Structural and Biochemical Characterisation of Phospho-Regulation in the DNA Damage Response

by Matthew James Stewart

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY



Department of Cancer and Genomic Sciences

College of Medical and Health Sciences

University of Birmingham

December 2024

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

The human phosphoproteome constitutes as many as 90,000 (pSer/pThr/pTyr) phosphorylation sites, however most phosphosites remain functionally uncharacterised. Phosphorylation events drive allosteric protein regulation, recruitment of phospho-binding domains, and assembly of multi-domain complexes that serve as a direct link between phosphorylation, ubiquitination, and SUMOylation pathways. The DNA damage response (DDR) is an essential signal transduction pathway that maintains genome stability and demonstrates the precision and efficiency of phospho-protein regulation. This thesis investigates the structural and biochemical basis of phospho-regulation of MDC1-TQxF motifs, Rad50-T690, and Sae1-S185 in DDR signalling. In the first instance, binding experiments and the crystal structure of the $Pellino_{FHA}$:$MDC1_{pT765}$ phosphopeptide complex supported a revised Pellino DNA damage recruitment model, which is dependent on MDC1-TQxF phosphorylation status, rather than γH2AX. We found no consistent evidence for a phosphorylation-dependent interaction between Rad50 and the PLK1 Polo Box domain, despite extensive biophysical binding experiments and attempts at complex purification. Despite this setback, site-specific incorporation of a phosphothreonine residue at Rad50-T690 into a recombinantly purified zinc-hook coiled-coil fragment was successful. This reagent represents a technical breakthrough for further structural and biochemical analysis of phospho-regulation of Rad50 and the MRN. Finally, we found no significant phospho-S185-dependent difference in SUMO E1 activity. Nonetheless, purification of recombinant Sae1 with site-specific incorporation of phosphoserine at S185, and the corresponding Sae1-pS185:Sae2 heterodimer, was successful. Both tools will provide useful for future investigations into alternate modes of phospho-regulation of the SUMO E1.

Dedicated in loving memory of my first teachers, Grandma Maureen and Granddad Terry.

Grandma, it may not be the red dress I promised you when I was little,

but I hope this work conveys the depth of the unspoken love I have for you both.

Acknowledgements

The past four years, culiminating in this thesis, have been the most challenging yet rewarding experiences of my life. It truly has been a rollercoaster, one that I often doubted I would ever complete. I would not been in this position, would it not be for the inspiration and support from the incredible scientists around me, as well as the love and laughter from my extended friends and family.

Firstly, I'd like to express my heartfelt thanks to my supervisor, Professor. Stephen Smerdon. Working with Steve has been the highlight of my scientific career. His immense knowledge of biochemistry and structural biology has been a constant source of inspiration, and I am deeply grateful to him for how he has pushed me to expand the boundaries of my own understanding. Beyond his academic guidance, I have always greatly appreciated his ability to support me personally, often through his hobby of making light of me and the many unique situations that I've found myself in along the way. Steve, thank you for everything- especially your infinite patience with me and my writing style. I would also like to extend this gratitude to my many scientific mentors, including Professor. Grant Stewart, Professor. Mat Coleman, Professor. Jo Morris, and Dr. Roger Grand, who have all been highly supportive and instrumental in my scientific development. To the Coleman, Tuxworth, and especially Morris lab members, thank you for always chatting with me and tolerating my annoying presence whenever I had a break between experiments. A special thanks goes to my main scientific co-conspirators: Dr Gareth 'Gazza' Hughes, Yara Aghabi, and Matt 'Big Matt' Mackintosh. Working with you has always been great source of fun and full of laughter. In particular, a special thank you to Yara for being my best friend throughout the MIBTP PhD. Yara, I'm grateful you didn't run away after our first meeting in Canon Hill Park! Thank you as well for always being there for me and for being the better friend. I'll always be there for you whenever you need me- especially if you've spilt something on your lab coat.

Thank you to the Pershore Road family, including Jakub Grzelak, Cristian-Silviu Rosca, and the boss of our house, Zaff. To Kuba and Criss, thank you for enduring my many scientific monologues and for the many evenings and bothy adventures filled with unforgettable shenanigans. To Zaff, cheers for the delightful curries and words of wisdom. I am also deeply grateful to my brothers from other mothers, including Anish 'Bