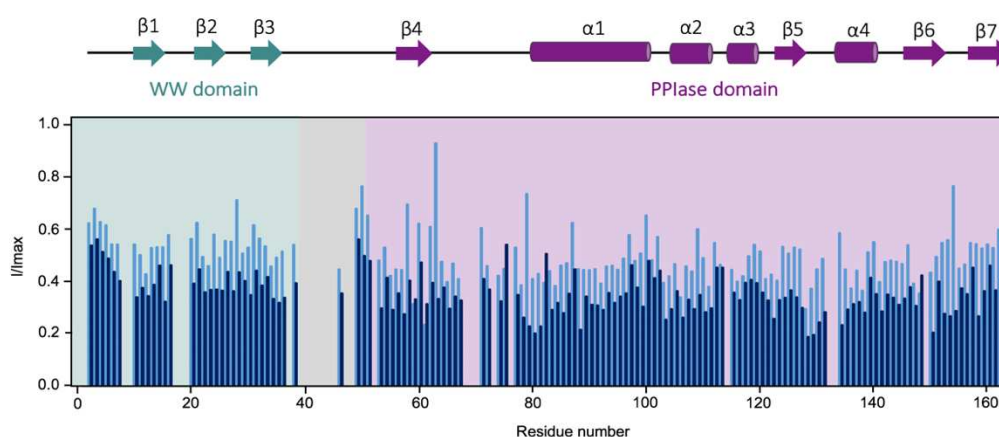


Figure 96: Analysis of the relative signal intensity of amide cross-peaks in ¹H-¹⁵N HSQC spectra recorded on apo-Pin1 and in the presence of 1 (pale blue) or 2 (dark blue) molar equivalents of Cdc25C⁴⁴⁻²²⁸. In the presence of 1 and 2 equivalents of Cdc25C⁴⁴⁻²²⁸, all peaks display a similar reduction in intensity. Residues with a value of 0 represent prolines or unassigned residues in the apo-Pin1 spectrum.

Another explanation of the widespread signal loss upon Cdc25C⁴⁴⁻²²⁸ addition may be the dilution of the Pin1 sample throughout the titration. To validate that the observed reduction in signal intensity was independent of any dilution effect on the ¹⁵N-labeled Pin1 sample during the titration, a cross-titration experiment was performed on a 500 MHz spectrometer. This approach maintained a constant concentration of Pin1 throughout the titration. The experiment involved the preparation of two NMR tubes, one containing 50 μ M ¹⁵N-labelled Pin1 in Tris

buffer (pH 7) with 10% D₂O, and the second tube containing an additional 300 μM of unlabelled Cdc25C⁴⁴⁻²²⁸ (equivalent to 6 molar equivalents). By mixing these tubes at different ratios, intermediate concentrations of 0.5, 1.0, 1.5, 2.0, 3.0 and 4.0 molar equivalents of Cdc25C⁴⁴⁻²²⁸ were achieved during the course of the cross-titration. The resulting NMR spectra exhibited a comparable line-broadening phenomenon across all ¹H-¹⁵N cross-peaks after addition of 1 equivalent of Cdc25C, ruling out the hypothesis of signal loss due to sample dilution.

In summary, the observation of signal disappearance and weak CSPs in both the WW and PPIase domains suggests that the two Pin1 domains have the ability to interact with the unphosphorylated form of WT Cdc25C⁴⁴⁻²²⁸, albeit with a very low affinity. To gain more structural insight into the interaction between Pin1 and Cdc25C, it became imperative to generate phosphorylated forms of our Cdc25C⁴⁴⁻²²⁸ construct, as these modifications are expected to significantly increase the binding affinity and are likely to shed light on the intricacies of the binding interface that might have been overlooked in this preliminary assessment.

8.3 NMR investigation of the Cdc25C⁴⁴⁻²²⁸ phosphomimetic mutant

The production of phosphorylated forms of the Cdc25C IDR for NMR studies poses several challenges linked to the need to produce high yields of a homogeneous population of specifically phosphorylated protein. As an alternative and simpler approach, a phosphomimetic mutant of hyperphosphorylated Cdc25C⁴⁴⁻²²⁸ (pM-Cdc25C⁴⁴⁻²²⁸) was generated, in which each of the serine/threonine residues within the six S/T-P sites were substituted with glutamate. This phosphomimetic strategy aimed to provide insights into the potential effects of multiple phosphorylations on the interaction with Pin1, offering a preliminary understanding of the interaction before undertaking the complex task of site-specific phosphorylation.

8.3.1. The use of phosphomimetics to study phospho-proteins

Mimicking phosphorylated serine and threonine residues using the natural negatively charged amino acids, aspartate and glutamate, is a well-established method to study the effects of protein phosphorylation (Thorsness & Koshland, 1987). Introducing these amino acids in

place of native serine/threonine residues can be accomplished through novel plasmid design or side-directed mutagenesis (SDM). Their substitution introduces a localised negative charge, which can mimic phosphate-mediated electrostatic effects (figure 97). In fact, a theory has been suggested that Asp and Glu residues could serve as ancestral precursors to phosphorylation sites (Pearlman *et al*, 2011). According to this hypothesis, these residues have, over time, evolved into phosphorylated Ser or Thr residues to allow the reversible control over their functional effects. While both groups have proved to act as successful phosphomimetic substitutions in various instances, from comparing their chemical structures to that of a phosphorylated serine residue, glutamate should represent a better structural mimic due to its longer sidechain.

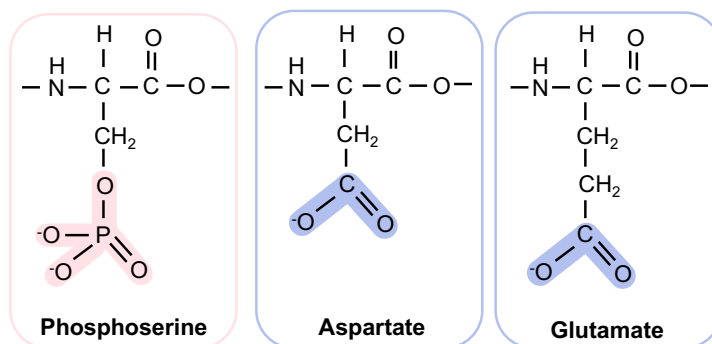


Figure 97: Comparison of the structures of phosphoserine, aspartate and glutamate.

This phosphomimetic approach presents several notable advantages. Primarily, it facilitates sample production significantly compared to traditional phosphorylation methods. Additionally, while achieving homogenous populations through protein phosphorylation is often challenging, phosphomimetic proteins ensure a uniform ‘phosphorylated’ sample. Furthermore, this method grants precise control over the sites that undergo mutation, facilitating the creation of specific phosphomimetic populations which could be hard to produce using kinases or cell extracts. However, the use of phosphomimetics has proved to be somewhat inconsistent in the ability to mimic the effects of Ser/Thr phosphorylation. This discrepancy is likely due to the fact that at physiological pH, phosphorylated residues primarily carry a negative charge of $-2e$ (with pKa values of 5.6 for pSer and 5.9 for pThr), while Asp and Glu residues are singly charged. While this technique does not consistently yield results, it serves as a valuable initial approach for investigating the hyperphosphorylated state of Cdc25C⁴⁴⁻²²⁸ and could allow the more rapid acquisition of data that better reflects what occurs during the interaction with Pin1.

8.3.2. NMR characterisation of pM-Cdc25C⁴⁴⁻²²⁸

To produce a phosphomimetic mutant of hyperphosphorylated Cdc25C⁴⁴⁻²²⁸, a novel plasmid was designed in which the six Ser/Thr-Pro sites were mutated to Glu-Pro. The production and purification protocol for WT Cdc25C⁴⁴⁻²²⁸ was applied to successfully produce unlabelled and uniformly labelled ¹⁵N and ¹³C/¹⁵N samples of pM-Cdc25C⁴⁴⁻²²⁸ (see chapter 6.2.4). Before investigating the interaction of Pin1 with the phosphomimetic mutant, a crucial step involved the assignment of backbone amide resonances in the free form of pM-Cdc25C⁴⁴⁻²²⁸. Indeed, the overlay of ¹H-¹⁵N BEST-TROSY spectra of WT Cdc25C⁴⁴⁻²²⁸ and pM-Cdc25C⁴⁴⁻²²⁸ showed notable chemical shift differences (figure 98), underscoring the need for a backbone assignment, particularly to assign resonances in the regions harbouring the mutated sites.

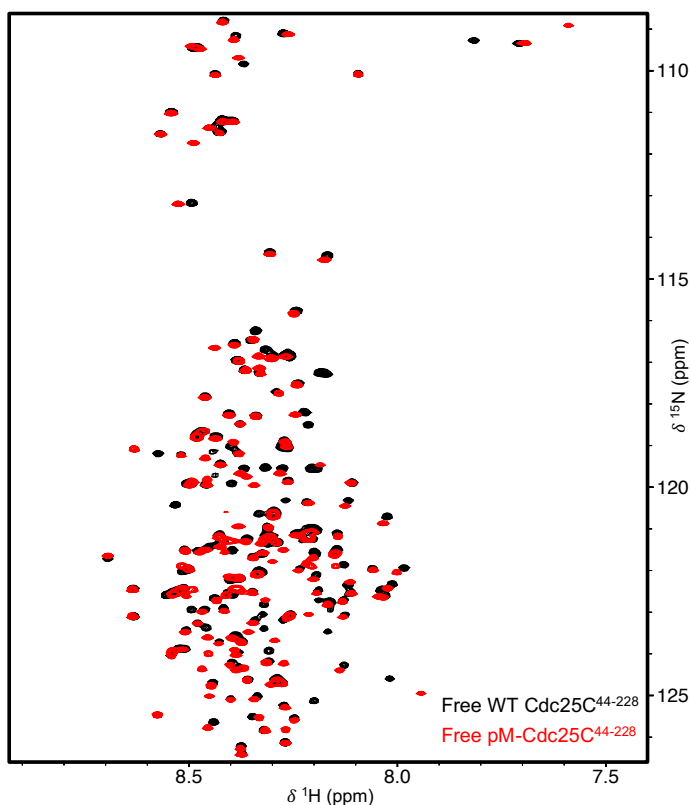


Figure 98: Overlay of ¹H-¹⁵N BEST-TROSY folded spectra of WT (black) and pM-Cdc25C⁴⁴⁻²²⁸ (red). Experiments were recorded at pH 6.5, 283 K using a 950 MHz spectrometer equipped with cryoprobe.

To assign HN, NH, CO, C α , and C β resonances, seven ¹H/¹⁵N/¹³C triple resonance 3D experiments were recorded: HNC(O), HNCA, HNCACB, HN(CA)CO, CBCA(CO)NH,

HN(COCA)N, and HN(CA)N. Experiments were recorded at pH 6.5 and 283 K using a 950 MHz spectrometer equipped with a cryoprobe. The analysis of 3D spectra allowed for the complete assignment of all 174 non-proline amide resonances of the pM-Cdc25C⁴⁴⁻²²⁸ sequence at pH 6.50, 283 K (figure 99a) (annex 1). The assignment confirmed that the chemical shift differences observed between WT Cdc25C⁴⁴⁻²²⁸ and pM-Cdc25C⁴⁴⁻²²⁸ were only concentrated around the mutated (EP) sites (see figure 99b).

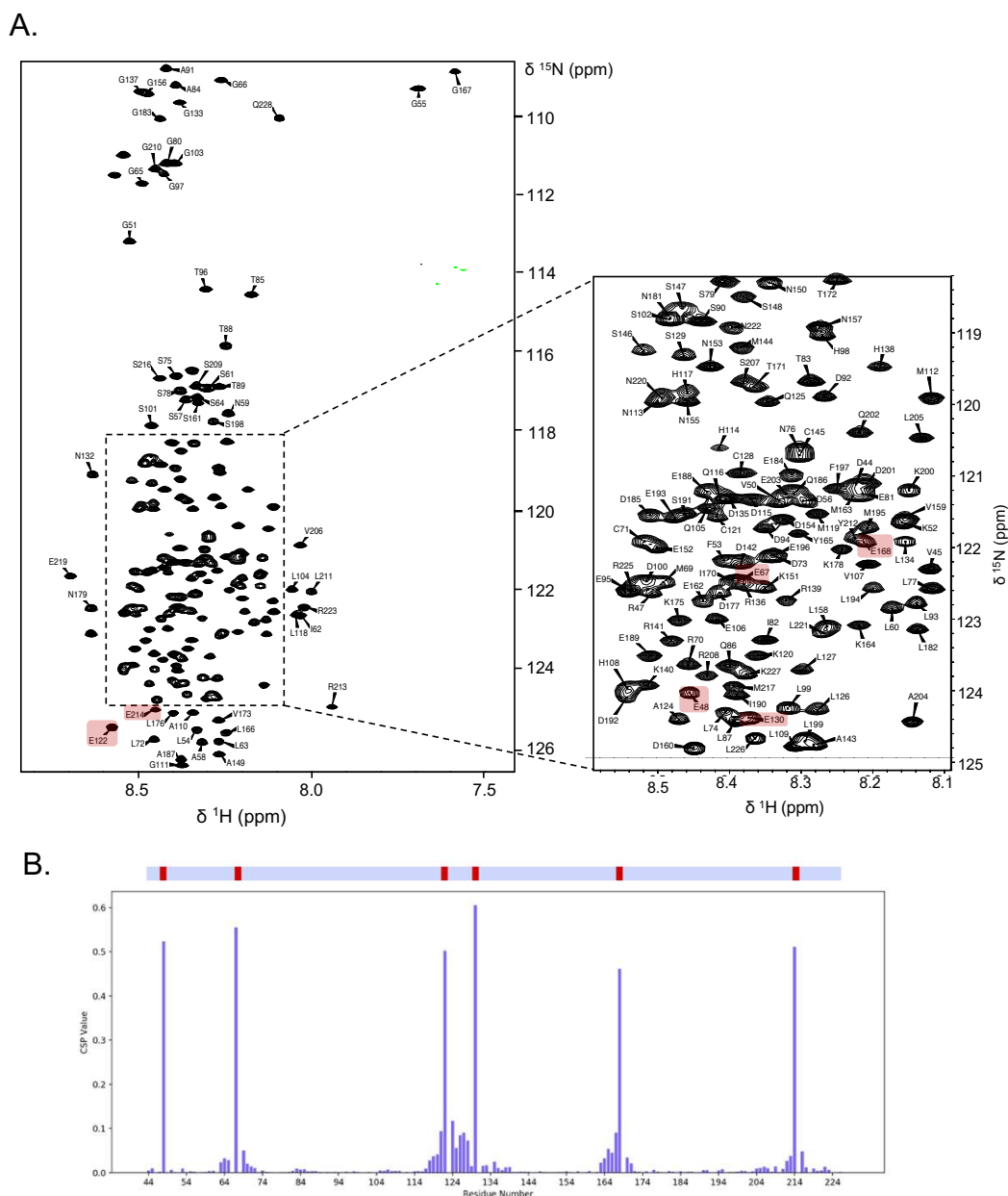


Figure 99: Complete backbone amide assignment of Cdc25C⁴⁴⁻²²⁸-pM at pH 6.5, 283 K. A. Assigned ¹H-¹⁵N BEST-TROSY spectrum of pM-Cdc25C⁴⁴⁻²²⁸ recorded on a 950 MHz spectrometer. Unlabelled resonances correspond to remaining residues of the PreScission cleavage site after GST-tag removal. The six mutated residues (S/T6E) are highlighted in red. **B.** Amide chemical shift differences between WT Cdc25C⁴⁴⁻²²⁸ and pM-Cdc25C⁴⁴⁻²²⁸. The position of the six mutated (EP) sites are highlighted in red in the bar above the chart.

This indicates that the substitution of the 6 Ser/Thr residues with glutamate does not modify the overall structure of Cdc25C⁴⁴⁻²²⁸, which remains disordered, as also supported by the very similar C α chemical shifts between the 2 proteins. As for WT Cdc25C⁴⁴⁻²²⁸, since interaction studies with Pin1 will be performed under different conditions (pH 7, 298K), temperature and pH titrations were performed on ¹⁵N-labelled samples of pM-Cdc25C⁴⁴⁻²²⁸ at 950 MHz, allowing the assignment of 140 amide resonances over the 174 non-proline amide resonances of pM-Cdc25C⁴⁴⁻²²⁸.

8.3.3. The interaction between Pin1 and pM-Cdc25C⁴⁴⁻²²⁸

To investigate whether the phosphomimetic mutant of Cdc25C⁴⁴⁻²²⁸ enhances the interaction with Pin1, NMR titrations were conducted. First, a CSP experiment was performed by recording ¹H-¹⁵N HSQC spectra on a ¹⁵N-labelled sample of 150 μ M Pin1 in the absence and presence of up to 2 molar equivalents of pM-Cdc25C⁴⁴⁻²²⁸ (figure 100).

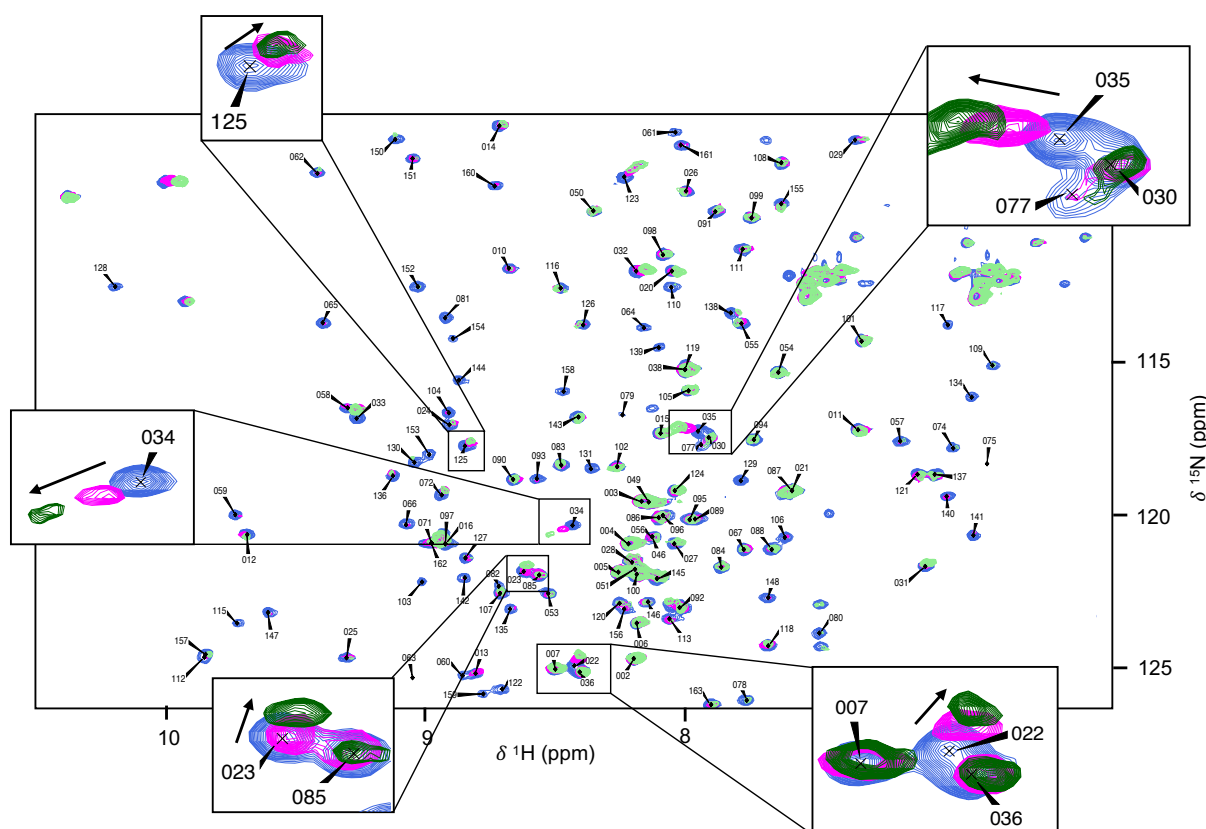


Figure 100: Overlay of ¹H-¹⁵N HSQC spectra of apo-Pin1 (blue) and Pin1 in the presence of 2 molar equivalents of WT (magenta) or phosphomimetic (green) Cdc25C⁴⁴⁻²²⁸. The same Pin1 resonances show CSPs upon binding to phosphomimetic and WT Cdc25C⁴⁴⁻²²⁸ and followed similar trajectories (as indicated by arrows).

Interestingly, the titration of ^{15}N -Pin1 with pM-Cdc25C⁴⁴⁻²²⁸ gave rise to very similar observations to those of the titration with WT Cdc25C⁴⁴⁻²²⁸, with weak CSPs combined with the overall line broadening of Pin1 resonances beyond the addition of 1 molar equivalent of pM-Cdc25C⁴⁴⁻²²⁸. In particular, a comparable pattern of CSPs was observed as seen upon binding to WT Cdc25C⁴⁴⁻²²⁸. The same residues were affected, and their shifts followed similar trajectories, although the CSPs were more pronounced for the phosphomimetic mutant (figure 101).

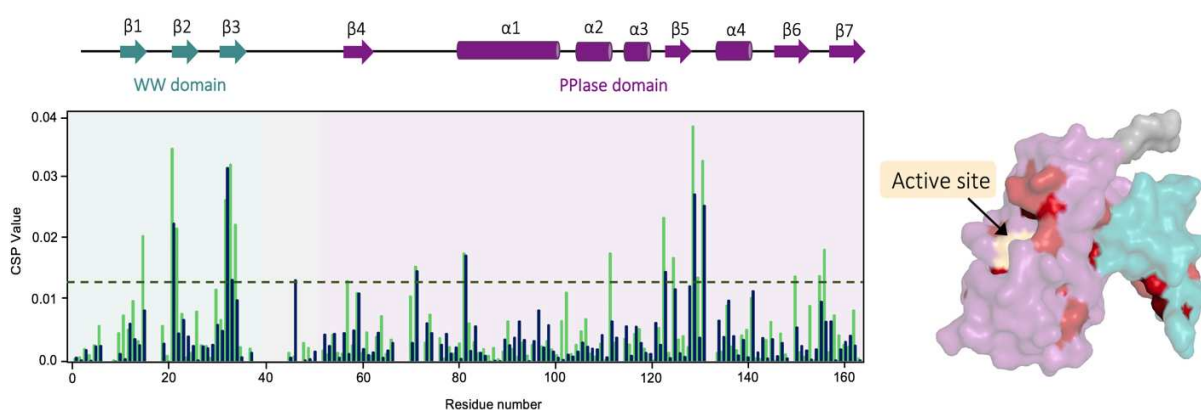


Figure 101: Analysis of amide CSPs observed in Pin1 ^1H - ^{15}N HSQC spectra upon binding to 2 molar equivalents of WT (dark green) or phosphomimetic (light green) Cdc25C⁴⁴⁻²²⁸. The CSP chart (left) reveals residues in the WW and PPIase domain which show weak CSP values upon binding to the WT or phosphomimetic Cdc25C⁴⁴⁻²²⁸. These residues are mapped onto the structure of Pin1 (right) in red. In light red are the residues which display CSPs above a threshold value of $m + \sigma$ upon binding both the WT and phosphomimetic mutant protein. In dark red are residues which only display CSPs over the threshold value upon binding to the phosphomimetic mutant.

The residues displaying the highest CSPs in the WW domain upon binding to the phosphomimetic mutant corresponded to residues Ser16, Val22, Tyr23, Gln33, Trp34, and Glu35 of the substrate binding site. Similarly, the PPIase domain showed notable CSPs, particularly within the hydrophobic proline binding pocket (Gly123, Phe125, Gly128, Gln129, and Met130). Additional perturbations were noted in the PPIase domain for residues of the catalytic loop (Ser72, Lys82), residue Val150 of the interdomain interface and Gly155 which neighbours the catalytic site. Notably, the additional CSPs observed in the phosphate-group binding site of the WW domain (Ser16 and Tyr23) suggest that the glutamates preceding proline residues in pM-Cdc25C⁴⁴⁻²²⁸ likely contribute to binding. However, there is a notable absence of significant CSPs for residues of the phosphate-binding site of the PPIase domain (residues 63-70), indicating limited interaction at this site. It is worth noting that residues Ser72

and Lys82 of the catalytic loop exhibit similar levels of perturbation in both the phosphomimetic mutant and WT Cdc25C⁴⁴⁻²²⁸ binding interactions. This could suggest that these particular residues are involved in interactions that are not heavily influenced by the phosphorylation state of Cdc25C.

Overall, the higher amplitude of amide CSPs induced by the phosphomimetic mutant suggests a higher binding affinity to Pin1 as compared to WT Cdc25C⁴⁴⁻²²⁸. However, the interaction of Pin1 with the phosphomimetic mutant was still not strong enough to allow the extraction of K_D values at the residue level as the CSP curves did not reach saturation. Yet the linear trend of CSPs curves from 0 to 2 molar equivalents of pM-Cdc25C⁴⁴⁻²²⁸ suggested a K_D value in the millimolar range. Achieving binding saturation was again limited by the significant line broadening of Pin1 resonances after the addition of 1 molar equivalent of pM-Cdc25C⁴⁴⁻²²⁸. To quantitatively assess the loss of peak intensity upon binding of the phosphomimetic mutant to Pin1, the relative intensities of each amide cross-peak (I) were compared to their reference peaks from apo-Pin1 (I_{max}) throughout the titration (figure 102).

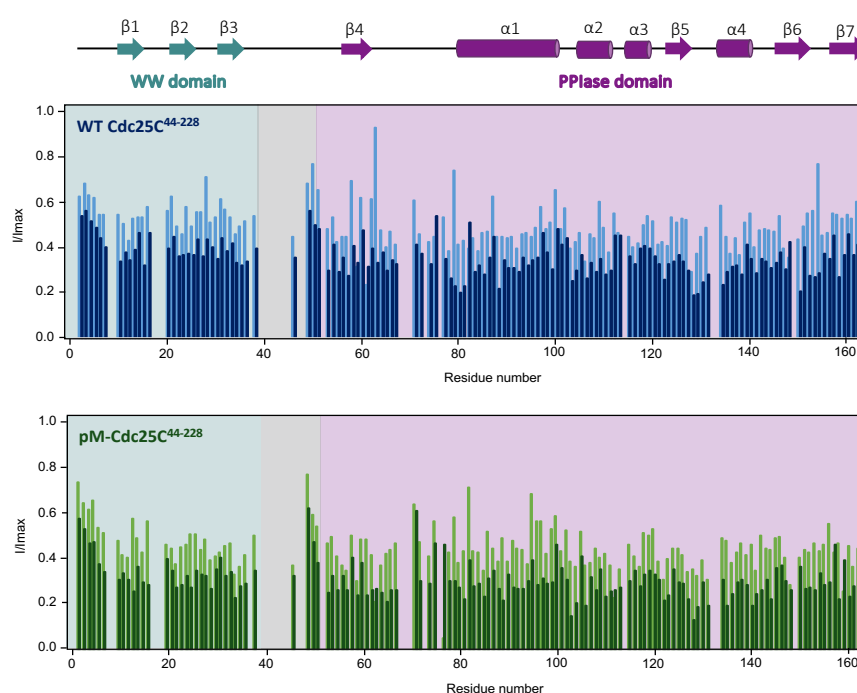


Figure 102: Analysis of the relative signal intensity of amide cross-peaks in ^1H - ^{15}N HSQC spectra recorded on apo-Pin1 and in the presence of 1-2 molar equivalents of WT (upper) or phosphomimetic (lower) Cdc25C⁴⁴⁻²²⁸. The I/I_{max} value is represented in pale blue/green for Pin1 in the presence of 1 molar equivalent of Cdc25C and dark blue/green for Pin1 in the presence of 2 equivalents of Cdc25C⁴⁴⁻²²⁸. All peaks display a similar reduction in intensity. Residues with a value of 0 represent prolines or unassigned residues in the apo-Pin1 spectrum.

Similar to the observations made for the binding of the WT protein, all amide cross-peaks in the ^1H - ^{15}N HSQC spectra of Pin1 exhibited a severe reduction in intensity upon binding to 1 and 2 molar equivalents of the pM-Cdc25C⁴⁴⁻²²⁸, likely due to the formation of a large complex or the existence of μs - ms dynamics within the complex. Taken together, this data suggests that, although Pin1 binds the phosphomimetic mutant with a slightly higher affinity than WT Cdc25C⁴⁴⁻²²⁸, its binding mode to the two Cdc25C constructs may be very similar.

To identify which sites of pM-Cdc25C⁴⁴⁻²²⁸ were involved in the binding to Pin1, a reverse titration was performed. The titration of up to 4 molar equivalents of unlabelled Pin1 into a ^{15}N -labelled sample of pM-Cdc25C⁴⁴⁻²²⁸ was monitored by a series of ^1H - ^{15}N BEST-TROSY spectra recorded on a 950 MHz spectrometer. Analysis of the resulting spectra revealed that while the majority of pM-Cdc25C⁴⁴⁻²²⁸ amide peaks were unaffected by the presence of Pin1, a few residues were shown to exhibit significant CSPs (figure 103). Furthermore, the addition of 1 to 4 molar equivalents of unlabelled Pin1 did not lead to severe line broadening of pM-Cdc25C⁴⁴⁻²²⁸ amide resonances, in contrast to what was observed when titrating ^{15}N -Pin1 with unlabelled pM-Cdc25C⁴⁴⁻²²⁸. These observations are consistent with a fast exchange regime on the chemical shift time scale.

Surprisingly, the residues showing the highest CSPs upon binding to Pin1 did not align with the six mutated Glu-Pro sites (figure 104), with the exception of residues surrounding Glu214-Pro215 that were significantly affected. Furthermore, Glu122 exhibited a CSP exceeding the threshold level, defined as the average CSP plus the standard deviation ($m + \sigma$), although adjacent residues did not demonstrate substantial CSP values. The most significant perturbations were instead clustered around residues 190-200, a region that coincides with the segment of the Cdc25C⁴⁴⁻²²⁸ sequence displaying a slight restriction in mobility and a helical propensity (see section 8.2.2). What is further surprising is that no significant CSPs were observed around the sites Glu48-Pro49 or Glu67-Pro68, or in fact any residue before position 100. Given that the literature points to phosphorylated Thr48-Pro49 and Thr67-Pro68 as the primary binding sites for Pin1, the evident lack of binding-induced perturbations at these sites suggests that the phosphomimetic mutant might not accurately mimic phosphorylation at these specific positions.

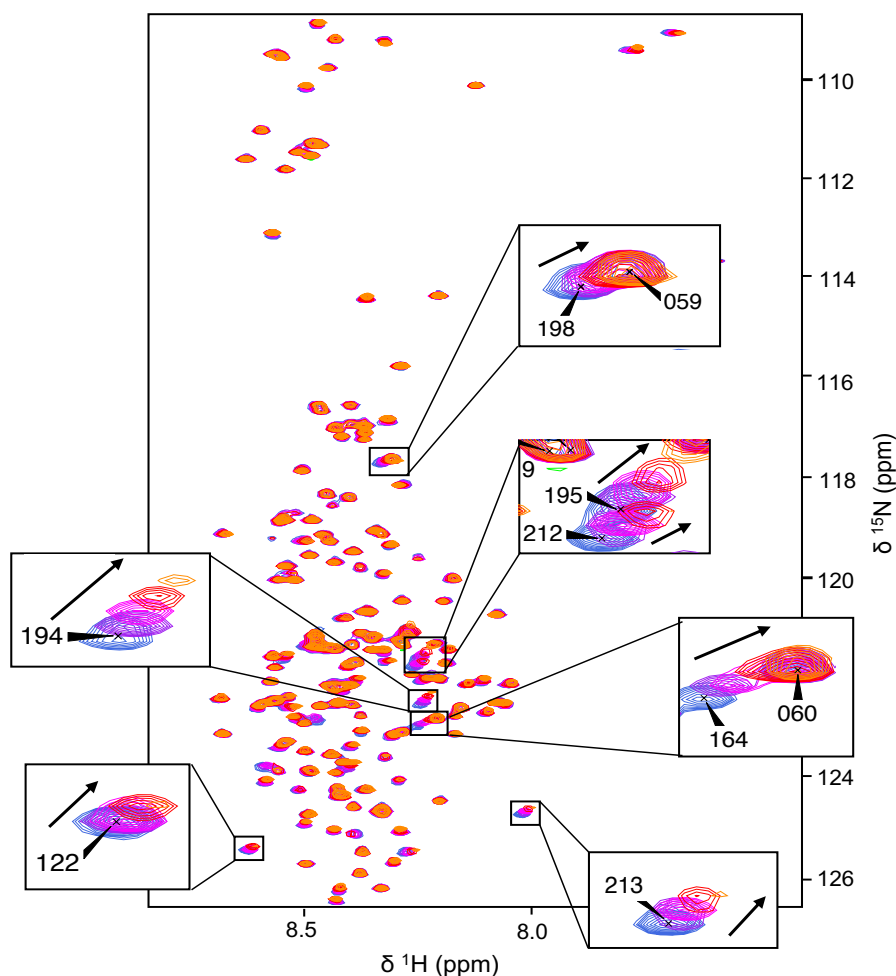


Figure 103: Overlay of ^1H - ^{15}N BEST-TROSY spectra of pM-Cdc25C⁴⁴ throughout a titration with up to 4 molar equivalents of unlabelled Pin1. The overlay displays spectra recorded at 0 (blue), 0.5 (purple), 1.0 (magenta), 2.0 (red) and 4.0 (orange) molar equivalents of Pin1. Direction of peak movement is indicated by arrows.

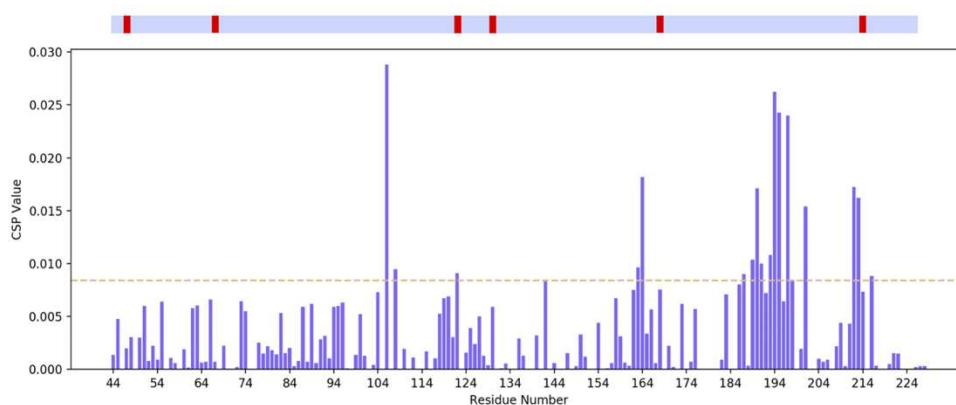


Figure 104: Binding-induced CSPs of amide cross-peaks in ^1H - ^{15}N BEST-TROSY spectra of the Cdc25C⁴⁴⁻²²⁸ phosphomimetic upon binding to 4 molar equivalents of Pin1. The bar above the CSP chart indicates the six mutated sites (red) on the sequence of Cdc25C⁴⁴⁻²²⁸ (blue). CSP values above the threshold CSP value ($m + \sigma$) are represented by the orange dashed line.

In terms of binding affinity, although binding saturation was not attained for residues displaying the highest CSPs (even in the presence of 4 equivalents of Pin1), K_D values could be estimated for these residues from CSP curves. Variable K_D values in the micromolar-millimolar range were determined, the highest binding affinities being observed for residues in the region 190-200 with an average K_D of $400 \pm 150 \mu\text{M}$. The C-terminal residues surrounding Glu214-Pro215 exhibited an average K_D of $1.7 \pm 1.0 \text{ mM}$ and bound Pin1 with a significantly lower affinity. In the case of residues 106 and 108, the binding curves were characteristic of very weak interactions with K_D values beyond 10 mM, despite the large CSP measured for residue 106. These data highlighted the region 190-200 as the predominant site for the binding to Pin1.

In terms of structure and dynamics, the observation of a fast exchange regime with weak CSPs and limited line broadening indicates that the interacting residues of pM-Cdc25C⁴⁴⁻²²⁸ remain disordered upon binding to Pin1. Indeed, in the case of a folding-upon-binding mechanism, one would expect severe line broadening or signal disappearance for residues in the interacting regions that would adopt the correlation time of a folding protein within the complex. The differences in chemical shifts are also expected to be more pronounced between the free and the bound forms in the case of folding-upon-binding, which could lead to the intermediate exchange regime. The persistence of disorder in the bound form of pM-Cdc25C⁴⁴⁻²²⁸ is also likely to explain why amide resonances of the interacting residues are still observed when Pin1 is in excess, even if the interaction results in the formation of a high molecular weight complex.

Taken together, these NMR titrations showed that the substitution of the 6 Thr/Ser-Pro motifs by Glu-Pro in the Cdc25C⁴⁴⁻²²⁸ protein did not increase substantially its binding affinity for Pin1. More importantly, while the predominant Pin1 binding site in pM-Cdc25C⁴⁴⁻²²⁸ is localised in the C-terminal region 190-200 that lacks Glu-Pro, the 6 Glu-Pro motifs were not shown to contribute significantly to the interaction. As a result, the phosphomimetic mutant enriched in glutamate residues does not appear to be a relevant mimic of the hyperphosphorylated form of Cdc25C⁴⁴⁻²²⁸, underscoring the need for the generation of phosphorylated forms of Cdc25C.

8.4 Phosphorylation of Cdc25C⁴⁴⁻²²⁸

8.4.1. Challenges with multi-site-specific phosphorylation

The investigation of proteins containing multiple phosphorylation sites introduces several notable challenges. Foremost among these challenges is the task of achieving selective phosphorylation of specific sites. Phosphorylation of serine, threonine or tyrosine residues can be achieved through various means, including recombinant kinases, cell extracts, or amber codon suppression, each method carrying its own advantages and disadvantages. Despite the chosen method, acquiring a homogeneous population of phosphorylated proteins is another important challenge. In this regard, characterising of the phospho-form, i.e., the precise identification of the phosphorylated sites, is a crucial step before studying the interaction of phosphorylated proteins. The use of mass spectrometry and antibodies, while common tools, often lack the specificity required for pinpointing these sites accurately. In anticipation of addressing these complexities while investigating the phosphorylation of Cdc25C⁴⁴⁻²²⁸, an overview of several phosphorylation and detection methods is given below.

Phosphorylation methods

The conventional method for generating phosphorylated proteins involves *in vitro* phosphorylation reactions utilising **recombinant kinases**, offering simplicity and broad applicability. The native upstream kinase of the protein of interest can either be produced and activated in-house or obtained commercially. However, this approach is constrained by the vast number of kinase-substrate pairings that remain unidentified. Even when the upstream kinase is known, cellular phosphorylation patterns are often more intricate than those elicited by a single kinase. Thus, obtaining the functional phosphoprotein can prove challenging. The intricacies surrounding Cdc25C underscore these complexities. Notably, phosphorylation by CDK1 alone was initially implicated in Cdc25C inactivation by Pin1. However, subsequent investigations suggested that Pin1-mediated activation of Cdc25C requires phosphorylation by both CDK1 and Plx1 (Stukenberg & Kirschner, 2001). Hence phosphorylation with a different kinase can largely impact the nature of the interaction between Cdc25C and Pin1.

An alternative approach to the use of specific kinases is the *in vitro* phosphorylation of proteins using crude **cell extracts** and the subsequent purification of the phosphorylated protein of

interest, often using affinity chromatography. This method better simulates the native cellular context, utilising the full spectrum of cellular kinases for target phosphorylation. In the case of Cdc25C, mitotic cell extracts have been successfully used to produce the hyperphosphorylated mitotic form of Cdc25C, which is a characterised Pin1 substrate (Lu *et al*, 1999b; Zhou *et al*, 2000). However, such cell extracts are expected to harbour the Pin1 protein, which could already exert its regulatory effect upon the phosphorylated Cdc25C⁴⁴⁻²²⁸ protein prior to its isolation. As a result, subsequent studies aimed at investigating the interaction between Pin1 and Cdc25C⁴⁴⁻²²⁸ may yield misleading outcomes. To avoid this, an alternative strategy involves employing mitotic cell extracts wherein the Pin1 protein has been selectively depleted (Stukenberg & Kirschner, 2001). It is important to acknowledge that the production of such extracts demands a considerable investment of time and falls outside the specialised focus of this laboratory's expertise. Both recombinant kinase and crude cell extract-based *in vitro* phosphorylation methods additionally encounter limitations in generating homogenous phosphorylated protein populations. Moreover, controlling the phosphorylated sites proves challenging, necessitating mutation of undesired phosphorylation sites to alanine—a strategy impractical for Cdc25C⁴⁴⁻²²⁸, which houses 22 serine and 10 threonine residues.

An additional possibility for producing phospho-proteins is the use of **amber codon suppression**, a technique which makes use of the genetic code to incorporate unnatural amino acids into recombinant proteins using an *E. coli* expression system (Park *et al*, 2011). This technique makes use of the least used STOP codon in *E. coli*, the amber codon (UAG), to encode an unnatural amino acid of choice provided that the corresponding orthogonal tRNA bearing the phosphorylated residue and the tRNA synthetase enzyme are supplied. This method has been successfully applied for the incorporation of pSer and pThr residues at desired sites (Park *et al*, 2011; Zhang *et al*, 2017). It allows the generation of samples with highly controllable phosphorylation patterns and, in the case of the Cdc25C IDR, could allow the production of specific phospho-variants, for instance with all or specific combinations of the six S/T-P sites phosphorylated. However, its implementation necessitates specialised expertise and the production of multiple mutants. Given the challenges in obtaining high yields of unphosphorylated WT Cdc25C⁴⁴⁻²²⁸ samples and the time constraints of this thesis, we did not choose this strategy to phosphorylate our Cdc25C construct.

Detection methods

Another challenge associated with the structural study of multi-site phosphorylation is the ability to detect individual phosphorylation events with a resolution that provides comprehensive information about each site. Various methodologies have thus been developed to detect and characterise the phosphorylated forms of proteins. Commonly, **SDS-PAGE** is used to track protein phosphorylation as many proteins display a phosphorylation-dependent electrophoretic mobility shift (PDEMS) (Lee *et al*, 2019). Phosphorylation of certain protein sequences generally results in reduced binding to SDS, hence slowing migration of phosphoproteins through the polyacrylamide gel compared to their unphosphorylated forms. Specifically, the decreased binding of SDS is associated with the presence of electrophoretic mobility shift (EMS) motifs: \emptyset -Xaa₁₋₃- \emptyset -Xaa₁₋₃- \emptyset , whereby \emptyset represents a negatively charged or phosphorylated amino acid, and Xaa represents any amino acid. Notably, proteins displaying a higher net negative charge can exhibit this phenomenon despite phosphorylation, hence the effect of PDEMS can be difficult to observe. Although this technique does not provide any information about which specific sites are phosphorylated, it can be useful for confirming the ratio of phosphorylated:unphosphorylated protein.

A common method for the detection of protein phosphorylation at specific sites is via Western blot analysis using phospho-amino acid-specific **antibodies** (Kaufmann *et al*, 2001). However, this method carries several limitations. Firstly, it requires the raising of antibodies against a specific phospho-site, meaning the expected phosphorylation sites must be anticipated. Significantly, the production of such antibodies can be an expensive and time-consuming process. Moreover, this means that additional phosphorylation at unexpected sites will remain undetected.

More recently, quantitative **mass spectrometry (MS)** has become a popular tool for the characterisation of phosphorylated proteins (Yu & Veenstra, 2021). Identifying specific phospho-sites first involves the enzymatic cleavage of the phosphorylated protein into short peptide sequences. These peptides are then analysed by MS and their resulting masses compared to the measured mass of peptides obtained from enzymatic cleavage of the unphosphorylated protein. Given that a single phosphorylation event contributes around 80 Da of additional mass, it is possible to identify the number of phosphorylations in the protein

sequence as well as their position, provided a MS instrument with high accuracy is used. Another aspect of phosphorylation detection using antibodies or MS is that it is not possible to monitor the phosphorylation reaction in real-time, with both methods requiring the obtention of fully-phosphorylated protein before detection of phospho-sites.

NMR spectroscopy has arisen as a powerful method to provide residue-specific information about protein phosphorylation. The phosphorylation of serine and threonine residues results in significant downfield chemical shift changes of Ser/Thr backbone ^1H and ^{15}N amide peaks which can be monitored using 2D ^1H - ^{15}N NMR (Julien *et al.*, 2020). This is particularly useful for studying the phosphorylation of IDPs, which display a characteristic weak dispersion of signals in the proton dimension of 2D ^1H - ^{15}N spectra, hence the downfield shift of ‘phospho-peaks’ is very easily detected. In addition, this technique permits the real-time monitoring of the phosphorylation reaction, provided that the reaction is slow enough to be monitored by sequential experiments. While the traditional HSQC experiment could take too long for efficient resolution of kinetics, faster experiments for instance the SOFAST-HMQC or BEST-HSQC/TROSY, can be applied to obtain high resolution spectra in less time.

8.4.2. Selection of the kinase for *in vitro* phosphorylation of Cdc25C⁴⁴⁻²²⁸

For preliminary studies, it was decided that the *in vitro* phosphorylation of Cdc25C⁴⁴⁻²²⁸ using a Pro-directed Ser/Thr kinase would be the most simple and time-efficient method for producing the hyperphosphorylated form, since the optimisation of the production and purification of WT Cdc25C⁴⁴⁻²²⁸ had been a time-consuming process. An advantage of this method of phosphorylation is that it would allow the time resolution of the phosphorylation reaction by NMR. From literature studies, it is clear that a good starting point would be the CDK1/cyclin B mitotic partner of Cdc25C which is known to phosphorylate the five S/T-P sites which characterise the hyperphosphorylated mitotic form of Cdc25C (Thr48, Thr67, Ser122, Tr130 and Ser214). However, the recombinant production of the CDK1 kinase in insect cells and its subsequent purification and activation would be a time-consuming process, and hence was not possible within the scope of this thesis. Additionally, the purchase of commercial CDK1/Cyclin B is expensive (548€/10 ng) and would not allow large-scale phosphorylation trials. Thus, other Pro-directed Ser/Thr kinases were considered, such as p38 and ERK2, which were both kindly provided by Dr. François-Xavier Theillet from the Institute

for Integrative Biology of the Cell (I2BC) at Gif-sur-Yvette. The MAPK, p38, has previously been shown to phosphorylate Cdc25C *in vitro* at the Pro215-Ser216 site (Bulavin *et al*, 2001). Intriguingly this site does not correspond to the typical recognition site of MAPKs, which consists of a proline at the +1 or -2 position with respect to the S/T phosphorylation site (López-Avilés *et al*, 2005). The phosphorylation of Ser216 has been described as the most predominant inhibitory phosphorylation of Cdc25C, forming an essential binding site for 14-3-3 proteins which sequester Cdc25C in the cytoplasm during interphase (see chapter 2.3). Hence, for the present study, phosphorylation of this residue was not desired. By contrast, *in vitro* assays showed the ability of the ERK2 kinase to successfully phosphorylate three Ser/Thr-Pro sites within the IDR of Cdc25C, namely Thr48, Thr130 and Ser168 (Wang *et al*, 2007). ERK2 was also shown to actively phosphorylate Thr48 at mitosis in mammalian cell lines, and hence is suspected to play a role in activating Cdc25C at the G2 to M transition. While the phosphorylation of Ser168 was also suggested as an inhibitory phosphorylation of Cdc25C, phosphorylation of the other two sites, in particular Thr48, has been reported to be significant for the interaction between Cdc25C and Pin1. Therefore, ERK2 was selected for the *in vitro* phosphorylation of Cdc25C⁴⁴⁻²²⁸ to assess if phosphorylation of these three sites were sufficient for significant binding to Pin1, and if Pin1 displayed a binding preference between the three sites.

8.4.3. Monitoring *in vitro* phosphorylation of Cdc25C⁴⁴⁻²²⁸ by SDS-PAGE

To optimise the phosphorylation conditions, the phosphorylation reaction was first monitored by SDS-PAGE. It was indeed necessary to determine the kinase concentration to use that would allow complete phosphorylation of the protein within 5 hours, as kinases typically display a significant loss of activity after 5 hours at room temperature (Julien *et al*, 2020). Samples of 25 µM unlabelled Cdc25C⁴⁴⁻²²⁸ were prepared in 25 µL of a suitable buffer for *in vitro* phosphorylation (see details in the Material and Methods section). Three concentrations of ERK2 were originally trialled – 250, 125 and 25 nM – corresponding to respective ERK2:Cdc25C ratios of 1:100, 1:200 and 1:1000. Reactions were run at room temperature and aliquots taken at 0, 10, 30, 60 and 240 minutes were heated to 95°C to stop the reaction. Subsequent SDS-PAGE analysis revealed the presence of a visible electrophoretic mobility shift after incubation with 250 or 125 nM ERK2 for 5 hours (see figure 105).

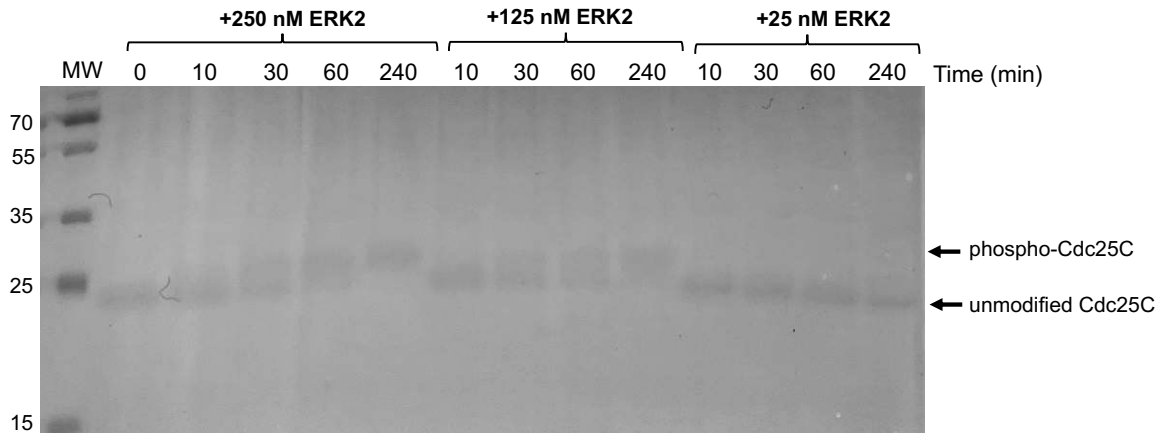


Figure 105: SDS-PAGE analysis of phosphorylation reactions of 25 μ M Cdc25C⁴⁴⁻²²⁸ using 250, 125 or 25 nM ERK2. Samples run on a 1.5 mm SDS-PAGE gel.

Since the resulting SDS-PAGE gel was not clear enough to quantify the ratio of phosphorylated:unphosphorylated protein after 240 mins, the reaction was repeated for the sample containing a 1:100 ratio of ERK2:Cdc25C. In this second trial, aliquots were taken at 0, 10, 30, 60, 120 and 300 minutes, and the amount of protein loaded onto the SDS-PAGE gel was increased. Analysis of the resulting SDS-PAGE gel using ImageJ software (figure 106) confirmed that ~90% of the population was phosphorylated after incubation for 5 hours at room temperature with ERK2 at a ERK2:Cdc25C ratio of 1:100.

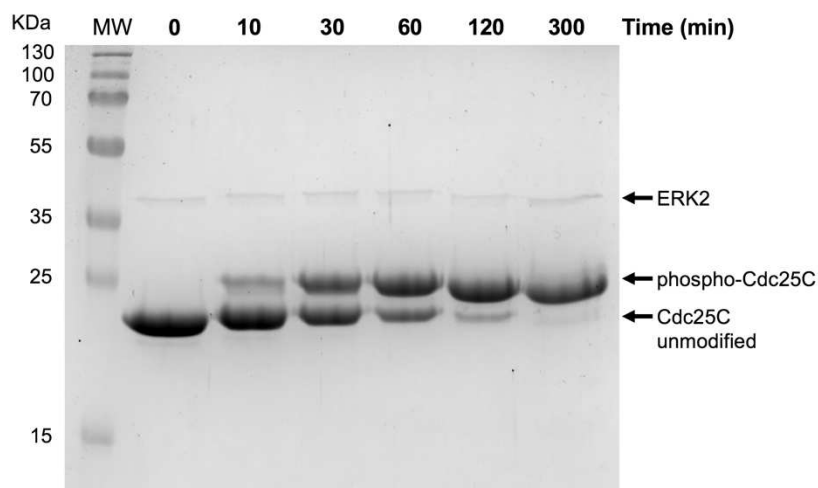


Figure 106: SDS-PAGE analysis of phosphorylation reactions of 25 μ M Cdc25C⁴⁴⁻²²⁸ using 250 nM ERK2. Samples run on a 0.75 mm SDS-PAGE gel.

8.4.4. Monitoring *in vitro* phosphorylation of Cdc25C⁴⁴⁻²²⁸ by NMR

To identify residues of Cdc25C⁴⁴⁻²²⁸ that were phosphorylated by ERK2, the phosphorylation reaction was monitored using 1D ¹H and 2D ¹H-¹⁵N SOFAST-HMQC experiments. These experiments were performed at pH 7.0, 298K on a 100 μM sample of ¹⁵N-Cdc25C⁴⁴⁻²²⁸ using a 500 MHz spectrometer equipped with a cryoprobe. In these conditions, the use of SOFAST-HMQC experiments allowed acquisition of high-resolution 2D ¹H-¹⁵N spectra within 20 minutes compared to over 1 hour for the standard HSQC experiment, while maintaining a good signal to noise ratio. By enabling the acquisition of 2D ¹H-¹⁵N spectra every 20 minutes, it was possible to determine kinetics parameters associated with each phosphorylation event.

Preliminary trials monitoring phosphorylation using SOFAST-HMQC experiments

Based on the results obtained by SDS-PAGE, the phosphorylation reaction was initially monitored using a 1:100 ratio of Cdc25C⁴⁴⁻²²⁸:ERK2. Reference 1D spectra were recorded at the start and end of the reaction to monitor the stability of the sample and the ATP:ADP ratio, ensuring that ATP was still in excess (figure 107a). Analysis of the resulting SOFAST-HMQC spectra revealed the apparition of several peaks after incubation with 1 μM ERK2 for 6 hours (figure 107b). Four of these peaks exhibited a ¹H frequency >8.7 ppm characteristic of phosphorylated residues. In addition, several peaks disappeared over time, likely corresponding to residues of the unphosphorylated protein which are in close proximity to the phospho-sites. The prior assignment of WT Cdc25C⁴⁴⁻²²⁸ at pH 7, 298 K (detailed in section 8.2.1) enabled confirmation that the peaks corresponding to unphosphorylated Thr48, Thr130 and Ser168 completely disappeared over the course of the reaction. However, as there were multiple phosphorylation events, an assignment of the 2D ¹H-¹⁵N spectrum of the phosphorylated-Cdc25C⁴⁴⁻²²⁸ protein was necessary to determine to which residue corresponds each appearing phospho-peaks. Notably, no large-scale folding was evident upon phosphorylation.

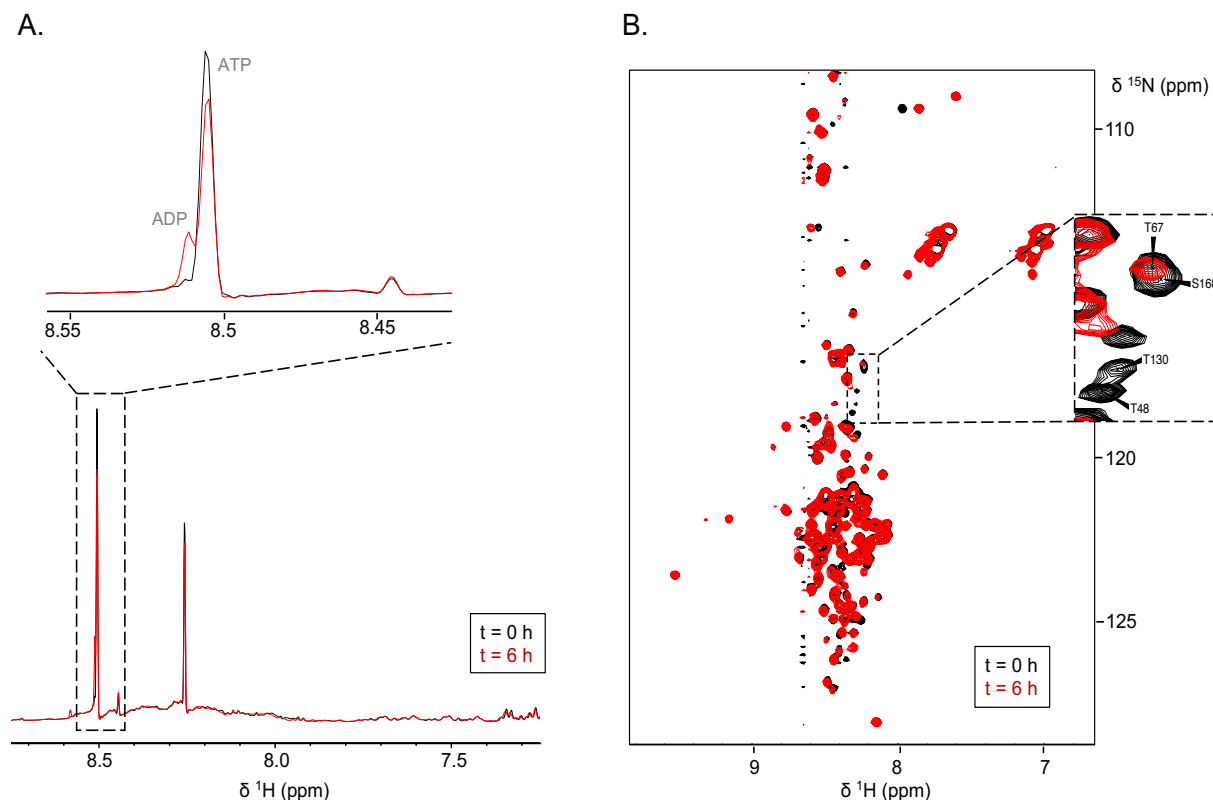


Figure 107: Overlay of 1D ^1H (A) and 2D ^1H - ^{15}N SOFAST-HMQC (B) spectra recorded at $t=0$ h and $t=6$ h after incubation of ^{15}N -Cdc25C⁴⁴⁻²²⁸ with ERK2 kinase (at a ratio of 1:100 protein:kinase). A. The zoom displays the region of the 1D spectra where the peaks corresponding to ATP (8.51 ppm) and ADP (8.52 ppm) can be visualised. The ATP is still in excess after 6 hours (red spectrum) as indicated by a higher intensity for the ATP peak than the ADP peak. **B.** The appearance of four downfield peaks is evident in spectra recorded after incubation for 6 hours with ERK2, likely corresponding to peaks of phosphorylated residues. The zoom shows the disappearance of residues corresponding to unphosphorylated T48, T130 and S168.

Backbone amide assignment of phospho-Cdc25C⁴⁴⁻²²⁸

In order to perform a backbone assignment, a double-labelled $^{13}\text{C}/^{15}\text{N}$ sample of Cdc25C⁴⁴⁻²²⁸ was produced and phosphorylated by ERK2. Triple resonance $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ 3D experiments were performed using a 500 MHz spectrometer. Experiments were recorded at pH 6.5, 283 K so that the assignment could be facilitated by the previous assignments of H^{N} , N^{H} , CO, $\text{C}\alpha$, and $\text{C}\beta$ resonances of WT Cdc25C⁴⁴⁻²²⁸ under these conditions. HNC(O) and HNCACB spectra were indeed sufficient to assign the backbone resonances of phospho-Cdc25C⁴⁴⁻²²⁸ starting from the chemical shifts of WT Cdc25C⁴⁴⁻²²⁸ (figure 108) (annex 1). The assignment confirmed the phosphorylation of three residues by ERK2: Thr48, Thr130 and Ser168. The

fourth significant downfield shift observed upon phosphorylation corresponded to residue Asn132 neighbouring the pThr130 site.

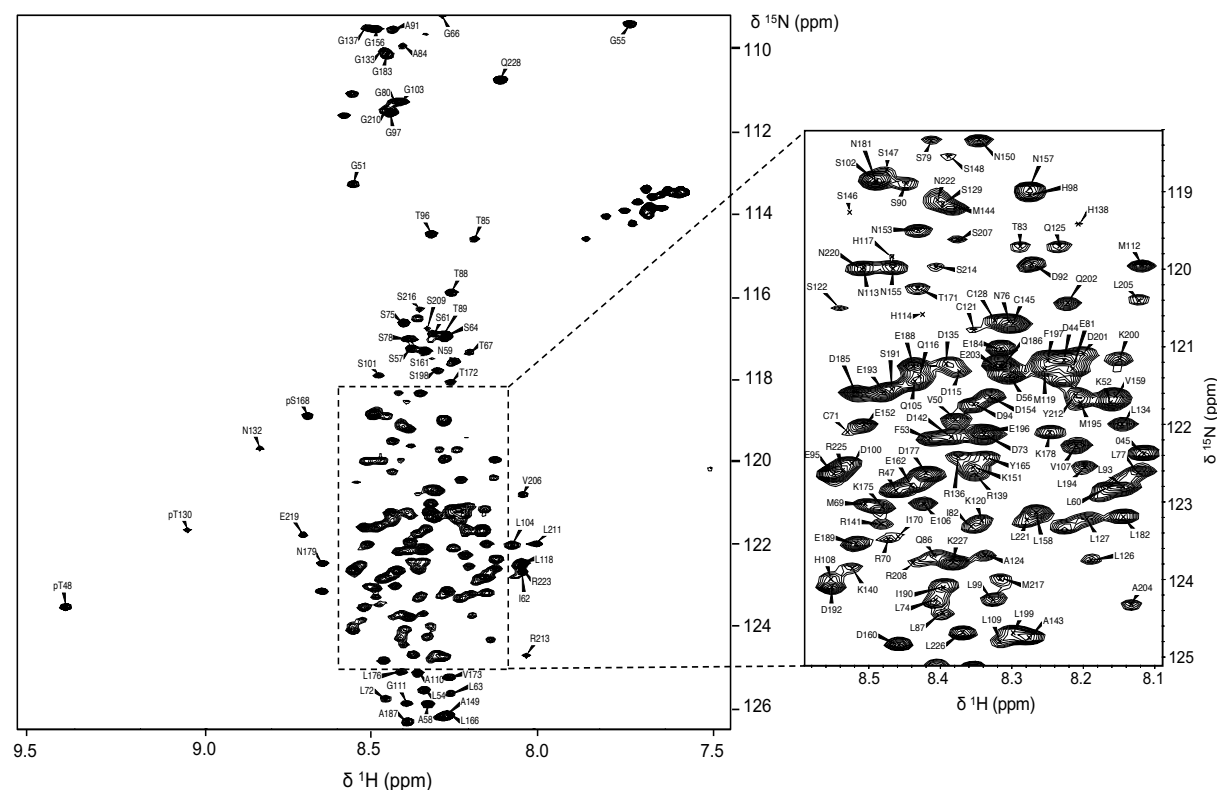


Figure 108: Assigned ^1H - ^{15}N -HSQC spectrum of tri-phosphorylated Cdc25C⁴⁴⁻²²⁸, recorded at pH 6.5, 298 K, using a 500MHz spectrometer equipped with cryoprobe. Three residues, T48, T130 and S168, showed complete phosphorylation after incubation with 1:100 concentration of ERK2: Cdc25C⁴⁴⁻²²⁸ for 6 hours.

Additional low-intensity peaks with a ^1H frequency >8.7 ppm were also observed in the ^1H - ^{15}N spectrum of phospho-Cdc25C⁴⁴⁻²²⁸ after 6-hours incubation with ERK2. This minor population of phospho-Cdc25C⁴⁴⁻²²⁸ was attributed to the partial phosphorylation of residues Thr67 and Thr83 (see figure 109). Remarkably, pThr67-Pro68 is a potential Pin1 binding site unlike Thr83 which is followed by Ala84 in the sequence of Cdc25C. While elevating kinase concentration could potentially achieve complete phosphorylation of these sites, such an approach was dismissed due to concerns about the potential interference of pT83 with the Pin1 interaction. Furthermore, the chosen kinase concentration was already substantial, and a further increase would entail significant kinase usage. Consequently, we used this tri-phosphorylated form of Cdc25C⁴⁴⁻²²⁸ to characterise its interaction with Pin1.

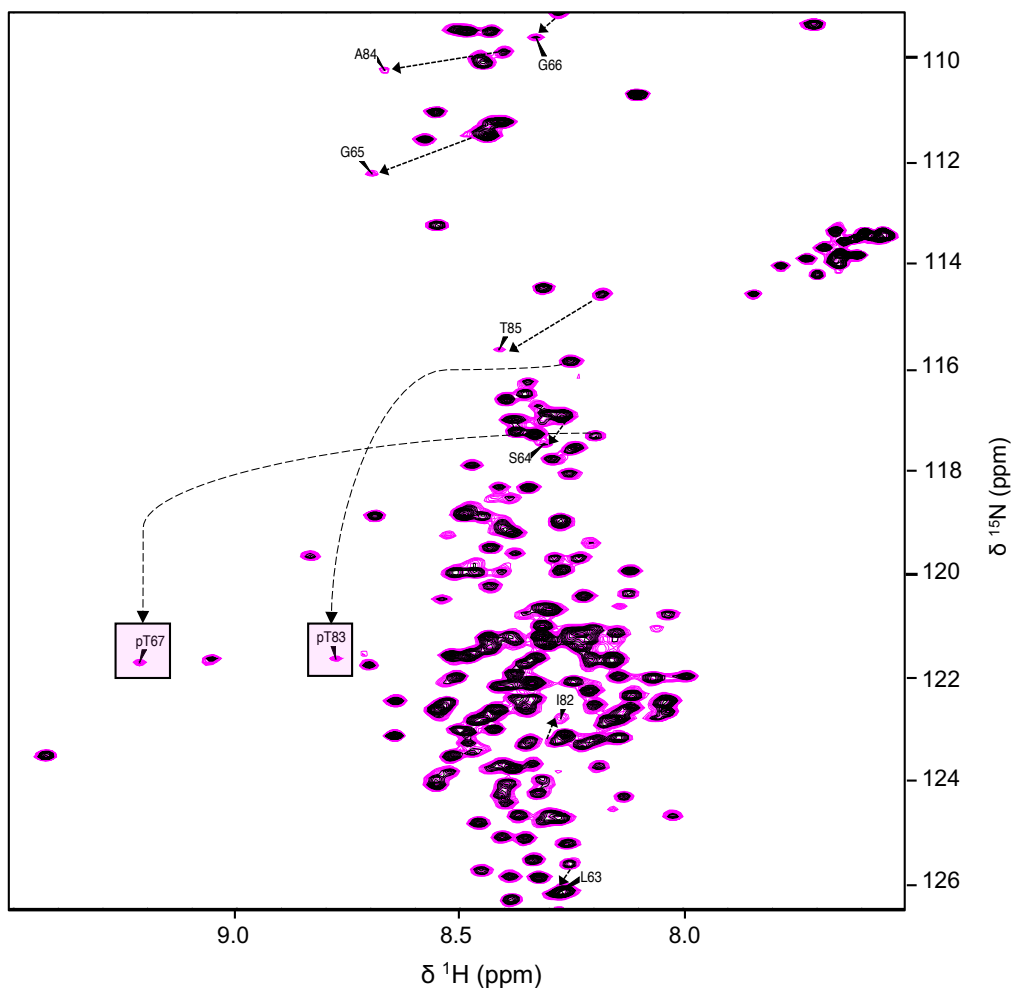


Figure 109: Minor phosphorylated population of Cdc25C⁴⁴⁻²²⁸. The overlay shows the assigned ¹H-¹⁵N HSQC spectrum of phospho-Cdc25C⁴⁴⁻²²⁸ in black with the contour level increased (in magenta) to reveal additional peaks representing a minor population with additional phosphorylation of residues T67 and T83 (highlighted by pink squares). Both of these residues and their neighbouring residues display a large downfield shift in the ¹H frequency indicated by the dashed arrows. The spectrum was recorded at pH 6.5, 298 K, using a 500 MHz spectrometer.

To align with the experimental conditions of interaction studies with Pin1 (298 K and pH 7), temperature and pH titrations were conducted on the tri-phosphorylated sample of Cdc25C⁴⁴⁻²²⁸, utilising a series of HSQC experiments recorded at 500 MHz. After completing both titrations, a total of 104 resonances were confidently assigned, accounting for 60% of all non-proline residues. Notably, some resonances in close proximity to the phospho-sites remained unassigned (including D44, V45, R47, K50, F52, D53, S129, L134, K164, Y165, I170, and T171). As these residues could play a role in the interaction with Pin1, an additional backbone assignment of triphospho-Cdc25C⁴⁴⁻²²⁸ was pursued at pH 7 and 298 K. By analysing

HNCO, HNCA and HNCACB spectra recorded at 950 MHz, an additional 41 backbone amide resonances were successfully assigned, including those in close proximity to the phospho-sites. An overall assignment of 145 out of 174 non-proline resonances was possible at pH 7, 298K, representing an overall sequence coverage of 83 %.

Real-time monitoring of phosphorylation kinetics using SOFAST-HMQC experiments

The conversion of unphosphorylated Cdc25C⁴⁴⁻²²⁸ to phosphorylated Cdc25C⁴⁴⁻²²⁸ exhibits a slow exchange regime on the NMR timescale, meaning that integration of disappearing ‘unphosphorylated’ and appearing ‘phosphorylated’ cross-peaks can provide information about the rate of the phosphorylation process. To enhance the temporal resolution of the phosphorylation reaction, a series of SOFAST-HMQC experiments were conducted on 100 μ M samples of Cdc25C⁴⁴⁻²²⁸ using ERK2:Cdc25C⁴⁴⁻²²⁸ ratios of 1:200, 1:1000, and 1:2000. The phosphorylation of Thr48 was too rapid when using ERK2:Cdc25C⁴⁴⁻²²⁸ ratios of 1:200 and 1:1000, which did not allow to accurately elucidate the initial reaction rate for this residue. Reducing the ERK2 concentration to 50 nM (yielding a ERK2:Cdc25C⁴⁴⁻²²⁸ ratio of 1:2000) was most appropriate and the reaction was monitored for a duration of 20 hours.

Phosphorylation analysis was conducted by focusing on spectral peaks corresponding to unphosphorylated and phosphorylated forms of Thr48, Thr130, and Ser168, as well as well-isolated peaks of neighbouring residues (Gly51, Asn132, and Gly167) that displayed chemical shift changes due to phosphorylation (figure 110a). These peaks were integrated and normalised to serve as reporters for each phosphorylation event. Normalisation was achieved by dividing each value by the peak’s maximum intensity (I/I_{\max}). Subsequently, the normalised values were plotted over time for each site, as depicted in Figure 110b.

For the calculation of initial phosphorylation rates, the volumes of the ‘unphosphorylated’ peaks, which exhibited consistent decreases over time, were used. This choice was motivated by the fact that phosphorylation reactions rarely reach 100%, rendering I_{\max} less robust for phosphopeaks and, consequently, less accurate for curve fitting and initial rate extraction. Specifically, unphosphorylated peaks corresponding to T48/G51 and S168/G167 were identified as suitable candidates for rate extraction, as they exhibited consistent decreases over

time. However, for Thr130 phosphorylation, the unphosphorylated cross-peak did not display an accurate decay trend due to a lower signal-to-noise ratio, we thus utilised the phospho-peaks corresponding to pT130 and N132_(pT130) for extracting the initial reaction rate at this specific site. Notably these peaks display higher signal-to-noise ratios which should increase the accuracy of the measurement.

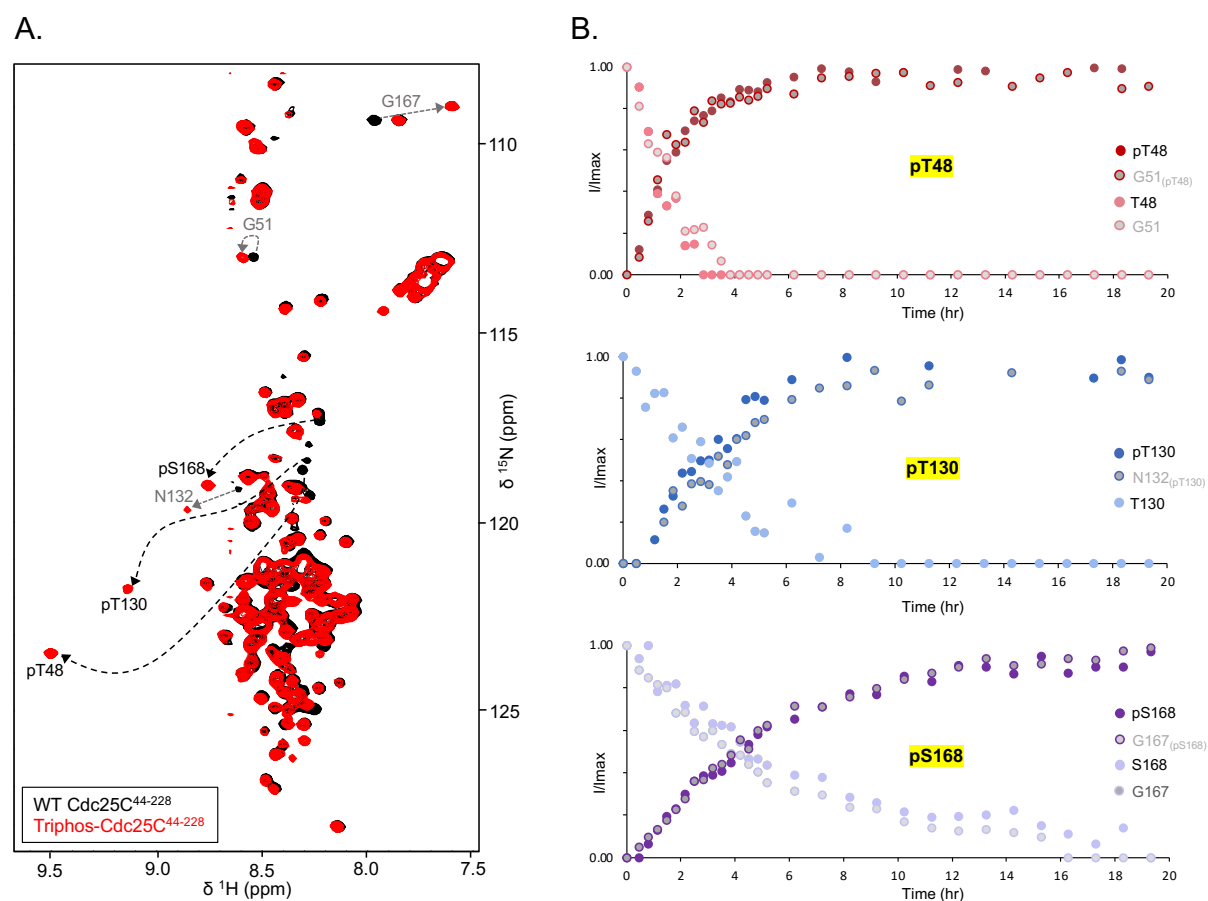


Figure 110: Monitoring phosphorylation of Thr48, Thr130 and Ser168 of Cdc25C⁴⁴⁻²²⁸ by ERK2 using ¹H-¹⁵N SOFAST-HMQC experiments. A. The ¹H-¹⁵N SOFAST-HMQC spectra of unphosphorylated (black) and phosphorylated (red) Cdc25C⁴⁴⁻²²⁸ at pH 7, 298 K recorded using a 500 MHz spectrometer. Phosphorylation-induced chemical shift perturbations of key resonances are indicated by arrows. B. The volumes of peaks corresponding to unphosphorylated residues (peaks in black spectrum) T48, G51, T130, S168, and G167 and phosphorylated residues (peaks in red spectrum) pT48, G51_(pT48), pT130, N132_(pT130), pS168 and G167_(pS168) were integrated and normalised by their maximum intensity values. The values were plotted against time for each phospho-site.

To plot phosphorylation build-up curves for T48 and S168, first the raw intensities of the corresponding unphosphorylated peaks in spectra recorded throughout the reaction were added

to the raw intensities of their corresponding reference peaks, G51 and G167, to enhance the signal-to-noise ratio. The summed intensities were plotted against time and the initial slope of the resulting graphs were used to calculate I_{\max} values at $t=0$. The percentage of phosphorylation at a given time is estimated by normalising summed intensity values using the following equation:

$$\% \text{ phosphorylation } (t) = 1 - \frac{I(t)}{I_{\max}(t=0)}$$

The phosphorylation build-up curve for T130 was plotted using the summed intensity values of pT130 and N132_(pT130) cross-peaks normalised by the summed maximal peak intensity (I_{\max}). The I_{\max} value was taken from the final timepoint of the experiment ($t=20\text{h}$), under the approximation of 100% phosphorylation.

$$\% \text{ phosphorylation } (t) = \frac{I(t)}{I_{\max}(t=20)}$$

The resulting values were plotted against time, displaying a mono-exponential trend (see figure 111). The curves were fit using MATLAB, and the initial rates of phosphorylation were extracted from the initial slope.

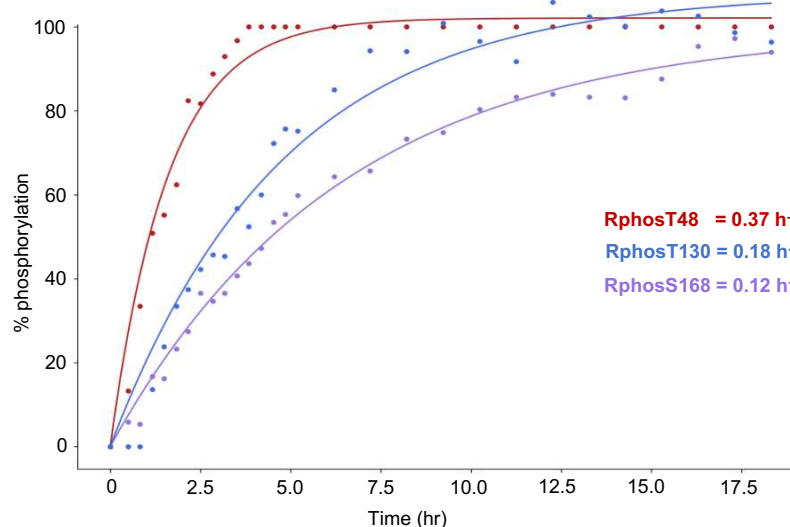


Figure 111: Phosphorylation build-up curves and initial rates calculated for each phosphorylation event.

The provided data supports the conclusion that phosphorylation initiates at the T48 site, suggesting its preferential phosphorylation by ERK2 due to a notably faster rate (0.37 h^{-1}) compared to T130 (0.18 h^{-1}) and S168 (0.12 h^{-1}). T48 reaches near-complete phosphorylation prior to phosphorylation of the other sites. For T130 and S168, their phosphorylation kinetics appear relatively comparable.

Challenges in sample production and stability of phosphorylated Cdc25C⁴⁴⁻²²⁸

A serious limitation encountered for the production of phosphorylated samples of Cdc25C⁴⁴⁻²²⁸ was the substantial loss of protein throughout the phosphorylation process and subsequent buffer exchange (see detailed protocols in the Material and Method section). This was observed by comparing the respective yields of Cdc25C⁴⁴⁻²²⁸ after initial purification of the unphosphorylated protein and after buffer exchange of the phosphorylated form (table 19).

Labelling	Yield after initial purification (mg)	Yield after phosphorylation (mg)	% loss
unlabelled	6	-	100
unlabelled	10	2.2	78
¹⁵ N	7	-	100
¹⁵ N	7.4	1.2	83
¹⁵ N ¹³ C	10	3.7	63

Table 19: Analysis of protein loss throughout the phosphorylation process of several protein batches. Yields are given in mg of purified protein obtained per L of bacterial culture after initial purification and mg of purified protein remaining after subsequent phosphorylation and buffer exchange.

The overall process was shown to lead to a protein loss ranging from 60% to 80%. Repurifying the phosphorylated Cdc25C⁴⁴⁻²²⁸ sample using gel filtration was tested as an alternative to buffer exchange, however this resulted in a comparable loss of protein. The substantial loss of protein is likely to arise from the lower stability of the phosphorylated protein in our buffer conditions or its higher sensitivity to proteases. To limit degradation of the phosphorylated samples, fresh antiproteases were prepared and added to samples at different stages of the

phosphorylation process (namely: after phosphorylation, after buffer exchange, and in the NMR tube prior to recording experiments), which unfortunately did not completely alleviate the extent of protein loss, however did prevent complete sample loss.

Due to substantial protein loss, significant protein production and purification efforts were undertaken to obtain sufficient phosphorylated protein for the subsequent interaction study with Pin1, necessitating additional time and kinase quantities.

8.4.5. Interaction between Pin1 and phosphorylated Cdc25C⁴⁴⁻²²⁸

NMR titration of ¹⁵N-Pin1 by unlabelled tri-phosphoCdc25C⁴⁴⁻²²⁸

To begin the exploration of this interaction, a CSP experiment was performed by recording ¹H-¹⁵N BEST-TROSY spectra on a ¹⁵N-labelled sample of Pin1 in the absence and presence of up to 4 molar equivalents of unlabelled triphospho-Cdc25C⁴⁴⁻²²⁸ (figure 112). For this purpose, a Pin1 NMR sample at a reduced concentration of 50 μM was prepared at pH 7.0, a strategic choice due to the limited yield of triphospho-Cdc25C⁴⁴⁻²²⁸.

As previously observed when titrating ¹⁵N-Pin1 with WT and phosphomimetic Cdc25C⁴⁴⁻²²⁸, the presence of triphospho-Cdc25C⁴⁴⁻²²⁸ induced the overall line broadening of Pin1 amide resonances. Distinct behaviours were however observed depending on the residue position in the sequence of Cdc25C and the Pin1:Cdc25C⁴⁴⁻²²⁸ ratio (figure 113). Many peaks were engaged in a fast to intermediate exchange regime with variable CSPs prior to the severe line broadening of resonances (or full signal vanishing) upon binding to more than 0.5 equivalents of triphospho-Cdc25C⁴⁴⁻²²⁸. In the WW domain, residues of the proline binding site, Gln33, Trp34, and Glu35 were highly broadened and fully disappeared at stoichiometric amounts of triphospho-Cdc25C⁴⁴⁻²²⁸. These residues align with those frequently affected upon binding to peptide ligands of the WW domain, signifying their importance in substrate binding. Residues of loop 1 of the WW domain (Arg14, Ser16, and Gly20), which is involved in phosphate group coordination, exhibited a similar level of line-broadening. A comparable loss of NMR signal was also observed for WW domain residues Asn26 and His27 that are localised at the interdomain interface, as well as residue Thr79 of the PPIase catalytic loop, and residue Ser115

of the PPIase active site. While the highest CSPs were observed in the WW domain, significant CSPs were also observed in the PPIase domain, in particular for residues Leu141 at the interdomain interface, Val150, Gly155, and Ile159 within the beta-sheet core, and Gly123, Gly128, Gln129, and Met130 within the proline binding pocket.

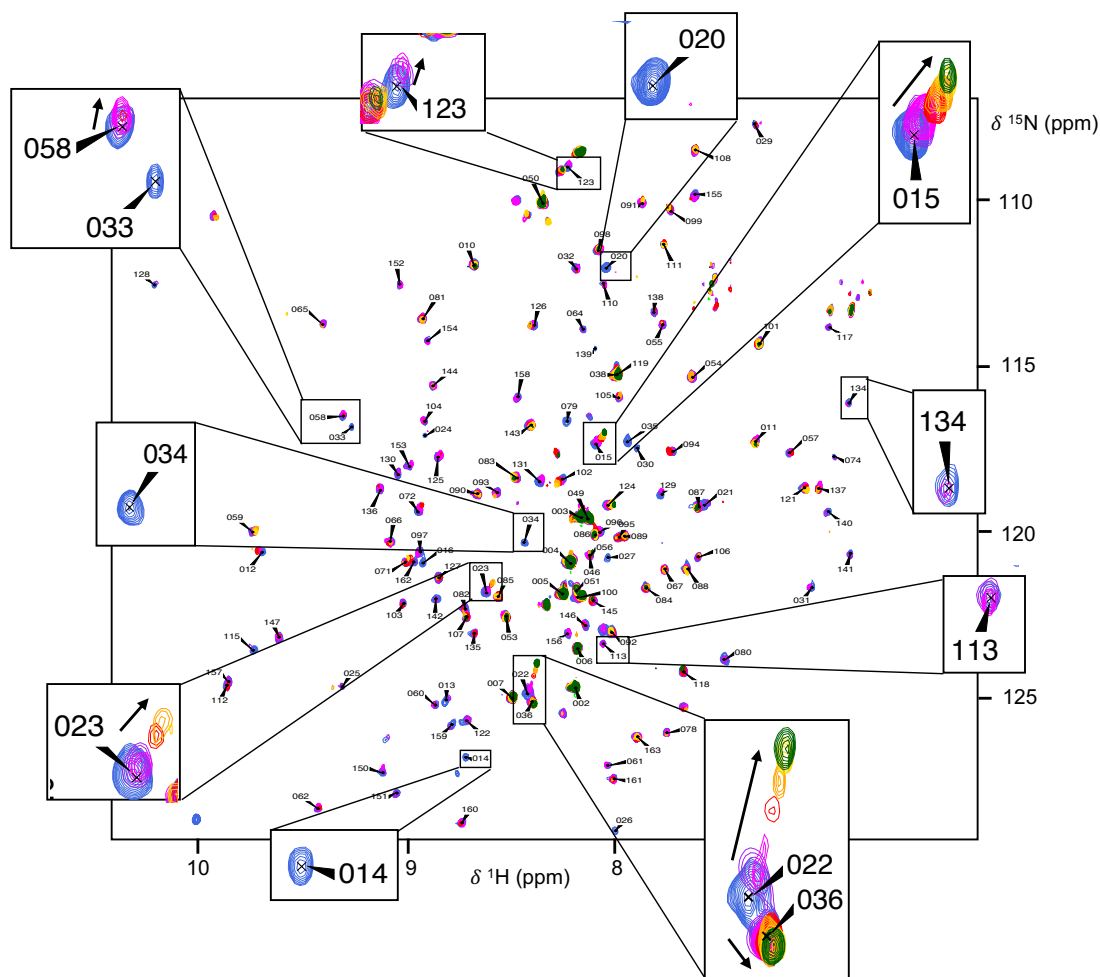


Figure 112: Overlay of ^1H - ^{15}N BEST-TROSY spectra of Pin1 recorded throughout a titration with up to 4 molar equivalents of triphospho-Cdc25C⁴⁴⁻²²⁸ using a 950MHz spectrometer equipped with cryoprobe. The overlay displays spectra recorded in the presence of increasing concentrations of triphospho-Cdc25C⁴⁴⁻²²⁸ (0, 0.1, 0.2, 0.5, 1.0, 2.0, 4.0) from blue to green. Direction of peak movement is indicated by arrows.

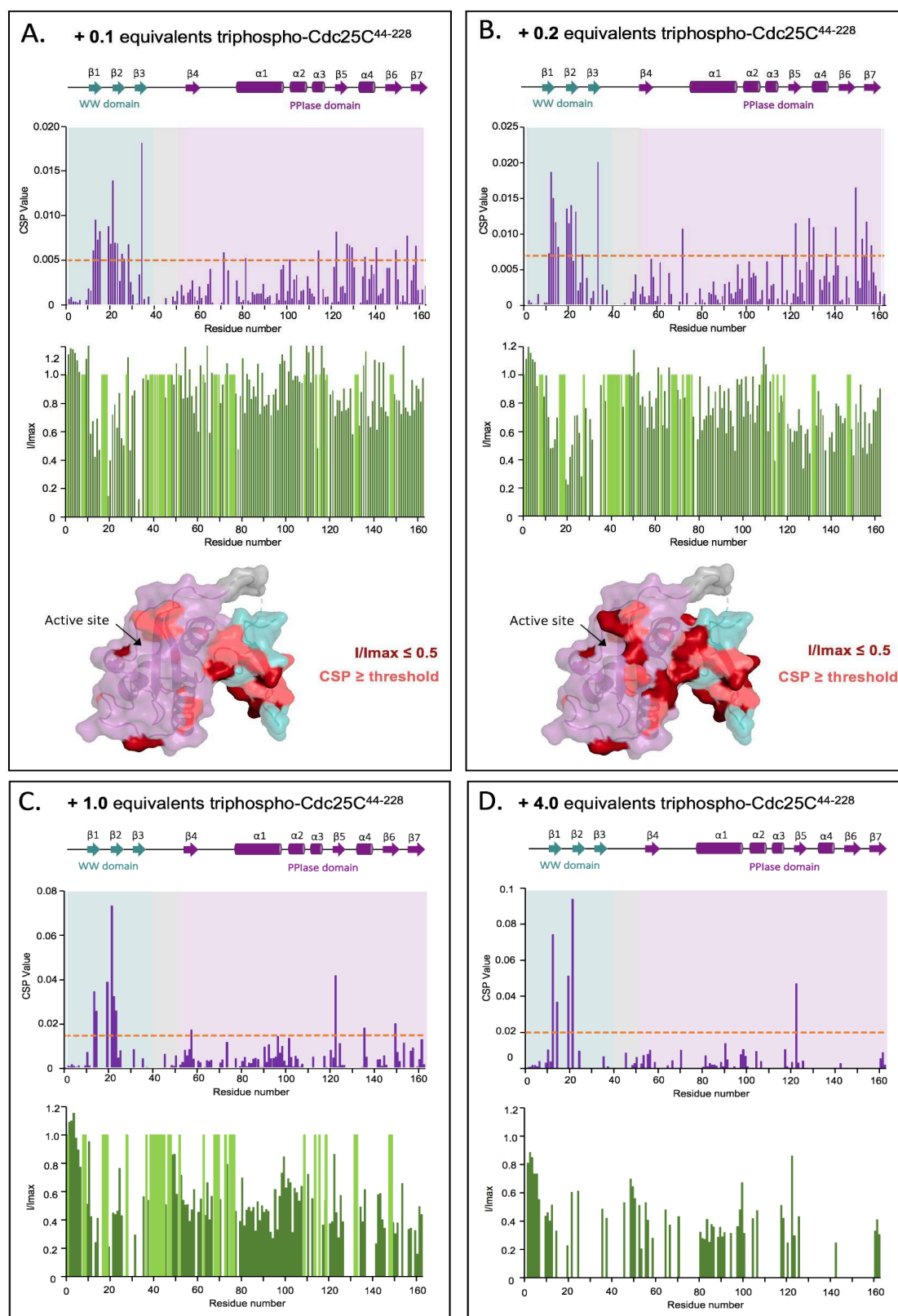


Figure 113: Analysis of CSPs and signal loss (I/I_{max}) observed for Pin1 amide resonances throughout titration with increasing concentrations of triphospho-Cdc25C⁴⁴⁻²²⁸ from 0.1 (A), 0.2 (B), 1.0 (C) and 4.0 (D) equivalents. The orange dashed line in CSP charts represents the threshold value ($m + \sigma$). Residues which were unassigned in reference spectra are represented in light green with an I/I_{max} value of 1. Significant CSPs (over the threshold value) and I/I_{max} changes (under 0.5) are mapped onto the structure of Pin1 in pale red and dark red respectively.

Overall, these results suggest that the WW domain acts as the preferential binding module for triphospho-Cdc25C⁴⁴⁻²²⁸. The perturbations observed in the PPIase domain are indeed likely to be due to an allosteric effect caused by Cdc25C binding to the WW domain rather than a direct participation to the binding. This hypothesis is corroborated by mapping the residues displaying the most significant CSPs (pale red) and residues displaying the most significant signal loss (dark red) on the structure of Pin1 (figure 113a-b). The impacted residues predominantly cluster around the WW domain's substrate binding site, interdomain interface, and specific PPIase domain residues such as those within the catalytic loop and surrounding the active site. Importantly, the amide cross-peaks of residues neighbouring the PPIase phosphate-binding site (residues 64-70) did not exhibit significant CSPs, regardless of the Pin1:triphospho-Cdc25C⁴⁴⁻²²⁸ ratio. In cases of specific binding of phosphorylated ligands to the PPIase domain, these residues often exhibit significant chemical shift changes, similar to the observations made for ligand VB-66 discussed in chapter 7.1.2. It should be noted, however, that severe line broadening was also observed at stoichiometric amounts of triphospho-Cdc25C⁴⁴⁻²²⁸ for the residues of the phosphate binding region in the PPIase domain. While this observation might suggest that the PPIase domain only binds to triphospho-Cdc25C⁴⁴⁻²²⁸ once all WW domain sites are fully occupied, the severe line broadening may also be explained by the existence of conformational dynamics in the μ s-ms time scale that may be induced by the binding of the WW domain to Cdc25C.

In terms of binding affinity, K_D values could be determined from binding curves only for amide cross-peaks exhibiting significant CSPs and modest line broadening. This was the case of residues Met15, Gly20, and Val22 in the WW domain that displayed K_D values around 2 μ M, a binding affinity substantially higher as compared to K_D values measured for the binding of phosphomimetic Cdc25C⁴⁴⁻²²⁸ to Pin1.

NMR titration of ¹⁵N-tri-phosphoCdc25C⁴⁴⁻²²⁸ by unlabelled Pin1

The enhanced interaction between the Cdc25C IDR and Pin1 subsequent to the phosphorylation of three key sites (Thr48, Thr130, and Ser168) prompted the hypothesis of direct binding between Pin1 and these phosphorylated residues. To validate this hypothesis and investigate binding preference, a reverse NMR titration experiment was performed by

recording a series of ^1H - ^{15}N BEST-TROSY spectra on a ^{15}N -labelled sample of triphospho-Cdc25C⁴⁴⁻²²⁸ in the absence of presence of up to 4 equivalents of unlabelled Pin1 (figure 114).

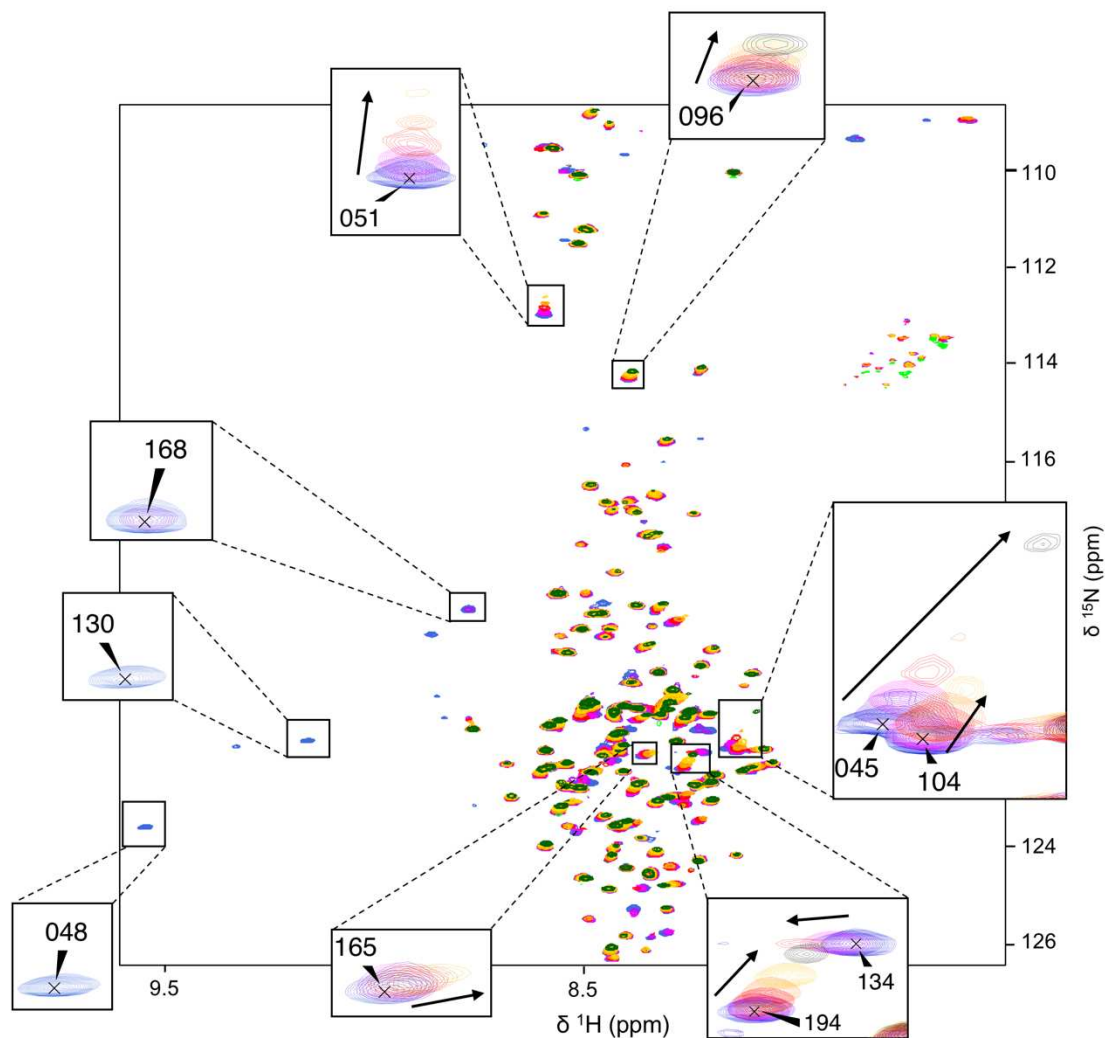


Figure 114: Overlay of ^1H - ^{15}N BEST-TROSY spectra of triphospho-Cdc25C⁴⁴⁻²²⁸ recorded throughout a titration with up to 4 molar equivalents of Pin1. The overlay displays spectra recorded in the presence of increasing concentrations of Pin1 (0, 0.1, 0.2, 0.5, 1.0, 2.0, 4.0) from blue to green. Direction of peak movement is indicated by arrows.

To elucidate the sequence of binding events, the data from the titration were quantitatively analysed at different Pin1:Cdc25C ratios, taking into account both chemical shift changes and line-broadening of individual resonances (Figure 115a-f). Four regions of the Cdc25C IDR displayed significant CSPs and substantial line broadening in the presence of 0.1 to 0.3 molar

equivalent of Pin1 (figure 115a-c). These regions correspond to residues surrounding the three phospho-sites pThr48, pThr130, and pSer168, as well as residues 104-110. In the region flanking pThr48, the signal nearly entirely disappeared in the presence of 0.1 molar eq. of Pin1 and became entirely undetectable at 0.2 molar eq. of Pin1. A similar trend was observed for residues surrounding pThr130, with full signal disappearance in the presence of 0.3 molar eq. of Pin1. Interestingly, for the pSer168 site and neighbouring residue Ile170, line broadening was less pronounced after the addition of 0.1 molar eq. of Pin1, with full signal disappearance only achieved after 0.5 molar eq. of Pin1. These data indicate the binding of Pin1 to the three pSer/pThr-Pro motifs of triphospho-Cdc25C⁴⁴⁻²²⁸, with a slight binding preference for the pThr48-Pro49 and pThr130-Pro131 motifs.

Since we previously showed that the WW domain acts as the preferential binding module for triphospho-Cdc25C⁴⁴⁻²²⁸, the severe line broadening observed in the regions flanking the three pSer/pThr-Pro motifs is likely due to the exchange of the WW domain between these three motifs in the complex, that would occur in the μ s-ms time scale. Indeed, the modest binding affinity of Pin1 for triphospho-Cdc25C⁴⁴⁻²²⁸, with a K_D value of approximately 2 μ M, is expected to result in a binding equilibrium in the fast to intermediate exchange regime, as illustrated by the perturbation of residue 45 in figure 114. However, the formation of a high-molecular weight complex combined to a decreased mobility of the interacting regions in the bound form may also contribute to signal disappearance when Pin1 is in excess.

The observation of both line broadening and significant CSPs for residues in the region ¹⁰⁴LQEVHLA¹¹⁰ is more unexpected as this region does not align with anticipated Pin1 binding sites due to the complete absence of serine, threonine and proline residues in this sequence. Hence, it is unlikely that this region of Cdc25C is directly involved in the interaction with Pin1. Interestingly, the perturbations observed in this region, in particular the severe line broadening of amide resonances, are concomitant with the binding of the pThr48-Pro49 and pSer168-Pro169 motifs. This observation suggests that the region ¹⁰⁴LQEVHLA¹¹⁰ of Cdc25C may be involved in intramolecular interactions with the pThr48-Pro49 or pSer168-Pro169 motif in the complex with Pin1.

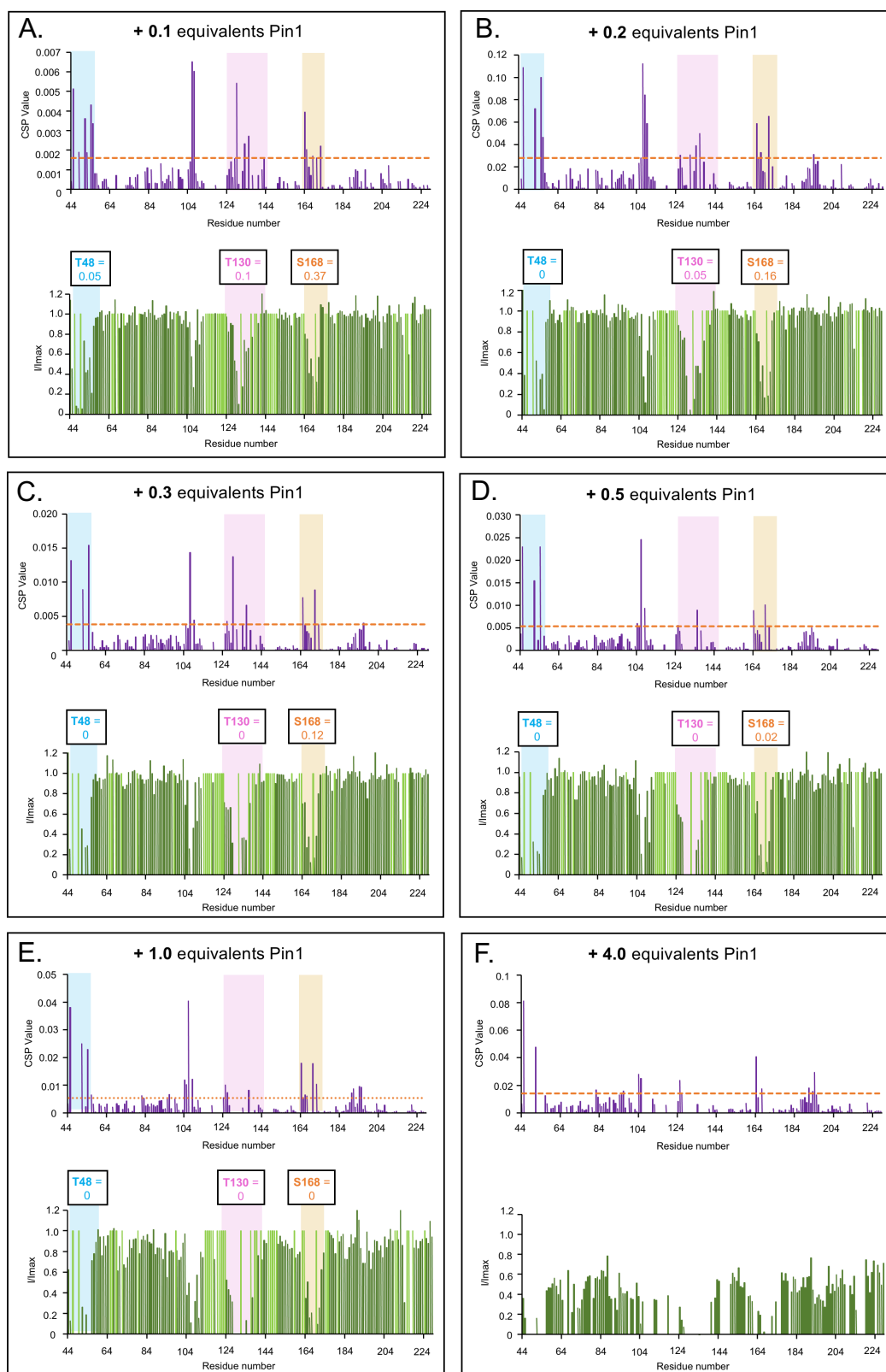


Figure 115: Analysis of CSPs and signal loss (I/I_{max}) observed for triphospho-Cdc25C⁴⁴⁻²²⁸ resonances throughout titration with increasing concentrations of Pin1 from 0.1 (A), 0.2 (B), 0.3 (C), 0.5 (D), 1.0 (E) and 4.0 equivalents (F). The orange dashed line in CSP charts represents the threshold value ($m + \sigma$). Residues which were unassigned in reference spectra are represented in light green with an I/I_{max} value of 1.

8.5 Discussion and perspectives

Taken together, the examination of the interaction between Pin1 and the three forms of Cdc25C discussed in this chapter revealed that a high binding affinity was only obtained upon phosphorylation of the Cdc25C⁴⁴⁻²²⁸ protein. Indeed, while very weak interactions with K_D values higher than 400 μM were observed with the unphosphorylated Cdc25C⁴⁴⁻²²⁸ and pM-Cdc25C⁴⁴⁻²²⁸ proteins, the binding affinity was substantially higher with triphospho-Cdc25C⁴⁴⁻²²⁸, with a K_D value around 2 μM . Unexpectedly, the results obtained with the phosphomimetic mutant of Cdc25C⁴⁴⁻²²⁸ did not align with those of the phosphorylated form of Cdc25C⁴⁴⁻²²⁸. When Pin1 interacted with pM-Cdc25C⁴⁴⁻²²⁸, none of the six mutated sites appeared to be significantly involved in binding. Instead, a region encompassing residues 190-200 exhibited the most pronounced binding effects, albeit with a relatively low affinity for Pin1 (K_D of ~ 400 μM). Notably, this region lacks serine, threonine, or proline residues but exhibited a slight helical propensity based on $\Delta C\alpha-C\beta$ values and SSP scores, and supported by ¹H-¹⁵N hNOE experiments which indicated this region displays a slightly restricted mobility in the ps-ns timescale. By contrast, upon Pin1 binding to tri-phospho Cdc25C⁴⁴⁻²²⁸, the involvement of this region in the interaction was negligible. Instead, the three phosphorylated motifs, pThr48-Pro49, pThr130-Pro131, and pSer168-Pro169, as well as a region spanning residues 104-110, displayed the most significant binding-induced changes. Strikingly, none of these regions exhibited substantial interaction with Pin1 in the phosphomimetic mutant. These results collectively emphasise that the use of glutamate as a mimic did not faithfully replicate the effect of phosphorylation in the case of the interaction between Cdc25C and Pin1. Furthermore, they highlight the necessity of substrate phosphorylation for interaction with Pin1.

In this work, although we could not produce the hyperphosphorylated mitotic form of Cdc25C, some initial insights can be hypothesised regarding the binding mode of Pin1 from our findings with the triphospho-Cdc25C⁴⁴⁻²²⁸ protein. In Chapter 3.4, we introduced several binding models that have been proposed to explain Pin1's interactions with substrates. Among these models, the simultaneous binding model suggests that both the WW and PPIase domains bind simultaneously to two distinct motifs within the same substrate. However, our data did not support a direct binding of the PPIase domain to the triphospho-Cdc25C⁴⁴⁻²²⁸ protein, the interaction being predominantly mediated by the WW domain. The simultaneous binding of

the WW and PPIase domains to two distinct sites is thus unlikely, especially considering the lack of observable chemical shift changes in residues of the phosphate-binding site of the PPIase domain, unlike the apparent strong interaction with the phosphate-binding loop of the WW domain. Our data also suggest that the catalysis-first model of Pin1 substrate binding, with the PPIase domain first catalysing isomerisation and the WW domain then binding to prevent the reverse reaction occurring, is unlikely as a stronger interaction with the PPIase domain would be also expected. Regarding the multimeric binding model, which involves interactions with an active kinase in a multiprotein complex, it remains beyond the scope of our study to evaluate.

Hence, among the four proposed models of Pin1-substrate interaction, the sequential binding model appears to be the most suitable for explaining the data discussed in this chapter. This model suggests that the WW domain acts as a targeting module, increasing the local concentration of the PPIase domain, enabling it to bind to other sites on the protein. Although our titration data at low concentrations of phospho-Cdc25C⁴⁴⁻²²⁸ indicate that the WW domain binds first, the severe line broadening of Pin1 resonances at stoichiometric ratios of phospho-Cdc25C⁴⁴⁻²²⁸ prevented us from determining whether the PPIase domain effectively binds at higher substrate concentrations. In addition, the apparent absence of a clear binding preference of Pin1 between the three phosphorylated motifs of phospho-Cdc25C⁴⁴⁻²²⁸ makes it more complex to propose a complete binding model for this system. It can be noted however, that the binding to Pin1 did not induce the folding of phospho-Cdc25C⁴⁴⁻²²⁸ interacting regions.

Since the NMR investigation of the interaction between FL Pin1 and the IDR of Cdc25C proved to be challenging, an alternative strategy based on the dissection of the interaction is likely to provide additional molecular insights. In particular, the use of shorter Cdc25C fragments may result in the formation of smaller complexes, which may address line broadening caused by potential large complex formation. Such studies could include competition experiments among Cdc25C phosphorylation sites to provide more detail about binding preferences, using both full-length Pin1 and its isolated domains. In this regard, the production and use of the hyperphosphorylated form of Cdc25C (with all six pSer/Thr-Pro motifs phosphorylated) would be highly relevant, not solely focussing on the tri-phosphoform.

In addition, the dynamic study of the free and bound form of Cdc25C⁴⁴⁻²²⁸ should be completed. For apo unphosphorylated WT Cdc25C⁴⁴⁻²²⁸, the extraction of R₁ and R₂ relaxation rates would provide valuable insights into the timescales of molecular motions, shedding more light on the intrinsic flexibility and conformational dynamics of the protein. Moreover, the investigation should extend to the bound phosphorylated form of Cdc25C⁴⁴⁻²²⁸ to characterise any changes in dynamics induced by binding to Pin1. Furthermore, to gain more insights into the conformational ensemble explored by the Cdc25C IDR in the free and bound forms, incorporating a paramagnetic probe at different positions in the Cdc25C IDR and measuring paramagnetic relaxation enhancements (PRE) experiments would provide valuable information on long-range interactions. In particular, such experiments would give additional insights on the involvement of the region ¹⁰⁴LQEVHLA¹¹⁰ of Cdc25C⁴⁴⁻²²⁸ in the interaction with Pin1 and potential intramolecular interactions with the phosphorylated motifs.

In the future, now that a phosphorylation protocol for the Cdc25C⁴⁴⁻²²⁸ construct has been established, there is potential to explore its phosphorylation by different kinases. An intriguing avenue of investigation involves characterising the phospho-form of Cdc25C IDR when phosphorylated solely by CDK1, in comparison to dual phosphorylation by CDK1 and Plk1. While previous studies have hinted at the significant impact of these scenarios on Pin1's regulatory effect on Cdc25C (Stukenberg & Kirschner, 2001), no research has elucidated the specific phosphorylation sites targeted by CDK1 and Plk1 in combination. Upon characterising the different phosphorylation status of Cdc25C⁴⁴⁻²²⁸ in these two cases, the binding of CDK1-phosphorylated Cdc25C⁴⁴⁻²²⁸ and CDK1/Plk1-phosphorylated Cdc25C⁴⁴⁻²²⁸ to Pin1 could be compared. Using NMR to study this interaction at the atomic level could reveal major insight into how Pin1 is able to activate CDK1/Plk1-phosphorylated Cdc25C⁴⁴⁻²²⁸ and inactivate CDK1-phosphorylated Cdc25C⁴⁴⁻²²⁸. Alternatively, due to the major disadvantage of sample loss observed upon *in vitro* phosphorylation and subsequent buffer exchange of Cdc25C⁴⁴⁻²²⁸, other phosphorylation methods could be explored. In particular, the use of amber codon suppression to produce highly controllable phosphorylated forms of Cdc25C⁴⁴⁻²²⁸ without the need for subsequent phosphorylation and buffer exchange, would also potentially enhance the yield of phosphorylated sample. This method should certainly be explored in future for the production of specific phosphorylated forms of the IDR of Cdc25C.



*MATERIALS &
METHODS*

Chapter 9 Materials and Methods

9.1 Reagents and Recipes

The composition of the buffers and media mentioned in the following sections are detailed below. LB media, SDS-PAGE gels and Running Buffer were prepared using distilled water. M9 media and all buffer solutions were prepared using MilliQ grade deionised water.

9.1.1. Bacterial Growth Media

Luria-Bertani agar (**LB agar**) was prepared as follows:

- ◆ 10 g/L peptone
- ◆ 5 g/L yeast extract
- ◆ 10 g/L NaCl
- ◆ 15 g/L agar

The pH was adjusted to 7.5 before sterilisation using an autoclave

Luria-Bertani (**LB**) media was prepared as follows:

- ◆ 10 g/L peptone
- ◆ 5 g/L yeast extract
- ◆ 10 g/L NaCl

The pH was adjusted to 7.5 before sterilisation using an autoclave

M9 minimal media was prepared as follows:

- ◆ 6.8 g/L Na₂HPO₄
- ◆ 3 g/L KH₂PO₄
- ◆ 0.5 g/L NaCl
- ◆ 0.1 mM CaCl₂
- ◆ 1 mM MgSO₄
- ◆ 100 µg/mL ampicillin
- ◆ 30 µg/mL chloramphenicol
- ◆ 10 µg/mL thiamine
- ◆ 1.7 g/L yeast nitrogen base

All components were pre-sterilised using an autoclave or 0.2 μ M filter and the M9 media was prepared in a Laminar flow cleanbench airflow.

9.1.2. Buffers and Solutions

Pin1 Binding Buffer was prepared as follows:

- ◆ 25 mM $\text{H}_2\text{NaO}_4\text{P}$
- ◆ 500 mM NaCl
- ◆ 30 mM imidazole
- ◆ 1 mM DTT

The pH was adjusted to 7.8 before the solution was filter-sterilised.

Pin1 Elution Buffer was prepared as follows:

- ◆ 25 mM $\text{H}_2\text{NaO}_4\text{P}$
- ◆ 500 mM NaCl
- ◆ 500 mM imidazole
- ◆ 1 mM DTT

The pH was adjusted to 7.8 before the solution was filter-sterilised.

Pin1 NMR Buffer was prepared as follows:

- ◆ 30 mM Tris
- ◆ 50 mM NaCl
- ◆ 2 mM DTT

The pH was adjusted to 7 before the solution was filter-sterilised.

Pin1 EXSY Buffer was prepared as follows:

- ◆ 50 mM Tris
- ◆ 50 mM NaCl
- ◆ 2 mM DTT

The pH was adjusted to 7 before the solution was filter-sterilised.

Cdc25C Binding Buffer was prepared as follows:

- ◆ 25 mM Tris
- ◆ 500 mM NaCl

- ◆ 1 mM EDTA
- ◆ 1 mM DTT
- ◆ 0.1% triton X-100
- ◆ 5% glycerol

The pH was adjusted to 8 before the solution was filter-sterilised.

Cdc25C Cleavage Buffer was prepared as follows:

- ◆ 25 mM Tris
- ◆ 150 mM NaCl
- ◆ 1 mM EDTA
- ◆ 2 mM DTT
- ◆ 5% glycerol

The pH was adjusted to 8 before the solution was filter-sterilised.

Cdc25C Elution Buffer was prepared as follows:

- ◆ 25 mM Tris
- ◆ 150 mM NaCl
- ◆ 1 mM EDTA
- ◆ 2 mM DTT
- ◆ 5% glycerol
- ◆ 10 mM glutathione

The pH was adjusted to 8 before the solution was filter-sterilised.

Cdc25C Storage Buffer was prepared as follows:

- ◆ 50 mM Tris
- ◆ 100 mM NaCl
- ◆ 2 mM DTT

The pH was adjusted to 7 before the solution was filter-sterilised.

Cdc25C NMR Buffer was prepared as follows:

- ◆ 40 mM Tris
- ◆ 100 mM NaCl
- ◆ 1 mM EDTA
- ◆ 2 mM DTT

The pH was adjusted to 7 before the solution was filter-sterilised.

Cdc25C Assignment Buffer was prepared as follows:

- ◆ 40 mM Tris
- ◆ 100 mM NaCl
- ◆ 1 mM EDTA
- ◆ 2 mM DTT

The pH was adjusted to 6.5 before the solution was filter-sterilised.

Phosphorylation Buffer (Phos-B) was prepared as follows:

- ◆ 50 mM Bis-Tris
- ◆ 100 mM NaCl
- ◆ 10 mM MgCl₂
- ◆ 4 mM ATP
- ◆ 4 mM DTT

The pH was adjusted to 7 before the solution was filter-sterilised.

PreScission Binding Buffer was prepared as follows:

- ◆ 20 mM H₂NaO₄P
- ◆ 500 mM NaCl
- ◆ 30 mM imidazole
- ◆ 1 mM DTT

The pH was adjusted to 7.5 before the solution was filter-sterilised.

PreScission Elution Buffer was prepared as follows:

- ◆ 20 mM H₂NaO₄P
- ◆ 500 mM NaCl
- ◆ 500 mM imidazole
- ◆ 1 mM DTT

The pH was adjusted to 7.5 before the solution was filter-sterilised.

SDS-PAGE Stacking Gel Buffer was prepared as follows:

- ◆ 0.5 M Tris

The pH was adjusted to 6.8 before the solution was filter-sterilised.

SDS-PAGE Resolving Gel Buffer was prepared as follows:

- ◆ 3 M Tris

The pH was adjusted to 8.8 before the solution was filter-sterilised.

SDS-PAGE Stacking Gel (4.3% BisAcrylamide) was prepared as follows:

- ◆ 1.88 mL Stacking Gel Buffer
- ◆ 1.13 mL 30% (w/v) BisAcrylamide
- ◆ 75 μ L 10% SDS

This solution was diluted to 7.75 mL before adding the following:

- ◆ 75 μ L 10% (w/v) Ammonium Persulphate (APS) (filter-sterilised)
- ◆ 18 μ L Tetramethylethylenediamine (TEMED)

The solution was gently inverted 6 times to thoroughly mix before pouring into gel apparatus to set.

SDS-PAGE Resolving Gel (15% BisAcrylamide) was prepared as follows:

- ◆ 2.92 mL Resolving Gel Buffer
- ◆ 11.66 mL 30% (w/v) BisAcrylamide
- ◆ 232 μ L 10% SDS

This solution was diluted to 23 mL before adding the following:

- ◆ 160 μ L 10% (w/v) APS (filter-sterilised)
- ◆ 24 μ L TEMED

The solution was gently inverted 6 times to thoroughly mix before pouring into gel apparatus to set.

4X SDS-PAGE Running Buffer was prepared as follows:

- ◆ 12.12 g/L Tris
- ◆ 57.6 g/L glycine
- ◆ 4 g/L SDS

2X SDS-PAGE Laemmli Buffer was prepared as follows:

- ◆ 4% SDS
- ◆ 20% glycerol
- ◆ 0.004% bromphenol blue
- ◆ 0.125 M Tris-Cl, pH 6.8

Stock solution is stored and 10% beta-mercaptoethanol is added before use

9.2 General Techniques

9.2.1. UV/Vis Absorbance Spectrophotometry

Measuring optical density at a wavelength of 600 nm (OD_{600}) was used to assess bacterial culture density using a UV/Vis spectrophotometer. 1 mL of sample is prepared in plastic cuvettes with a path length of 1 cm. Distilled water is used to obtain a reference absorption. If necessary, the culture is diluted before measurements are taken so that the OD_{600} value would be in the region of 0-0.1, ensuring the accuracy of readings.

The absorption of UV light at a wavelength of 280 nm (A_{280}) by aromatic amino acids was exploited to calculate the protein concentration of samples using a Thermo Scientific™ NanoDrop™ 2000 Spectrophotometer. As this instrument is so sensitive, only 2 μ L of buffered protein solution is applied to the pedestal after blanking with buffer solution. The A_{280} reading is then converted to a concentration (c) employing the Beer-Lambert law:

$$A = \epsilon cl$$

Whereby a path length (l) of 1 mm is used and the extinction coefficient (ϵ) is calculated for each protein sequence using the PROTPARAM ExPASy tool. The respective extinction coefficients for Pin1 and Cdc25C⁴⁴⁻²²⁸ are 20,870 and 2980 $M^{-1} cm^{-1}$, under the assumption that all cysteine residues are reduced. Given that the Cdc25C⁴⁴⁻²²⁸ sequence lacks tryptophan residues, which typically exhibits the highest molar absorption coefficient of the aromatic amino acids, concentrations estimated from A_{280} measurements carry an approximate 10% margin of error. Therefore, more repeats were performed to calculate a more accurate average concentration.

9.2.2. Centrifugation

A variety of centrifugation devices were used depending on the volume and nature of the sample. For large volumes of cell culture, the culture is transferred into 1 L centrifuge pots and centrifuged using a Beckman centrifuge fitted with a JLA 9.1000 rotor. For smaller volumes after cell lysis, 50 mL centrifuge pots are used and applied to a JA 25.5 fixed angle rotor. Samples under 1 mL are centrifuged using a benchtop Beckman centrifuge.

Protein concentration and buffer exchange of Cdc25C⁴⁴⁻²²⁸ samples are carried out using Amicon® Ultra-15 Centrifugal Filter Units. Centrifugal units with a 3 kDa MW cut-off are used to ensure that Cdc25C⁴⁴⁻²²⁸ (MW = 20.5 kDa) would not pass through the membrane. Samples are centrifuged at 5000 xg at 4 °C using a JLA 16.250 fixed-angle rotor fitted with 50 mL falcon adaptors. Buffer exchange is carried out by repeated steps of concentrating the sample and dilution into a new buffer until the old buffer comprises less than 0.1 % of the solution.

9.2.3. SDS-PAGE

To assess the purity of proteins throughout purification and to monitor the phosphorylation status of Cdc25C, Sodium Dodecyl Sulphate and Polyacrylamide Gel Electrophoresis (SDS-PAGE) was employed using Bio-Rad Mini-Protean II equipment. Gels are prepared in-house according to the recipes described in section 9.1.2. The 15% resolving gel is prepared first and poured into gel-forming apparatus before applying a layer of isopropanol to ensure a uniform separation between the two gels. After the resolving gel has set, the layer of isopropanol is removed before pouring the stacking gel. A comb is inserted immediately to form the sample wells. Samples are prepared for SDS-PAGE analysis by combining 10 µl of protein sample with 10 µl of 2X SDS-PAGE Loading Buffer before boiling to 95 °C for 5 minutes. 10 µl of the resulting solution is loaded onto the gel alongside a pre-stained PeqGold protein marker. The gel is run in SDS-PAGE Running Buffer at 100 V for 10 minutes to align the proteins at the start of the resolving gel and then 150 V for 45-60 minutes until the layer of Bromophenol Blue from the loading buffer has reached the bottom. After removal of the gels from the tank, they are washed with distilled water for 10 minutes under gentle agitation. Subsequently gels are stained using GelCode™ Blue Safe Protein Stain for 15 minutes under gentle agitation. De-staining is achieved overnight by incubating the gels in distilled water under gentle agitation.

9.3 Protein Expression

All recombinant proteins detailed in this manuscript were produced using an *E. coli* bacterial cell expression system.

9.3.1. Plasmid Design and Preparation

The selected expression vector for each protein described here varies due to their distinct purification requirements. A pET-15b plasmid containing the gene for full-length Pin1 was employed to produce a fusion protein comprising an *N*-terminal His6 tag separated from the Pin1 sequence by a PreScission cleavage site. Similarly, a pET-15b plasmid containing the gene for PreScission enzyme was provided. Concerning the production of the IDR of Cdc25C, the pGEX-6P-1 plasmid was selected as an expression vector for the production of an GST fusion protein. The *N*-terminal GST-tag is separated from the Cdc25C fragment by a PreScission cleavage site. All plasmids contain the gene for ampicillin resistance and are under the control of the T7 promoter allowing controllable induction of expression using IPTG. To produce more high-purity plasmid, a QIAprep Spin Miniprep Kit is used according to the manufacturers protocol.

9.3.2. Bacterial Transformation

A general heat-shock protocol was followed for transformation of *E. coli* cells with plasmid DNA. Firstly, BL21-Gold (DE3) PlysS competent cells are thawed on ice. 1 μ L of plasmid DNA is then incubated with a 50 μ L aliquot of cells on ice for 30 minutes. Subsequently, a 25 second incubation at 42°C is performed followed by immediate transfer to ice for 2 minutes. The cells are grown in 1 mL of LB media and shaken at 37°C for 1 hour prior to plating on LB agar plates containing 100 μ g/mL ampicillin. The plates are incubated at 37°C overnight. To prepare glycerol stocks of transformed *E. coli* cells, a single colony from antibiotic selection plates is inoculated in 5 mL of sterilised LB media containing 1 mg/mL ampicillin and 30 μ g/mL chloramphenicol. The culture is incubated at 37°C, shaking at 220 rpm until an OD₆₀₀ of 0.3-0.7 is reached. The cultures are stored at -80°C as 400 μ L aliquots containing 20% glycerol.

9.3.3. Bacterial Protein Expression

First, 50 mL of sterilised LB media is inoculated with 400 μ L glycerol stock of transformed BL21 (DE3) cells containing the gene for Pin1/Cdc25C/PreScission enzyme. This preculture is incubated overnight at 37°C, shaking at 220 rpm to reach an OD₆₀₀ of 4-8. 4 L of sterilised LB media is split into 500 mL aliquots in 2 L flasks and each aliquot is inoculated

with the preculture to reach an OD₆₀₀ of 0.05. 100 µg/mL ampicillin and 30 µg/mL chloramphenicol are also added to each 500 mL aliquot. The cultures are incubated at 37°C (for Pin1 and Cdc25C⁴⁴⁻²²⁸) or 30°C (PreScission enzyme) shaking at 220 rpm until an OD₆₀₀ of 0.6-0.8 is reached. At this stage, if isotopic labelling is not required, protein expression can be induced directly with the addition of 0.5 mM IPTG. If isotopic labelling *is* desired, the cultures are pooled and centrifuged at 3000 xg for 20 minutes at 20°C (Marley et al., 2001). The cell pellet is then resuspended in 1 L of sterilised M9 media and incubated at 37°C, shaking at 220 rpm, for 30 minutes. Carbon and nitrogen sources are then provided via 1g ¹⁵NH₄Cl or ¹⁴NH₄Cl and 4 g ¹²C₆H₁₂O₆ or ¹³C₆H₁₂O₆ (according to the desired labelling) and the culture is again incubated at 37°C for 30 minutes. Protein expression can then be induced with the addition of 0.5 mM IPTG. The cultures are incubated with IPTG for 4 hours at 37°C (Pin1) or 12 hours at 20°C (Cdc25C⁴⁴⁻²²⁸ and PreScission enzyme), shaking at 220 rpm. They are then pooled and centrifuged at 6000 xg for 20 minutes at 4°C. The cell pellet is resuspended in 80 mL of the appropriate binding buffer (Pin1, Cdc25C or PreScission Buffer A) and stored at -80°C until purification.

9.4 Protein Purification

9.4.1. Isolating the CFE

A frozen pellet of *E. coli* cells containing the overexpressed protein in 80 mL of Binding Buffer (prepared as described in section 9.1.2) is thawed at room temperature and then transferred to ice. The suspended cells are separated into 2 aliquots for lysis. To each 40 mL of suspended cells, 0.1% Triton X-100 is added along with the following cocktail of antiproteases: 0.1 mM PMSF, 1 µg/mL leupeptin and 1 µg/mL pepstatin. The 2 aliquots of cells are then lysed on ice using a sonicator whereby 20 cycles of 15s pulses at an amplitude of 40% are applied, with a 59s break between bursts. To isolate the CFE, the lysate is centrifuged at 30,000 xg for 25 minutes at 4°C. The soluble fraction is filtered using a 0.45 µm membrane prior to purification.

9.4.2. Purification of Pin1

Ni-NTA affinity chromatography is performed as the first step in the purification of Pin1. For this, an AKTA purifier machine, equipped with a 5 mL Ni-NTA His-trap column, is

run using the UNICORN program. The column is first equilibrated in 5 column volumes (CV) of Pin1 Binding Buffer at a flow rate of 1 mL/min. The filtered CFE is then injected into the column at 1 mL/min and the flow-through (FT) collected as 45 mL fractions. The column is subsequently washed with 10 CV of binding buffer to remove weakly interacting proteins. Finally, a gradient of 80 % Pin1 Elution Buffer (30-300 mM imidazole) is applied over 40 minutes and 3 mL fractions are collected as the His-tagged Pin1 is eluted from the column. The fractions containing Pin1 (identified by SDS-PAGE and chromatogram analysis) are pooled and incubated overnight with 0.15 mg of PreScission enzyme. After His-tag cleavage, the eluate is diluted 3-fold in binding buffer to decrease imidazole concentration before re-injection into the Ni-NTA column. The FT (containing fully-cleaved Pin1) is collected as 3 mL fractions and pooled.

A pressure-based concentration method is then used to obtain a 5 mL sample of Pin1 using an Amicon stirred ultrafiltration cell equipped with a 3 kDa ultrafiltration membrane. The final step in Pin1 purification relies on size exclusion chromatography (gel filtration). A HiLoad Superdex 75 pg column is equilibrated with the desired buffer at a flow of 1 mL/min. The sample is injected into the column as 2.5 mL aliquots and emerges from the column after washing with ~70 mL of buffer. The fractions containing pure Pin1 (identified by SDS-PAGE and chromatogram analysis) are pooled and concentrated using a stirred ultrafiltration cell prior to storage at as aliquots -80°C.

9.4.3. Purification of Cdc25C⁴⁴⁻²²⁸

For the purification of WT Cdc25C⁴⁴⁻²²⁸ and mutants, affinity chromatography is performed using a gravity-flow purification method. 10 mL of glutathione resin is added to a 1.5 x 30 cm glass Econo-Column® and equilibrated with 1 CV of Cdc25C Binding Buffer (prepared as described in section 9.1.2). The resin is then incubated with the filtered CFE (containing GST-Cdc25C) for 4 hours at 7°C under constant rotation to encourage binding of GST-Cdc25C to the glutathione beads. The resin/CFE mixture is then reapplied to the column and the FT collected. The resin is washed with 4 CV of Cdc25C Binding Buffer then equilibrated with 1 CV Cdc25C Cleavage Buffer (prepared as described in section 9.1.2). The resin is incubated overnight with 0.15 mg PreScission enzyme at 7°C under constant rotation. After GST-tag cleavage, the resin is reapplied to the column, allowed to set, then the FT is

collected. To recharge the resin, the beads are incubated with 1 CV of Cdc25C Elution Buffer, to elute GST bound to the resin. The resin is washed with 2 CV of filtered MilliQ and stored in 20% ethanol. The FT containing the Cdc25C IDR is concentrated using an Amicon® Ultra-15 centrifugal 3 KDa filter to reach a sample volume of 5 mL. Finally, 2.5 mL aliquots of the protein are passed through a HiLoad Superdex 75 pg gel filtration column using an AKTA purifier machine and the column washed with the desired buffer until the protein emerges after ~50 mL. The purity of the protein is monitored throughout the purification process by SDS-PAGE analysis.

9.4.4. Purification of PreScission Enzyme

PreScission enzyme is purified in a single affinity chromatography step. Passage of the His-tagged protein through a 5 mL Ni-NTA His-trap column is achieved using an AKTA purifier machine. The column is first equilibrated with PreScission Binding Buffer (prepared as described in section 9.1.2) before the filtered CFE is injected into the column at a flow rate of 1 mL/min. The column is washed with 10 CV of PreScission Binding Buffer. Finally, the enzyme is eluted using a gradient of 0-60% PreScission Elution Buffer (prepared as described in section 9.1.2) as 3 mL fractions are collected. The fractions are analysed on a 1.5 mm 15% Bis-Acrylamide SDS-PAGE gel and stored at -80°C as aliquots of 0.5 mg containing 20 % glycerol.

9.5 Mass Spectrometry

Mass spectrometry (MS) was performed using the MALDI-TOF technique to assess the quality of some recombinant protein and commercial peptide samples. The analyte is mixed with a 4-HCCA UV-MALDI matrix to a molar ratio of 1:10 (analyte:matrix) and applied to a sample plate. The plate is loaded into a MALDI-TOF instrument and an MS profile generated.

9.6 NMR Spectroscopy

9.6.1. Instruments Used for data Acquisition

All NMR experiments detailed in this manuscript were performed using either a 500 MHz Bruker spectrometer equipped with TCI ($^1\text{H}/^{13}\text{C}/^{15}\text{N}/^2\text{H}$) cryoprobe available at the Laboratoire des Biomolécules, or a 950 MHz Bruker Ascend III spectrometer equipped with a TCI ($^1\text{H}/^{13}\text{C}/^{15}\text{N}/^2\text{H}$) cryoprobe available through the Infranalytics network at the CNRS, Gif-Sur-Yvette.

9.6.2. 1D ^1H Spectroscopy

1D ^1H spectra were used throughout this thesis to assess the conditions of protein/peptide samples and to confirm ligand stock concentrations. The NMR experiment utilises the `zggpw5` pulse program with gradient-based Watergate W5 to implement a water suppression technique, resulting in improved spectral quality. For assessing the quality of protein samples, 1D ^1H experiments are recorded on samples of varying concentrations from 100-500 μM Pin1/Cdc25C⁴⁴⁻²²⁸. Similar experiments are also recorded on the pCdc25C peptide to assess the quality of the sample before and after running EXSY experiments. The use of 1D experiments to assess ligand stock concentration is described in section 9.6.6. The acquisition parameters of experiments conducted on protein samples, pCdc25C samples and samples used to calculate ligand stock solutions are summarised in table 20.

Sample	TD	NS	D1 (s)	AQ (s)	SW (ppm)	O1P (ppm)
Pin1	16384	32	1	1.365	16	4.7
Cdc25C ⁴⁴⁻²²⁸	8192	128	1	0.455	18	4.7
pCdc25C EXSY sample	16384	64	5	1.365	12	4.7
Ligand stock assessment	16384	128	20	1.036	12	4.7

Table 20: Acquisition parameters for 1D ^1H NMR experiments

9.6.3. 2D ^1H - ^1H Spectroscopy

For the assignment of peptide sequences pCdc25C, Ala-substrate and Glu-substrate, ^1H - ^1H Total Correlation Spectroscopy (TOCSY) and Rotating Frame Overhauser Enhancement Spectroscopy (ROESY) experiments were performed on a 500 MHz spectrometer. TOCSY experiments, run using the `dipsi2ph` pulse program, facilitate the elucidation of scalar couplings between protons across covalent bonds, thereby aiding in the identification of the spin systems of each residue. This is accomplished through cross-relaxation processes, transferring magnetisation between coupled nuclei. Conversely, ROESY experiments, ran using the `roesyph` pulse sequence, detect dipolar couplings between nuclei in spatial proximity, yielding insights into spatial arrangements of each spin system. These complementary techniques are used synergistically for the stepwise assignment of peptide sequences. The composition of the NMR samples and acquisition parameters for each experiment are detailed in table 21.

Sample composition		Experiment type	TD		NS	D1 (s)	AQ (s)		SW (ppm)		O1P (ppm)	
			F2	F1			F2	F1	F2	F1	F2	F1
10 mg/ml Ala/Glu-substrate 0.47 M LiCl 100% TFE- d_3		TOCSY	2048	512	16	1.5	0.170	0.043	12	12	4.7	4.7
		ROESY	2048	512	16	1.5	0.170	0.043	12	12	4.7	4.7
2.5 mM pCdc25C 10 % D $_2$ O 1 mM DSS MilliQ	2.5 mM pCdc25C 10 % D $_2$ O	TOCSY	8192	450	8	1.0	0.682	0.041	12	12	4.7	4.7
	60 μ M Pin1 Pin1 EXSY buffer	ROESY	4096	450	48	1.0	0.341	0.041	12	12	4.7	4.7

Table 21: Acquisition parameters for 2D ^1H NMR experiments

9.6.4. 2D ^1H - ^{13}C Spectroscopy

For the assignment of peptide sequences pCdc25C, Ala-substrate and Glu-substrate, 2D ^1H - ^{13}C Heteronuclear Single Quantum Coherence (HSQC) NMR experiments were conducted using the `hsqcetgp` pulse program. The composition of the NMR samples and acquisition parameters for each experiment are detailed in table 22.

Sample composition		TD		NS	D1 (s)	AQ (s)		SW (ppm)		OIP (ppm)	
		F2	F1			F2	F1	F2	F1	F2	F1
10 mg/ml Ala/Glu-substrate 0.47 M LiCl 100% TFE- d_3		1024	512	64	1	0.085	0.012	12	165.3	4.7	80
2.5 mM pCdc25C 10 % D ₂ O 1 mM DSS MilliQ	2.5 mM pCdc25C 10 % D ₂ O 60 μ M Pin1 Pin1 EXSY buffer	2048	220	64	1	0.128	0.011	16	80	4.7	40

Table 22: Acquisition parameters for 2D ^1H - ^{13}C NMR experiments

9.6.5. 2D ^1H - ^{15}N Spectroscopy

2D ^1H - ^{15}N HSQC NMR experiments were conducted using the `hsqcetfpf3gpsi` pulse program. These experiments are recorded on ^{15}N -labelled samples of Pin1 and the Cdc25C⁴⁴⁻²²⁸ variants using a 500 MHz spectrometer. Band-Selective Excitation Short Transient Transverse Relaxation Time (BEST-TROSY) experiments are recorded at 950 MHz, to attain higher sensitivity and minimised artifacts, resulting in enhanced spectral quality. These experiments are typically applied to ^{15}N -labelled samples of the Cdc25C⁴⁴⁻²²⁸ variants as higher sensitivity is required for data interpretation. The majority of samples are prepared in 5 mm NMR tubes where possible. As an exception, the BEST-TROSY spectra of WT Cdc25C⁴⁴⁻²²⁸ were recorded on a double-labelled ^{15}N - ^{13}C sample prepared in a 3 mm NMR tube due to issues with sample yields. For monitoring the time-course of phosphorylation of Cdc25C⁴⁴⁻²²⁸, SOFAST Heteronuclear Single Quantum Coherence (HMQC) experiments were performed using a 500 MHz spectrometer. Employing the `sfhmqcf3gpqh` pulse program, which features consecutive 120° and 180° soft pulses in the proton dimension, facilitates rapid data acquisition. The sample conditions and acquisition parameters for all 2D ^1H - ^{15}N spectra recorded are detailed in table 23.

	Sample composition	Experiment type	TD		NS	D1 (s)	AQ (s)		SW (ppm)		O1P (ppm)	
			F2	F1			F2	F1	F2	F1	F2	F1
Pin1	150 μ M Pin1 10 % D ₂ O Pin1 NMR buffer	HSQC	2048	140	4	1	0.128	0.071	16	19.5	4.7	116.75
	50 μ M Pin1 7.5 % D ₂ O Pin1 NMR buffer	BEST-TROSY	1612	120	24	0.2	0.071	0.028	12	22	4.7	118.75
	50 μ M Pin1 10 % D ₂ O Pin1 NMR buffer	SOFAST-HMQC	1024	140	48	0.2	0.064	0.07	16	19.5	4.7	117.5
WT Cdc25C ⁴⁴⁻²²⁸	150 μ M Cdc25C 10 % D ₂ O Cdc25C NMR/Assignment Buffer	HSQC	2048	400	4	1	0.128	0.188	16	21	4.7	117.5
	500 μ M Cdc25C 7.5 % D ₂ O 5 mM DTT Cdc25C Assignment Buffer	BEST-TROSY	1828	200	4	0.2	0.08	0.057	12	18	4.7	117.3
	100 μ M Cdc25C 10 % D ₂ O Phos-B buffer (+ ERK2)	SOFAST-HMQC	1024	400	8	0.2	0.064	0.189	16	21	4.7	117.5
Cdc25C ⁴⁴⁻²²⁸ -pM	210 μ M Cdc25C 10 % D ₂ O Cdc25C NMR buffer	HSQC	2048	300	8	1	0.128	0.148	16	20	4.7	117.0
	100 μ M Cdc25C 7.5 % D ₂ O Cdc25C NMR/Assignment Buffer	BEST-TROSY	1620	550	4	0.2	0.071	0.159	12	18	4.7	117.3
Phospho-Cdc25C ⁴⁴⁻²²⁸	100 μ M Cdc25C 7.5 % D ₂ O Cdc25C NMR/Assignment Buffer	HSQC	2048	400	8	1	0.128	0.188	16	21	4.7	117.5
	100 μ M Cdc25C 7.5 % D ₂ O Cdc25C NMR Buffer	BEST-TROSY	1612	550	4	0.2	0.071	0.159	12	18	4.7	117.3

Table 23: Acquisition parameters for 2D ¹H-¹⁵N NMR experiments

9.6.6. NMR titrations

NMR titrations were performed to study the interaction of Pin1 with the second-generation series of peptide ligands and Cdc25C⁴⁴⁻²²⁸ constructs. Similarly reverse titrations

were also performed on Cdc25C⁴⁴⁻²²⁸ constructs to study the effects of Pin1 binding. Interaction studies were performed at physiological pH to mimic the native cellular environment; however, experiments were recorded at a temperature of 298 K to due to the instability of recombinant protein samples at higher temperatures. For each point of a titration, ¹H-¹⁵N HSQC or BEST-TROSY spectra with identical parameters are recorded.

¹⁵N-Pin1 with second-generation ligands

An NMR sample containing 150 μM ¹⁵N-labelled Pin1 and 10% D₂O is prepared in Pin1 NMR buffer to a final volume of 500 μL in a 5 mm NMR tube. A reference HSQC spectrum is recorded on this sample using a 500 MHz spectrometer. Ligand stock solutions are prepared at 3 mM in Pin1 NMR buffer and the pH adjusted to 7 before the titration.

The concentration of the ligand stock solution is confirmed prior to the titration by integrating isolated resonances of assigned protons in 1D ¹H NMR spectra (see section 9.6.2) and comparing the calculated volume to that of a calibration peak corresponding to a known concentration of DSS. To do this, an NMR sample is prepared containing 100 μL of the 3 mM ligand stock, 10 % D₂O, and 1 mM DSS and subsequently a 1D ¹H NMR spectrum is recorded using a 500 MHz spectrometer. The DSS signal is used as a reference for relative quantification of the ligand stock solution by integrating known proton peaks in ligand spectra and comparing to the volume of the DSS peak, calibrated at 1 mM.

Ligand is titrated into the NMR sample to reach up to 8 molar equivalents of Pin1. The volumes added at each titration point are calculated as a molar ratio of Pin1. Typically, the following series of molar equivalents are titrated into the sample: 0.1, 0.2, 0.3, 0.5, 0.7, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 6.0, 8.0. ¹H-¹⁵N HSQC experiments are recorded on the sample at each titration point and compared to a reference spectrum of apo-Pin1 using a 500 MHz spectrometer.

¹⁵N-Pin1 with unlabelled Cdc25C⁴⁴⁻²²⁸ constructs

The titration of Pin1 with Cdc25C⁴⁴⁻²²⁸ constructs was performed in a similar way. The unlabelled sample of Cdc25C⁴⁴⁻²²⁸ protein is concentrated to 700-800 μM in Pin1 NMR buffer.

For titrations performed using a 500 MHz spectrometer, an NMR sample containing 150 μM ^{15}N -labelled Pin1 and 10% D_2O is prepared in Pin1 NMR buffer to a final volume of 500 μL in a 5 mm NMR tube. A reference ^1H - ^{15}N HSQC spectrum is recorded on this sample using a 500 MHz spectrometer. For titrations performed using a 950 MHz spectrometer, an NMR sample containing 50 μM ^{15}N -labelled Pin1 and 7.5% D_2O is prepared in Pin1 NMR buffer and a reference ^1H - ^{15}N BEST-TROSY spectrum. Cdc25C⁴⁴⁻²²⁸ is titrated into the NMR sample to reach up to 2-4 molar equivalents of Pin1. The volumes added at each titration point are calculated as a molar ratio of Pin1. Typically, the following series of molar equivalents are titrated into the sample: 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0, 1.5, 2.0, 3.0, 4.0.

Cross-titration

To perform the cross-titration whereby Pin1 concentration remains constant throughout the titration, two NMR tubes are prepared as follows:

- ◆ Tube 1: 100 μM ^{15}N -labelled Pin1 and 10% D_2O in Pin1 NMR buffer
- ◆ Tube 2: 100 μM ^{15}N -labelled Pin1, 400 μM unlabelled Cdc25C⁴⁴⁻²²⁸ and 10% D_2O in Pin1 NMR buffer

By mixing the contents of the two tubes, intermediate ratios of Pin1:Cdc25C (protein:ligand, P:L) can be obtained. The volume for each point of the cross-titration required to reach each ratio of P:L is detailed in table 24. A scheme explaining the process of the cross-titration is represented in figure 116. Notably, only the concentration of Cdc25C (L) changes throughout the titration.

Repeat	Volume cross-titrated (μl)	Ratio of P:L	
		Tube 1	Tube 2
1	50	1:0.5	1:5.5
2	60	1:1	1:5
3	70	1:1.5	1:4.5
4	100	1:2	1:4
5	150	1:2.5	1:3.5
6	300	1:3	1:3

Table 24: Volumes of tube 1 and 2 that are mixed to reach each desired equivalent of Cdc25C.

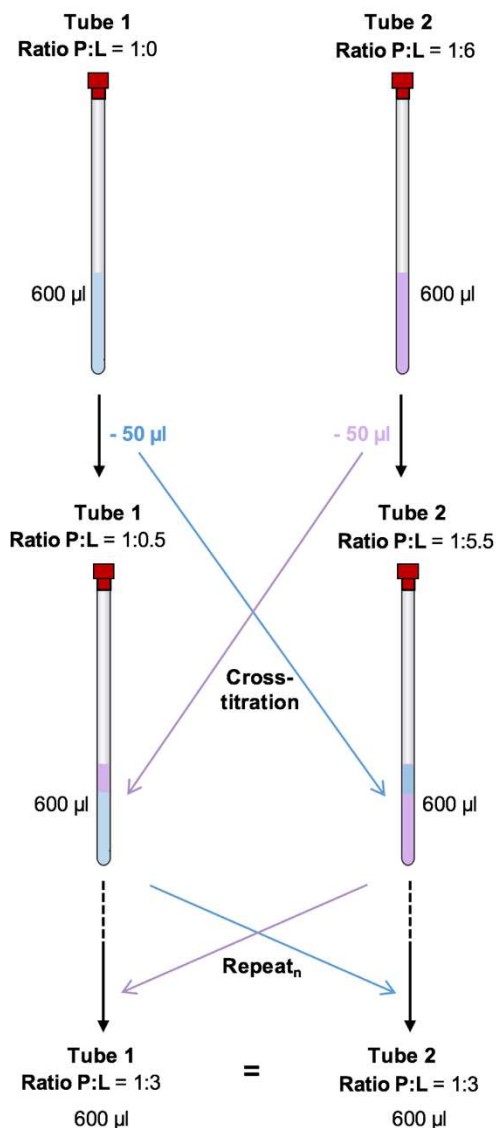


Figure 116: Scheme of the process of performing a cross-titration. The titration is repeated n number of times, by titrating the volumes calculated in table 5 until the ratio of Pin1:Cdc25C is equal in both tubes.

¹⁵N-Cdc25C⁴⁴⁻²²⁸ with unlabelled Pin1

The titration Cdc25C⁴⁴⁻²²⁸ constructs with Pin1 was performed using 950 MHz spectrometer. The unlabelled sample of Pin1 protein is concentrated to 700-800 µM in Cdc25C NMR buffer. A reference ¹H-¹⁵N BEST-TROSY spectrum is recorded on an NMR sample containing 100 µM ¹⁵N-labelled Cdc25C⁴⁴⁻²²⁸ and 7.5% D₂O in Cdc25C NMR buffer. Pin1 is titrated into the NMR sample to reach up to 2-4 molar equivalents of Cdc25C⁴⁴⁻²²⁸. The volumes added at each

titration point are calculated as a molar ratio of Cdc25C⁴⁴⁻²²⁸. Typically, the following series of molar equivalents are titrated into the sample: 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0.

Data-processing and analysis

The series of ¹H-¹⁵N 2D NMR experiments recorded at each titration point are processed using NMR Pipe. The spectra are analysed using the Sparky software to track the CSPs of each assigned backbone amide peak over the course of the titration. Lists containing the ¹H and ¹⁵N chemical shift values (and volumes) of each peak are extracted from Sparky and analysed in Microsoft excel. The following equation is used to calculate overall amide cross-peak CSP from the ¹H and ¹⁵N chemical shift values of each peak in the reference spectra (δ_f) compared to those of the spectra taken at the final titration point (δ_b):

$$\text{CSP} = \sqrt{\left(\frac{(\delta_b - \delta_f)_{^{15}\text{N}}}{10}\right)^2 + \left((\delta_b - \delta_f)_{^1\text{H}}\right)^2}$$

A CSP curve is established for each residue above the threshold value of $m + \sigma$ (where m is the average and σ is the standard deviation of the chemical shift change) by plotting CSP as a function of $[L]_t/[P]_t$. To calculate a K_D value for each residue, a Python algorithm is employed to fit the experimental data to the following equation:

$$\text{CSP} = \Delta\delta_{\text{max}} \times \frac{([P]_t + [L]_t + K_D) - \sqrt{([P]_t + [L]_t + K_D)^2 - 4[P]_t [L]_t}}{2 [P]_t}$$

To assess the individual peak volumes over the course of the titration for binding events in the fast to intermediate or intermediate exchange regime, amide cross-peak peak volumes are extracted from Sparky software. The relative peak intensity (I) at each titration point is

normalised by the intensity of this peak in the reference spectrum (I_{\max}) and plotted as (I/I_{\max}) against the protein sequence.

9.6.7. ^{15}N -(^1H) Heteronuclear NOE Spectroscopy

A ^1H -(^{15}N) NOE experiment is performed using the `hsqcnoef3gpsi` pulse program. The experiment was conducted at 298 K using the 500 MHz spectrometer equipped with a cryoprobe. A 350 μM sample of ^{15}N -labelled WT unphosphorylated Cdc25C⁴⁴⁻²²⁸ is prepared in Cdc25C storage buffer containing 7.5 % D_2O . The acquisition parameters for are detailed in table 25.

TD		NS	D1 (s)	AQ (s)		SW (ppm)		OIP (ppm)	
<i>F2</i>	<i>F1</i>			<i>F2</i>	<i>F1</i>	<i>F2</i>	<i>F1</i>	<i>F2</i>	<i>F1</i>
2048	700	24	4	0.128	0.397	16	17.4	4.7	117

Table 25: Acquisition parameters for ^1H -(^{15}N)-NOE NMR experiment

9.6.8. 3D backbone assignment experiments

A series of 3D experiments (HNCO, HNCA, HNCACB, HN(CA)CO, CBCA(CO)NH, HN(COCA)N and HN(CA)N) were recorded on double $^{13}\text{C}/^{15}\text{N}$ -labelled samples of the Cdc25C⁴⁴⁻²²⁸ variants to facilitate a backbone assignment. For the WT protein, experiments were performed at 950 MHz using a 3 mm NMR tube due to limitations in sample preparation. For the phosphomimetic mutant (Cdc25C⁴⁴⁻²²⁸-pM), the sample was prepared in a 5 mm NMR tube and experiments were recorded at 950 MHz. An assignment of the phospho-protein was performed at 500 MHz, with only HNCO and HNCACB experiments. All primary assignment experiments were performed at 283 K, pH 6.5. Additional assignment experiments were recorded on a double labelled sample of phospho-Cdc25C⁴⁴⁻²²⁸ at 298 K, pH 7 using a 950 MHz spectrometer to confirm the assignment under these conditions. The sample conditions and acquisition parameters for all 3D spectra recorded at 950 and 500 MHz are detailed in table 26. Spectra are processed using NMR Pipe and analysed using Sparky software.

Sample	Experiment type	TD			N S	D 1 (s)	AQ (s)			SW (ppm)			O1P (ppm)		
		F3	F2	F1			F3	F2	F1	F3	F2	F1	F3	F2	F1
500 μ M Cdc25C 7.5 % D ₂ O Cdc25C Assignment buffer 950 MHz	HNCO	182 4	80	18 0	4	0. 2	0.0 8	0.02 3	0.05 8	12	18	6.5	4. 7	117. 3	173
	HNCA	182 4	80	12 0	4	0. 2	0.0 8	0.02 3	0.01 2	12	18	21	4. 7	117. 3	51.5
	HNCACB	182 4	80	18 0	8	0. 2	0.0 8	0.02 3	0.00 6	12	18	61	4. 7	117. 3	41.5
	HN(CA)CO	182 4	80	18 0	8	0. 2	0.0 8	0.02 3	0.05 4	12	18	7	4. 7	117. 3	173
	CBCA(CO)NH	182 4	80	18 0	4	0. 2	0.0 8	0.02 3	0.00 6	12	18	61	4. 7	117. 3	41.5
	HN(COCA)N	182 4	80	80	16	0. 2	0.0 8	0.02 3	0.02 3	12	18	18	4. 7	117. 3	117. 3
	HN(CA)N	182 4	80	60	16	0. 2	0.0 8	0.02 3	0.01 7	12	18	18	4. 7	117. 3	117. 3
Cdc25C 7.5 % D ₂ O Cdc25C Assignment buffer 500 MHz	HNCO	161 6	80	18 0	4	0. 2	0.0 7	0.02 3	0.04 7	12	18	8	4. 7	117. 3	173
	HNCACB	161 6	80	16 0	4	0. 2	0.0 7	0.02 3	0.00 5	12	18	61. 5	4. 7	117. 3	41.5

Table 26: Acquisition parameters for 3D assignment NMR experiments

9.6.9. EXchange Spectroscopy (EXSY)

The isomerisation of Pin1 substrates (pCdc25C and pintide) is measured in the presence of Pin1 and ligands CZ115/CZ40. Experiments are conducted at 298 K using a 500 MHz spectrometer. Experiments are performed on a 2.5 mM sample of pCdc25C in Pin1 EXSY buffer, with 60 μ M Pin1, 10% D₂O, in a 5 mm NMR tube. To test the concentration dependence of this assay, additional experiments were performed in the presence of varying concentrations of Pin1 (5, 15, 30, and 45 μ M), in the absence of ligand. To test inhibitory activity of ligands, EXSY experiments are recorded in the presence of the following concentrations of CZ115 and CZ40: 0.25, 0.5, 0.75 and 1.25 mM. The EXSY spectra consist of a series of 2D ¹H-¹H NOESY spectra with acquisition parameters as summarised in table 27. NOESY spectra are recorded with the following series of mixing times: 0.5, 1, 1.5, 2, 3, 5, 7, 10, 15, 20, 30, 40, 50 and 70 ms.

Sample composition	TD		NS	D1 (s)	AQ (s)		SW (ppm)		OIP (ppm)	
	F2	F1			F2	F1	F2	F1	F2	F1
2.5 mM pCdc25C 60 μ M unlabelled Pin1 10 % D ₂ O Pin1 NMR buffer (+ CZ115/CZ40)	2048	256	4	5	0.17	0.021	12	12	4.7	4.7

Table 27: Acquisition parameters for EXSY NMR experiments

Spectra are processed and analysed using Topspin. Integration of cross-peaks corresponding to H^N pT5 is performed for spectra recorded at each mixing time. Diagonal peaks corresponding to H^N pT5 are integrated from the 2D ¹H-¹H NOESY spectra recorded at 0.5 ms, and the resulting volumes are used to normalise the intensities of their corresponding cross-peaks for the 14 mixing times. The resulting ratio is converted to a percentage of exchange cross-peak (normalised by diagonal peaks) at each mixing time:

$$\frac{\text{Summed volume of crosspeaks } (t_m = n)}{\text{Summed volume of diagonal peaks } (t_m = 0.5 \text{ ms})} \times 100$$

This percentage is plot against mixing time and the initial slope of the resulting graph is used as an estimate of Pin1 activity in the presence of difference concentrations of ligand (0.125, 0.25, 0.75 and 1.25 mM). To calculate the percentage of Pin1 activity in the presence of each ligand concentration, the calculated slope is divided by the slope obtained for experiments recorded in the absence of ligand (representative of 100% enzyme activity). A semilog plot of Pin1 activity as a function of ligand concentration is produced using GraphPad Prism software. An IC₅₀ value for each ligand is extracted after fitting the nonlinear regression with constant slope equations.

9.7 Pin1 UV activity assay

The UV activity assay was followed as described in (Janowski *et al*, 1997). Experiments were performed using a CD spectrometer. This machinery was selected as it is equipped with

a thermoelectric module for regulating the temperature. The reaction should be carried out at 10 °C to slow it down for more easy monitoring of the *cis* to *trans* isomerisation, particularly in the presence of a PPIase. The Pin1 substrate (Ala- or Glu-substrate) is first prepared as a 10 mg/mL stock solution in 0.47 M LiCl/TFE. To perform the solvent jump, 6.7 µL of stock solution is pipetted into a 200 µL cuvette containing 193.3 µL of 35 mM HEPES, pH 7. The mixture is pipetted up and down a few times to mix. A final substrate concentration of 0.3 mg/mL in the cuvette was found to be optimal, for preventing oversaturation ($OD > 1$). The same cuvette, lacking the substrate, is prepared and used as a blank.

To calculate the maximum absorption difference between the *cis* and *trans* forms of the Glu-substrate, a cuvette of '*cis*' substrate is prepared from 6.7 µL of substrate stock solution in a 200 µL cuvette containing 0.47 M LiCl/TFE and absorption measured at wavelengths of 250-400 nm. The same cuvette without substrate forms the blank. For the '*trans*' Glu-substrate absorption at wavelengths of 250-400 nm is measured 600s after a solvent jump. The relative absorption values are subtracted from each other to reveal at which wavelength is the highest difference in absorption between *cis* and *trans* forms.

For the Ala-substrate, absorption at wavelengths of 250-400 nm is monitored at 30 s and 600 s after a solvent jump. The relative absorption values of are subtracted from each other to reveal at which wavelength is the highest difference in absorption between *cis* and *trans* forms.

To measure the time course of *cis* to *trans* isomerisation, the solvent jump is performed. The cuvette is then immediately applied to the CD spectrometer. A dead time of ~25 seconds is anticipated. Absorption at 315 nm (Ala-substrate) or 300/330 nm (Glu-substrate) is monitored over the course of 20 minutes.

9.8 Phosphorylation Reactions

In vitro phosphorylation reactions are carried out using the ERK2 kinase which is stored in aliquots of 100 µM at -80 °C to prevent repeated freeze-thaw cycles which could affect activity. The phosphorylation protocol was adapted from (Julien *et al*, 2020).

For monitoring the phosphorylation reaction using SDS-PAGE analysis, 25 μM samples of unlabelled Cdc25C⁴⁴⁻²²⁸ are prepared in 25 μL of Phos-B buffer. The reaction begins with the addition of 250, 125 or 25 nM ERK2 (for 1:100, 1:200 and 1:1000 ratios of ERK2:Cdc25C⁴⁴⁻²²⁸). 4 μL of reaction mixture is taken after regular time intervals between 0 and 6 hours and directly mixed with 2x SDS-PAGE Laemmli Buffer and heated to 95 °C to stop the reaction. Samples are analysed by SDS-PAGE as detailed in section 9.2.3 using a 1.5 mm gel. For the second trial, and for the production of a gel for quantitative analysis, the reaction volume is doubled to 50 μL of 25 μM unlabelled Cdc25C⁴⁴⁻²²⁸ in Phos-B buffer. The reaction is run at room temperature in the presence of 250 nM ERK2 (ERK2:Cdc25C ratio of 1:100) for 5 hours. Aliquots were taken at 0, 10, 30, 60, 120 and 300 minutes and directly mixed with 2x SDS-PAGE Laemmli Buffer and heated to 95 °C to stop the reaction. Samples are analysed by SDS-PAGE as detailed in section 9.2.3 using a 0.75 mm gel. To estimate the ratio of unphosphorylated:phosphorylated protein after 5 hours, ImageJ software is used to calculate the relative intensity of bands corresponding to the phosphorylated and unphosphorylated form of Cdc25C⁴⁴⁻²²⁸.

For monitoring the phosphorylation reaction using NMR spectroscopy, an NMR sample is prepared containing 100 μM ¹⁵N-labelled WT Cdc25C⁴⁴⁻²²⁸ in 500 μL Phos-B buffer containing 10% D₂O. The sample is added to a 5 mm NMR tube and a reference ¹H-¹⁵N SOFAST HMQC spectrum recorded using a 500 MHz spectrometer. The samples removed from the NMR tube and the ERK2 kinase is added to the reaction mixture at a relative ratio of ERK2:Cdc25C of 1:100, 1:200, 1:1000 or 1:2000 and the sample is inserted back into the spectrometer. A lock and shim are performed before a series of ¹H-¹⁵N SOFAST HMQC experiments are recorded back-to-back for a duration of 5-20 hours. The resulting spectra are processed using NMRPipe and analysed in the following way. The peaks corresponding to unphosphorylated Thr48, and Ser168 are integrated in each recorded spectra using Sparky software. In addition, the unphosphorylated reporter peaks, corresponding to the neighbouring residues of each site (Gly50 and Gly167) are integrated using Sparky software in each recorded spectrum. The resulting volumes are summed for each site and plotted against time. The initial slope of the resulting plot is used to estimate maximum intensity (I_{max}) of the relevant peaks at $t=0$. Finally, a phosphorylation build-up curve is plotted as a function of time for each phosphorylation site:

$$\% \text{ phosphorylation } (t) = 1 - \frac{I(t)}{I_{\max}(t=0)}$$

For the Thr130 phosphorylation, build-up curves are prepared based on summed intensities of phosphopeaks corresponding to pThr130 and Asn132_(pThr130). The intensities are normalised using the summed maximal intensity value of both peaks at the end of the reaction (t=20hr) to calculate a percentage of phosphorylation at a given time:

$$\% \text{ phosphorylation } (t) = \frac{I(t)}{I_{\max}(t=20)}$$

The initial slope of the resulting graph is used to calculate the rate of phosphorylation at each site in units of h⁻¹. The curves were fit using the following equation:

$$\% \text{ phospho}(t) = \infty(1 - \exp(-R_{\text{phos}} \times t))$$

To prepare phosphorylated samples of unlabelled, ¹⁵N-labelled and ¹³C/¹⁵N-labelled Cdc25C⁴⁴⁻²²⁸, WT protein samples (in Cdc25C storage buffer) are first concentrated to 500 μM using an Amicon ultracentrifugation device with (MWCO = 3 kDa) (as described in section 9.2.2). Samples are diluted to 100 μM in Phos-B buffer. 1 μM of ERK2 kinase is added to the reaction mixture. The sample is mixed gently by pipetting up and down and left at room temperature for 6 hours. To stop the reaction, the sample is heated to 95 °C for 5 minutes using a heat block. To remove the kinase, the sample is centrifuged at 16,000 xg at 4 °C for 5 minutes. 1 μg/mL Leupeptin and Pepstatin are added to the supernatant before subjecting it to subsequent buffer exchange into appropriate buffer (e.g., Cdc25C assignment buffer for ¹³C/¹⁵N-labelled phospho-Cdc25C⁴⁴⁻²²⁸, Cdc25C NMR buffer for ¹⁵N-labelled Cdc25C⁴⁴⁻²²⁸ or Pin1 NMR buffer for unlabelled Cdc25C⁴⁴⁻²²⁸) and concentrated to the desired concentration for proceeding experiments using an Amicon ultrafiltration device (MWCO = 3 kDa) (as described in section 9.2.2). Alternatively to buffer exchange, the supernatant is injected directly as 2.5 ml aliquots into a HiLoad Superdex 75 pg gel filtration column, pre-equilibrated with the desired buffer, using an AKTA purifier machine. The column is washed with the desired

buffer until the protein emerges after ~45 mL. 1 µg/mL Leupeptin and Pepstatin are added to the sample before flash freezing and storage at -80°C.

Bibliography

- Ai Z, Zhang W, Luo W, Huang L & Pan Y (2009) Expression of Pin1 mRNA in Non-Small-Cell Lung Cancer Patients. *Asian Cardiovasc Thorac Ann* 17: 157–161
- Aplin AE, Gibb GM, Jacobsen JS, Gallo J-M & Anderton BH (2002) In Vitro Phosphorylation of the Cytoplasmic Domain of the Amyloid Precursor Protein by Glycogen Synthase Kinase-3 β . *Journal of Neurochemistry* 67: 699–707
- Arakawa T, Kita Y & Timasheff SN (2007) Protein precipitation and denaturation by dimethyl sulfoxide. *Biophysical Chemistry* 131: 62–70
- Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, Zhao C, Sonenberg N, Kay LE & Forman-Kay JD (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* 519: 106–109
- Bahassi EM, Hennigan RF, Myer DL & Stambrook PJ (2004) Cdc25C phosphorylation on serine 191 by Plk3 promotes its nuclear translocation. *Oncogene* 23: 2658–2663
- Bao L, Kimzey A, Sauter G, Sowadski JM, Lu KP & Wang D-G (2004) Prevalent Overexpression of Prolyl Isomerase Pin1 in Human Cancers. *The American Journal of Pathology* 164: 1727–1737
- Bayer E, Goetsch S, Mueller JW, Griewel B, Guiberman E, Mayr LM & Bayer P (2003) Structural Analysis of the Mitotic Regulator hPin1 in Solution. *Journal of Biological Chemistry* 278: 26183–26193
- Behrsin CD, Bailey ML, Bateman KS, Hamilton KS, Wahl LM, Brandl CJ, Shilton BH & Litchfield DW (2007) Functionally Important Residues in the Peptidyl-prolyl Isomerase Pin1 Revealed by Unigenic Evolution. *Journal of Molecular Biology* 365: 1143–1162
- Bianchi M & Manco M (2018) Pin1 Modulation in Physiological Status and Neurodegeneration. Any Contribution to the Pathogenesis of Type 3 Diabetes? *IJMS* 19: 2319
- Bondeson DP, Mares A, Smith IED, Ko E, Campos S, Miah AH, Mulholland KE, Routly N, Buckley DL, Gustafson JL, *et al* (2015) Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat Chem Biol* 11: 611–617
- Bonnet J, Mayonove P & Morris MC (2008) Differential phosphorylation of Cdc25C phosphatase in mitosis. *Biochemical and Biophysical Research Communications* 370: 483–488
- Born A, Henen MA & Vögeli B (2019) Activity and Affinity of Pin1 Variants. *Molecules* 25: 36
- Born A, Soetbeer J, Breitgoff F, Henen MA, Sgourakis N, Polyhach Y, Nichols PJ, Strotz D, Jeschke G & Vögeli B (2021) Reconstruction of Coupled Intra- and Interdomain Protein Motion from Nuclear and Electron Magnetic Resonance. *J Am Chem Soc* 143: 16055–16067
- Born A, Soetbeer J, Henen MA, Breitgoff F, Polyhach Y, Jeschke G & Vögeli B (2022) Ligand-specific conformational change drives interdomain allostery in Pin1. *Nat Commun* 13: 4546
- Bouchard JJ, Xia J, Case DA & Peng JW (2018) Enhanced sampling of interdomain motion using map-restrained Langevin Dynamics and NMR: Application to Pin1. *Journal of Molecular Biology* 430: 2164–2180
- Boutros R, Lobjois V & Ducommun B (2007) CDC25 phosphatases in cancer cells: key players? Good targets? *Nat Rev Cancer* 7: 495–507

- Brenner A, Reikvam H, Lavecchia A & Bruserud Ø (2014) Therapeutic Targeting the Cell Division Cycle 25 (CDC25) Phosphatases in Human Acute Myeloid Leukemia — The Possibility to Target Several Kinases through Inhibition of the Various CDC25 Isoforms. *Molecules* 19: 18414–18447
- Bulavin DV, Higashimoto Y, Demidenko ZN, Meek S, Graves P, Phillips C, Zhao H, Moody SA, Appella E, Piwnica-Worms H, *et al* (2003) Dual phosphorylation controls Cdc25 phosphatases and mitotic entry. *Nat Cell Biol* 5: 545–551
- Bulavin DV, Higashimoto Y, Popoff IJ, Gaarde WA, Basrur V, Potapova O, Appella E & Fornace AJ (2001) Initiation of a G2/M checkpoint after ultraviolet radiation requires p38 kinase. *Nature* 411: 102–107
- Castaño Z, Gordon-Weeks PR & Kypta RM (2010) The neuron-specific isoform of glycogen synthase kinase-3 β is required for axon growth. *Journal of Neurochemistry* 113: 117–130
- Chan DS-H, Kavanagh ME, McLean KJ, Munro Andrew W, Matak-Vinković D, Coyne AG & Abell C (2017) Effect of DMSO on Protein Structure and Interactions Assessed by Collision-Induced Dissociation and Unfolding. *Anal Chem* 89: 9976–9983
- Charlier C, Bouvignies G, Pelupessy P, Walrant A, Marquant R, Kozlov M, De Ioannes P, Bolik-Coulon N, Sagan S, Cortes P, *et al* (2017) Structure and Dynamics of an Intrinsically Disordered Protein Region That Partially Folds upon Binding by Chemical-Exchange NMR. *J Am Chem Soc* 139: 12219–12227
- Chaume G, Barbeau O, Lesot P & Brigaud T (2010) Synthesis of 2-Trifluoromethyl-1,3-oxazolidines as Hydrolytically Stable Pseudoprolines. *J Org Chem* 75: 4135–4145
- Chen Y, Wu Y, Yang H, Li X, Jie M, Hu C, Wu Y, Yang S & Yang Y (2018) Prolyl isomerase Pin1: a promoter of cancer and a target for therapy. *Cell Death Dis* 9: 883
- Cheng C-W, Leong K-W & Tse E (2016) Understanding the role of PIN1 in hepatocellular carcinoma. *WJG* 22: 9921
- Clair SSt & Manfredi JJ (2006) The Dual Specificity Phosphatase Cdc25C is a Direct Target for Transcriptional Repression by the Tumor Suppressor p53. *Cell Cycle* 5: 709–713
- Czechitzky W & Hamley P eds. (2016) Small molecule medicinal chemistry: strategies and technologies Hoboken, New Jersey: John Wiley & Sons, Inc
- Dalal SN, Schweitzer CM, Gan J & DeCaprio JA (1999) Cytoplasmic Localization of Human cdc25C during Interphase Requires an Intact 14-3-3 Binding Site. *Mol Cell Biol* 19: 4465–4479
- Di Martino O & Welch JS (2019) Retinoic Acid Receptors in Acute Myeloid Leukemia Therapy. *Cancers* 11: 1915
- Diehl JA, Zindy F & Sherr CJ (1997) Inhibition of cyclin D1 phosphorylation on threonine-286 prevents its rapid degradation via the ubiquitin-proteasome pathway. *Genes Dev* 11: 957–972
- Ding L, Cao J, Lin W, Chen H, Xiong X, Ao H, Yu M, Lin J & Cui Q (2020) The Roles of Cyclin-Dependent Kinases in Cell-Cycle Progression and Therapeutic Strategies in Human Breast Cancer. *IJMS* 21: 1960
- Dong L, Marakovits J, Hou X, Guo C, Greasley S, Dagostino E, Ferre R, Johnson MC, Kravynov E, Thomson J, *et al* (2010) Structure-based design of novel human Pin1 inhibitors (II). *Bioorganic & Medicinal Chemistry Letters* 20: 2210–2214
- Drakenberg T, Dahlqvist K-I & Forsen S The Barrier to Internal Rotation in Amides. IV. N,N-Dimethylamides; Substituent and Solvent Effects.

- Driver JA, Zhou XZ & Lu KP (2015) Pin1 dysregulation helps to explain the inverse association between cancer and Alzheimer's disease. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1850: 2069–2076
- Dumy P, Keller M, Ryan DE, Rohwedder B, Wöhr T & Mutter M (1997) Pseudo-Prolines as a Molecular Hinge: Reversible Induction of *cis* Amide Bonds into Peptide Backbones. *J Am Chem Soc* 119: 918–925
- Eberhardt ES, Loh SN, Hinck AP & Raines RT (1992) Solvent effects on the energetics of prolyl peptide bond isomerization. *J Am Chem Soc* 114: 5437–5439
- Fauman EB, Cogswell JP, Lovejoy B, Rocque WJ, Holmes W, Montana VG, Piwnica-Worms H, Rink MJ & Saper MA (1998) Crystal Structure of the Catalytic Domain of the Human Cell Cycle Control Phosphatase, Cdc25A. *Cell* 93: 617–625
- Felli IC & Pierattelli R (2014) Novel methods based on ¹³C detection to study intrinsically disordered proteins. *Journal of Magnetic Resonance* 241: 115–125
- Ferrari E, Bettuzzi S & Naponelli V (2022) The Potential of Epigallocatechin Gallate (EGCG) in Targeting Autophagy for Cancer Treatment: A Narrative Review. *IJMS* 23: 6075
- Fersht A (2017) Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding WORLD SCIENTIFIC
- Flaherty DB, Soria JP, Tomasiewicz HG & Wood JG (2000) Phosphorylation of human tau protein by microtubule-associated kinases: GSK3 β and cdk5 are key participants. *J Neurosci Res* 62: 463–472
- Forester CM, Maddox J, Louis JV, Goris J & Virshup DM (2007) Control of mitotic exit by PP2A regulation of Cdc25C and Cdk1. *Proc Natl Acad Sci USA* 104: 19867–19872
- Forood B, Feliciano EJ & Nambiar KP (1992) Stabilization of α -helical structures in short peptides via end capping. *Proc Natl Acad Sci* 90: 838–842
- Franckhauser C, Mamaeva D, Heron-Milhavet L, Fernandez A & Lamb NJC (2010) Distinct Pools of cdc25C Are Phosphorylated on Specific TP Sites and Differentially Localized in Human Mitotic Cells. *PLoS ONE* 5: e11798
- Gil S, Hošek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R & Felli IC (2013) NMR Spectroscopic Studies of Intrinsically Disordered Proteins at Near-Physiological Conditions. *Angew Chem Int Ed* 52: 11808–11812
- de Gooijer MC, van den Top A, Bockaj I, Beijnen JH, Würdinger T & van Tellingen O (2017) The G2 checkpoint—a node-based molecular switch. *FEBS Open Bio* 7: 439–455
- Greenwood AI, Rogals MJ, De S, Lu KP, Kovrigin EL & Nicholson LK (2011) Complete determination of the Pin1 catalytic domain thermodynamic cycle by NMR lineshape analysis. *J Biomol NMR* 51: 21–34
- Guo C, Hou X, Dong L, Dagostino E, Greasley S, Ferre R, Marakovits J, Johnson MC, Matthews D, Mroczkowski B, *et al* (2009) Structure-based design of novel human Pin1 inhibitors (I). *Bioorganic & Medicinal Chemistry Letters* 19: 5613–5616
- Guo J, Pang X & Zhou H-X (2015) Two Pathways Mediate Interdomain Allosteric Regulation in Pin1. *Structure* 23: 237–247
- Gutierrez GJ, Tsuji T, Cross JV, Davis RJ, Templeton DJ, Jiang W & Ronai ZA (2010) JNK-mediated Phosphorylation of Cdc25C Regulates Cell Cycle Entry and G2/M DNA Damage Checkpoint. *Journal of Biological Chemistry* 285: 14217–14228
- Hanahan D & Weinberg RA (2000) The Hallmarks of Cancer. *Cell* 100: 57–70

- Hanahan D & Weinberg RA (2011) Hallmarks of Cancer: The Next Generation. *Cell* 144: 646–674
- Hartwell LH & Weinert TA (1989) Checkpoints: Controls That Ensure the Order of Cell Cycle Events. *Science* 246: 629–634
- Hennig L, Christner C, Kipping M, Schelbert B, Rücknagel KP, Grabley S, Küllertz G & Fischer G (1998) Selective Inactivation of Parvulin-Like Peptidyl-Prolyl *cis/trans* Isomerases by Juglone. *Biochemistry* 37: 5953–5960
- Hoffmann I, Draetta G & Karsenti E (1994) Activation of the phosphatase activity of human cdc25A by a cdk2-cyclin E dependent phosphorylation at the G1/S transition. *The EMBO Journal* 13: 4302–4310
- Hutchins JRA & Clarke PR (2004) Many Fingers on the Mitotic Trigger: Post-Translational Regulation of the Cdc25C Phosphatase. *Cell Cycle* 3: 40–44
- Innes BT, Bailey ML, Brandl CJ, Shilton BH & Litchfield DW (2013) Non-catalytic participation of the Pin1 peptidyl-prolyl isomerase domain in target binding. *Frontiers in Physiology* 4: 18
- Inoue M, Sumii Y & Shibata N (2020) Contribution of Organofluorine Compounds to Pharmaceuticals. *ACS Omega* 5: 10633–10640
- Ito T, Ando H, Suzuki T, Ogura T, Hotta K, Imamura Y, Yamaguchi Y & Handa H (2010) Identification of a Primary Target of Thalidomide Teratogenicity. *Science* 327: 1345–1350
- Izumi T & Maller JL (1993) Elimination of cdc2 phosphorylation sites in the cdc25 phosphatase blocks initiation of M-phase. *MBoC* 4: 1337–1350
- Jacobs DM, Saxena K, Grimme S, Vogtherr M, Pescatore B, Langer T, Elshorst B & Fiebig KM (2002) Letter to the Editor: 1H, 13C and 15N backbone resonance assignment of the peptidyl-prolyl *cis-trans* isomerase Pin.
- Jacobs DM, Saxena K, Vogtherr M, Bernadó P, Pons M & Fiebig KM (2003) Peptide Binding Induces Large Scale Changes in Inter-domain Mobility in Human Pin1. *Journal of Biological Chemistry* 278: 26174–26182
- Janowski B, Wöllner S, Schutkowski M & Fischer G (1997) A Protease-Free Assay for Peptidyl Prolyl-*cis/trans* Isomerases Using Standard Peptide Substrates. *Analytical Biochemistry* 252: 299–307
- Jeener J, Meier BH, Bachmann P & Ernst RR (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. *The Journal of Chemical Physics* 71: 4546–4553
- Jeffries B, Wang Z, Felstead HR, Le Questel J-Y, Scott JS, Chiarparin E, Graton J & Linclau B (2020) Systematic Investigation of Lipophilicity Modulation by Aliphatic Fluorination Motifs. *J Med Chem* 63: 1002–1031
- Jin H, Jiang J, Sun L, Zheng F, Wu C, Peng L, Zhao Y & Wu X (2011) The prolyl isomerase Pin1 is overexpressed in human esophageal cancer. *Oncology Letters* 2: 1191–1196
- Julien M, Bouguechtouli C, Alik A, Ghouil R, Zinn-Justin S & Theillet F-X (2020) Multiple Site-Specific Phosphorylation of IDPs Monitored by NMR. In *Intrinsically Disordered Proteins: Methods and Protocols*, Kragelund BB & Skriver K (eds) pp 793–817. New York, NY: Springer US
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583–589
- Kaufmann H, Bailey JE & Fussenegger M (2001) Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Proteomics* 1: 194–199

- Keller M, Sager C, Dumy P, Schutkowski M, Fischer GS & Mutter M (1998) Enhancing the Proline Effect: Pseudo-Prolines for Tailoring *Cis* / *Trans* Isomerization. *J Am Chem Soc* 120: 2714–2720
- Kofron JL, Kuzmic P, Kishore V, Colon-Bonilla E & Rich DH (1991) Determination of kinetic constants for peptidyl prolyl cis-trans isomerases by an improved spectrophotometric assay. *Biochemistry* 30: 6127–6134
- Kosol S, Contreras-Martos S, Cedeño C & Tompa P (2013) Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* 18: 10802–10828
- Kumagai A & Dunphy WG (1996) Purification and Molecular Cloning of Plxl,a Cdc25-Regulatory Kinase from *Xenopus* Egg Extracts. 273
- Kunz C, Jahreis G, Günther R, Berger S, Fischer G & Hofmann H-J (2012) Influence of lithium cations on prolyl peptide bonds: INFLUENCE OF LITHIUM CATIONS. *J Pept Sci* 18: 400–404
- Lam PB, Burga LN, Wu BP, Hofstatter EW, Lu KP & Wulf GM (2008) Prolyl isomerase Pin1 is highly expressed in Her2-positive breast cancer and regulates erbB2 protein stability. *Mol Cancer* 7: 91
- Landrieu I, De Veylder L, Fruchart J-S, Odaert B, Casteels P, Portetelle D, Van Montagu M, Inzé D & Lippens G (2000) The Arabidopsis thaliana PIN1At Gene Encodes a Single-domain Phosphorylation-dependent Peptidyl Prolylcis/trans Isomerase. *Journal of Biological Chemistry* 275: 10577–10581
- Landrieu I, Smet C, Wieruszkeski J-, Sambo A, Wintjens R, Buee L & Lippens G (2006) Exploring the Molecular Function of PIN1 by Nuclear Magnetic Resonance. *CPPS* 7: 179–194
- Lane DP (1992) p53, guardian of the genome. *Nature* 358: 15–16
- Lee C-R, Park Y-H, Min H, Kim Y-R & Seok Y-J (2019) Determination of protein phosphorylation by polyacrylamide gel electrophoresis. *J Microbiol* 57: 93–100
- Lee YM & Liou Y-C (2018) Gears-In-Motion: The Interplay of WW and PPIase Domains in Pin1. *Front Oncol* 8: 469
- Lee YM, Teoh DE-J, Yeung K & Liou Y-C (2022) The kingdom of the prolyl-isomerase Pin1: The structural and functional convergence and divergence of Pin1. *Front Cell Dev Biol* 10: 956071
- Leung K-W, Tsai C-H, Hsiao M, Tseng C-J, Ger L-P, Lee K-H & Lu P-J (2009) Pin1 overexpression is associated with poor differentiation and survival in oral squamous cell carcinoma. *Oncol Rep* 21
- Li J, Mo C, Guo Y, Zhang B, Feng X, Si Q, Wu X, Zhao Z, Gong L, He D, *et al* (2021) Roles of peptidyl-prolyl isomerase Pin1 in disease pathogenesis. *Theranostics* 11: 3348–3358
- Lin J, Sahakian D, De Morais S, Xu J, Polzer R & Winter S (2003) The Role of Absorption, Distribution, Metabolism, Excretion and Toxicity in Drug Discovery. *CTMC* 3: 1125–1154
- Liou Y-C, Ryo A, Huang H-K, Lu P-J, Bronson R, Fujimori F, Uchida T, Hunter T & Lu KP (2002) Loss of Pin1 function in the mouse causes phenotypes resembling cyclin D1-null phenotypes. *Proc Natl Acad Sci USA* 99: 1335–1340
- Liou Y-C, Sun A, Ryo A, Zhou XZ, Yu Z-X, Huang H-K, Uchida T, Bronson R, Bing G, Li X, *et al* (2003) Role of the prolyl isomerase Pin1 in protecting against age-dependent neurodegeneration. 424
- Liou Y-C, Zhou XZ & Lu KP (2011) Prolyl isomerase Pin1 as a molecular switch to determine the fate of phosphoproteins. *Trends in Biochemical Sciences* 36: 501–514

- Liu K, Zheng M, Lu R, Du J, Zhao Q, Li Z, Li Y & Zhang S (2020) The role of CDC25C in cell cycle regulation and clinical cancer therapy: a systematic review. *Cancer Cell Int* 20: 213
- López-Avilés S, Grande M, González M, Helgesen A-L, Alemany V, Sanchez-Piris M, Bachs O, Millar JBA & Aligue R (2005) Inactivation of the Cdc25 Phosphatase by the Stress-Activated Srk1 Kinase in Fission Yeast. *Molecular Cell* 17: 49–59
- Lu KP, Finn G, Lee TH & Nicholson LK (2007) Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol* 3: 619–629
- Lu KP, Hanes SD & Hunter T (1996) A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* 380: 544–547
- Lu KP & Zhou XZ (2007) The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat Rev Mol Cell Biol* 8: 904–916
- Lu P-J, Wulf G, Zhou XZ, Davies P & Lu KP (1999a) The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein. *Nature* 399: 784–788
- Lu P-J, Zhou XZ, Shen M & Lu KP (1999b) Function of WW Domains as Phosphoserine- or Phosphothreonine-Binding Modules. *Science* 283: 1325–1328
- Lu Z & Hunter T (2014) Prolyl isomerase Pin1 in cancer. *Cell Res* 24: 1033–1049
- Luo J, Solimini NL & Elledge SJ (2009) Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. *Cell* 136: 823–837
- Ma SL, Pastorino L, Zhou XZ & Lu KP (2012) Prolyl Isomerase Pin1 Promotes Amyloid Precursor Protein (APP) Turnover by Inhibiting Glycogen Synthase Kinase-3 β (GSK3 β) Activity. *Journal of Biological Chemistry* 287: 6969–6973
- Malumbres M (2014) Cyclin-dependent kinases. *Genome Biology* 15
- Malumbres M & Barbacid M (2005) Mammalian cyclin-dependent kinases. *Trends in Biochemical Sciences* 30: 630–641
- Malumbres M, Harlow E, Hunt T, Hunter T, Lahti JM, Manning G, Morgan DO, Tsai L-H & Wolgemuth DJ (2009) Cyclin-dependent kinases: a family portrait. *Nat Cell Biol* 11: 1275–1276
- Margolis SS (2003) PP1 control of M phase entry exerted through 14-3-3-regulated Cdc25 dephosphorylation. *The EMBO Journal* 22: 5734–5745
- Margolis SS, Perry JA, Forester CM, Nutt LK, Guo Y, Jardim MJ, Thomenius MJ, Freel CD, Darbandi R, Ahn J-H, *et al* (2006) Role for the PP2A/B56 δ Phosphatase in Regulating 14-3-3 Release from Cdc25 to Control Mitosis. *Cell* 127: 759–773
- Marley J, Lu M & Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* 20: 71–75
- Marsh JA, Singh VK, Jia Z & Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between α - and γ -synuclein: Implications for fibrillation. *Protein Sci* 15: 2795–2804
- Matena A, Rehic E, Hönig D, Kamba B & Bayer P (2018) Structure and function of the human parvulins Pin1 and Par14/17. *Biological Chemistry* 399: 101–125
- Matthews HK, Bertoli C & De Bruin RAM (2022) Cell cycle control in cancer. *Nat Rev Mol Cell Biol* 23: 74–88

- McIntosh JR (2016) Mitosis. *Cold Spring Harb Perspect Biol* 8: a023218
- Mercedes-Camacho AY, Mullins AB, Mason MD, Xu GG, Mahoney BJ, Wang X, Peng JW & Etzkorn FA (2013) Kinetic Isotope Effects Support the Twisted Amide Mechanism of Pin1 Peptidyl-Prolyl Isomerase. *Biochemistry* 52: 7707–7713
- Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M & Forman-Kay JD (2010) Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure* 18: 494–506
- Motlagh HN, Wrabl JO, Li J & Hilser VJ (2014) The ensemble nature of allostery. *Nature* 508: 331–339
- Mullard A (2021) Targeted protein degraders crowd into the clinic. *Nat Rev Drug Discov* 20: 247–250
- Mutter M, Wöhr T, Gioria S & Keller M (1999) Pseudo-prolines: Induction of cis/trans-conformational interconversion by decreased transition state barriers. *Biopolymers* 51: 121–128
- Myer DL, Bahassi EM & Stambrook PJ (2005) The Plk3-Cdc25 circuit. *Oncogene* 24: 299–305
- Nakamura K, Greenwood A, Binder L, Bigio EH, Denial S, Nicholson L, Zhou XZ & Lu KP (2012) Proline Isomer-Specific Antibodies Reveal the Early Pathogenic Tau Conformation in Alzheimer's Disease. *Cell* 149: 232–244
- Namanja AT, Peng T, Zintsmaster JS, Elson AC, Shakour MG & Peng JW (2007) Substrate Recognition Reduces Side-Chain Flexibility for Conserved Hydrophobic Residues in Human Pin1. *Structure* 15: 313–327
- Namanja AT, Wang XJ, Xu B, Mercedes-Camacho AY, Wilson KA, Etzkorn FA & Peng JW (2011) Stereospecific gating of functional motions in Pin1. *Proc Natl Acad Sci USA* 108: 12289–12294
- Nasr R, Guillemain M-C, Ferhi O, Soilihi H, Peres L, Berthier C, Rousselot P, Robledo-Sarmiento M, Lallemand-Breitenbach V, Gourmel B, *et al* (2008) Eradication of acute promyelocytic leukemia-initiating cells through PML-RARA degradation. *Nat Med* 14: 1333–1342
- Nelson AL, Dhimolea E & Reichert JM (2010) Development trends for human monoclonal antibody therapeutics. *Nat Rev Drug Discov* 9: 767–774
- Newcombe EA, Delaforge E, Hartmann-Petersen R, Skriver K & Kragelund BB (2022) How phosphorylation impacts intrinsically disordered proteins and their function. *Essays in Biochemistry* 66: 901–913
- Norbury C & Nurse P (1992) Animal Cell Cycles and Their Control. *Annu Rev Biochem* 61: 441–470
- Nováček J, Židek L & Sklenář V (2014) Toward optimal-resolution NMR of intrinsically disordered proteins. *Journal of Magnetic Resonance* 241: 41–52
- Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztányi Z, Uversky VN, Obradovic Z, Kurgan L, *et al* (2012) D2P2: database of disordered protein predictions. *Nucleic Acids Research* 41: D508–D516
- Orlicky S, Tang X, Willems A, Tyers M & Sicheri F (2003) Structural Basis for Phosphodependent Substrate Selection and Orientation by the SCFCdc4 Ubiquitin Ligase. *Cell* 112: 243–256
- Otto T & Sicinski P (2017) Cell cycle proteins as promising targets in cancer therapy. *Nat Rev Cancer* 17: 93–115
- Panagopoulos A & Altmeyer M (2021) The Hammer and the Dance of Cell Cycle Control. *Trends in Biochemical Sciences* 46: 301–314

- Pardee AB (1974) A Restriction Point for Control of Normal Animal Cell Proliferation. *Proc Natl Acad Sci USA* 71: 1286–1290
- Park BK, Kitteringham NR & O'Neill PM (2001) Metabolism of Fluorine-Containing Drugs. *Annu Rev Pharmacol Toxicol* 41: 443–470
- Park H-S, Hohn MJ, Umehara T, Guo L-T, Osborne EM, Benner J, Noren CJ, Rinehart J & Söll D (2011) Expanding the Genetic Code of *Escherichia coli* with Phosphoserine. *Science* 333: 1151–1154
- Pastorino L, Sun A, Lu P-J, Zhou XZ, Balastik M, Finn G, Wulf G, Lim J, Li S-H, Li X, *et al* (2006) The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid- β production. *Nature* 440: 528–534
- Pawson T & Scott JD (2005) Protein phosphorylation in signaling – 50 years and counting. *Trends in Biochemical Sciences* 30: 286–290
- Pearlman SM, Serber Z & Ferrell JE (2011) A Mechanism for the Evolution of Phosphorylation Sites. *Cell* 147: 934–946
- Peng C-Y, Graves PR, Thoma RS, Wu Z, Shaw AS & Piwnica-Worms H (1997) Mitotic and G₂ Checkpoint Control: Regulation of 14-3-3 Protein Binding by Phosphorylation of Cdc25C on Serine-216. *Science* 277: 1501–1505
- Peng JW (2015) Investigating dynamic interdomain allostery in Pin1. *Biophys Rev* 7: 239–249
- Pennycook BR & Barr AR (2020) Restriction point regulation at the crossroads between quiescence and cell proliferation. *FEBS Lett* 594: 2046–2060
- Perry JA & Kornbluth S (2007) Cdc25 and Wee1: analogous opposites? *Cell Div* 2: 12
- Petsalaki E & Zachos G (2020) DNA damage response proteins regulating mitotic cell division: double agents preserving genome stability. *FEBS J* 287: 1700–1721
- Pines J (1995) Cyclins and Cyclin-Dependent Kinases: Theme and Variations. In *Advances in Cancer Research* pp 181–212. Elsevier
- Presta LG & Rose GD (1988) Helix Signals in Proteins. *Science* 240: 1632–1641
- Pu W, Zheng Y & Peng Y (2020) Prolyl Isomerase Pin1 in Human Cancer: Function, Mechanism, and Significance. *Front Cell Dev Biol* 8: 168
- Rajbhandari P, Finn G, Solodin NM, Singarapu KK, Sahu SC, Markley JL, Kadunc KJ, Ellison-Zelski SJ, Kariagina A, Haslam SZ, *et al* (2012) Regulation of Estrogen Receptor α N-Terminus Conformation and Function by Peptidyl Prolyl Isomerase Pin1. *Molecular and Cellular Biology* 32: 445–457
- Rajbhandari P, Ozers MS, Solodin NM, Warren CL & Alarid ET (2015) Peptidylprolyl Isomerase Pin1 Directly Enhances the DNA Binding Functions of Estrogen Receptor α . *Journal of Biological Chemistry* 290: 13749–13762
- Rajbhandari P, Schalper KA, Solodin NM, Ellison-Zelski SJ, Ping Lu K, Rimm DL & Alarid ET (2014) Pin1 modulates ER α levels in breast cancer through inhibition of phosphorylation-dependent ubiquitination and degradation. *Oncogene* 33: 1438–1447
- Ranganathan R, Lu KP, Hunter T & Noel JP (1997) Structural and Functional Analysis of the Mitotic Rotamase Pin1 Suggests Substrate Recognition Is Phosphorylation Dependent. *Cell* 89: 875–886

- Reynolds RA et al (1999) Crystal Structure of the Catalytic Subunit of Cdc25B Required for G2/M Phase Transition of the Cell Cycle. *293*: 559–568
- Rock KL, Gramm C, Rothstein L, Clark K, Stein R, Dick L, Hwang D & Goldberg AL (1994) Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* 78: 761–771
- Rudolph J (2007) Cdc25 Phosphatases: Structure, Specificity, and Mechanism. *Biochemistry* 46: 3595–3604
- Ruppenthal S, Noll A, Götz C & Montenarh M (2007) Interference between p53 and cdc25C in cell cycle regulation. *Int J Oncol*
- Rustighi A, Zannini A, Campaner E, Ciani Y, Piazza S & Del Sal G (2017) PIN1 in breast development and cancer: a clinical perspective. *Cell Death Differ* 24: 200–211
- Salwiczek M, Nyakatura EK, Gerling UIM, Ye S & Kokscha B (2012) Fluorinated amino acids: compatibility with native protein structures and effects on protein–protein interactions. *Chem Soc Rev* 41: 2135–2171
- Schwartz AL & Ciechanover A (2009) Targeting Proteins for Destruction by the Ubiquitin System: Implications for Human Pathobiology. *Annu Rev Pharmacol Toxicol* 49: 73–96
- Serpell LC (1999) Alzheimer's amyloid β fibrils: structure and assembly. *Biochimica et Biophysica Acta* 1502: 16–30
- Shah P & Westwell AD (2007) The role of fluorine in medicinal chemistry: Review Article. *Journal of Enzyme Inhibition and Medicinal Chemistry* 22: 527–540
- Shen M, Stukenberg PT, Kirschner MW & Lu KP (1998) The essential mitotic peptidyl-prolyl isomerase Pin1 binds and regulates mitosis-specific phosphoproteins. *Genes & Development* 12: 706–720
- Shi M, Chen L, Ji J, Cai Q, Yu Y, Liu B, Zhu Z & Zhang J (2015) Pin1 is Overexpressed and Correlates with Poor Prognosis in Gastric Cancer. *Cell Biochem Biophys* 71: 857–864
- Smerdon SJ & Yaffe MB (2010) Recognition of Phospho-Serine/Threonine Phosphorylated Proteins by Phospho-Serine/Threonine-Binding Domains. In *Handbook of Cell Signaling* pp 539–550. Elsevier
- Smet C, Wieruszkeski J-M, Buée L, Landrieu I & Lippens G (2005) Regulation of Pin1 peptidyl-prolyl *cis/trans* isomerase activity by its WW binding module on a multi-phosphorylated peptide of Tau protein. *FEBS Letters* 579: 4159–4164
- Strausfeld U, Fernandez A, Capony JP, Girard F, Lautredou N, Derancourt J, Labbe JC & Lamb NJ (1994) Activation of p34cdc2 protein kinase by microinjection of human cdc25C into mammalian cells. Requirement for prior phosphorylation of cdc25C by p34cdc2 on sites phosphorylated at mitosis. *Journal of Biological Chemistry* 269: 5989–6000
- Stukenberg PT & Kirschner MW (2001) Pin1 Acts Catalytically to Promote a Conformational Change in Cdc25. *Molecular Cell* 7: 1071–1083
- Sur S & Agrawal DK (2016) Phosphatases and kinases regulating CDC25 activity in the cell cycle: clinical implications of CDC25 overexpression and potential treatment strategies. *Mol Cell Biochem* 416: 33–46
- Taylor WR & Stark GR (2001) Regulation of the G2/M transition by p53. *Oncogene* 20: 1803–1815
- Thorsness PE & Koshland DE (1987) Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate. *Journal of Biological Chemistry* 262: 10422–10425

- Tompa P & Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends in Biochemical Sciences* 33: 2–8
- Toyoshima-Morimoto F, Taniguchi E & Nishida E (2002) Plk1 promotes nuclear translocation of human Cdc25C during prophase. *EMBO Rep* 3: 341–348
- Uchida T, Takamiya M, Takahashi M, Miyashita H, Ikeda H, Terada T, Matsuo Y, Shirouzu M, Yokoyama S, Fujimori F, *et al* (2003) Pin1 and Par14 Peptidyl Prolyl Isomerase Inhibitors Block Cell Proliferation. *Chemistry & Biology* 10: 15–24
- Urusova DV, Shim J-H, Kim DJ, Jung SK, Zykova TA, Carper A, Bode AM & Dong Z (2011) Epigallocatechin-gallate Suppresses Tumorigenesis by Directly Targeting Pin1. *Cancer Prevention Research* 4: 1366–1377
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, *et al* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50: D439–D444
- Verdecia MA, Bowman ME, Ping K, Hunter T & Noel JP (2000) Structural basis for phosphoserine-proline recognition by group IV WW domains. *nature structural biology* 7
- Vermeulen K, Van Bockstaele DR & Berneman ZN (2003) The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer: *Cell cycle regulation and deregulation*. *Cell Proliferation* 36: 131–149
- Wang R, He G, Nelman-Gonzalez M, Ashorn CL, Gallick GE, Stukenberg PT, Kirschner MW & Kuang J (2007) Regulation of Cdc25C by ERK-MAP Kinases during the G2/M Transition. *Cell* 128: 1119–1132
- Wang X, Mahoney BJ, Zhang M, Zintsmaster JS & Peng JW (2015) Negative Regulation of Peptidyl-Prolyl Isomerase Activity by Interdomain Contact in Human Pin1. *Structure* 23: 2224–2233
- Wang XJ, Xu B, Mullins AB, Neiler FK & Etzkorn FA (2004) Conformationally Locked Isostere of PhosphoSer–*cis*-Pro Inhibits Pin1 23-Fold Better than PhosphoSer–*trans*-Pro Isostere. *J Am Chem Soc* 126: 15533–15542
- Wei S, Kozono S, Kats L, Nechama M, Li W, Guarnerio J, Luo M, You M-H, Yao Y, Kondo A, *et al* (2015) Active Pin1 is a key target of all-trans retinoic acid in acute promyelocytic leukemia and breast cancer. *Nat Med* 21: 457–466
- Wildemann D, Erdmann F, Alvarez BH, Stoller G, Zhou XZ, Fanghänel J, Schutkowski M, Lu KP & Fischer G (2006) Nanomolar Inhibitors of the Peptidyl Prolyl Cis/Trans Isomerase Pin1 from Combinatorial Peptide Libraries. *J Med Chem* 49: 2147–2150
- Williams GH & Stoeber K (2012) The cell cycle and cancer. *J Pathol* 226: 352–364
- Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* 73: 1–16
- Wilson KA, Bouchard JJ & Peng JW (2013) Interdomain Interactions Support Interdomain Communication in Human Pin1. *Biochemistry* 52: 6968–6981
- Wishart David S & Sykes Brian D (1994) The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* 4
- Wu W, Xue X, Chen Y, Zheng N & Wang J (2022) Targeting prolyl isomerase Pin1 as a promising strategy to overcome resistance to cancer therapies. *Pharmacological Research* 184: 106456

- Wulf G, Garg P, Liou Y-C, Iglehart D & Lu KP (2004) Modeling breast cancer in vivo and ex vivo reveals an essential role of Pin1 in tumorigenesis. *EMBO J* 23: 3397–3407
- Wulf GM (2001) Pin1 is overexpressed in breast cancer and cooperates with Ras signaling in increasing the transcriptional activity of c-Jun towards cyclin D1. *The EMBO Journal* 20: 3459–3472
- Xu GG, Sledobnick C & Etkorn FA (2012) Cyclohexyl Ketone Inhibitors of Pin1 Dock in a Trans-Diaxial Cyclohexane Conformation. *PLoS ONE* 7: e44226
- Xu GG, Zhang Y, Mercedes-Camacho AY & Etkorn FA (2011) A reduced-amide inhibitor of Pin1 binds in a conformation resembling a twisted-amide transition state. *Biochemistry* 50: 9545–9550
- Xu N, Tochio N, Wang J, Tamari Y, Uewaki J, Utsunomiya-Tate N, Igarashi K, Shiraki T, Kobayashi N & Tate S (2014) The C113D Mutation in Human Pin1 Causes Allosteric Structural Changes in the Phosphate Binding Pocket of the PPIase Domain through the Tug of War in the Dual-Histidine Motif. *Biochemistry* 53: 5568–5578
- Yang X, Lee WH, Sobott F, Papagrigoriou E, Robinson CV, Grossmann JG, Sundström M, Doyle DA & Elkins JM (2006) Structural basis for protein–protein interactions in the 14-3-3 protein family. *Proc Natl Acad Sci USA* 103: 17237–17242
- Yu JH, Im CY & Min S-H (2020) Function of PIN1 in Cancer Development and Its Inhibitors as Cancer Therapeutics. *Front Cell Dev Biol* 8: 120
- Yu L-R & Veenstra TD (2021) Characterization of Phosphorylated Proteins Using Mass Spectrometry. *CPPS* 22: 148–157
- Zhang H, Ma Q, Zhang Y & Xu H (2012a) Proteolytic processing of Alzheimer’s β -amyloid precursor protein. *Journal of Neurochemistry* 120: 9–21
- Zhang M, Wang XJ, Chen X, Bowman ME, Luo Y, Noel JP, Ellington AD, Etkorn FA & Zhang Y (2012b) Structural and Kinetic Analysis of Prolyl-isomerization/Phosphorylation Cross-Talk in the CTD Code. *ACS Chem Biol* 7: 1462–1470
- Zhang MS, Brunner SF, Huguenin-Dezot N, Liang AD, Schmied WH, Rogerson DT & Chin JW (2017) Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat Methods* 14: 729–736
- Zhang Y, Daum S, Wildemann D, Zhou XZ, Verdecia MA, Bowman ME, Lücke C, Hunter T, Lu K-P, Fischer G, *et al* (2009) Structural Basis for High-Affinity Peptide Inhibition of Human Pin.
- Zhou XZ, Kops O, Werner A, Lu P-J, Shen M & Stoller G (2000) Pin1-Dependent Prolyl Isomerization Regulates Dephosphorylation of Cdc25C and Tau Proteins. *Molecular Cell* 6: 873–883
- Zhou XZ & Lu KP (2016) The isomerase PIN1 controls numerous cancer-driving pathways and is a unique drug target. *Nat Rev Cancer* 16: 463–478
- Zhu Z, Zhang H, Lang F, Liu G, Gao D, Li B & Liu Y (2016) Pin1 promotes prostate cancer cell proliferation and migration through activation of Wnt/ β -catenin signaling. *Clin Transl Oncol* 18: 792–797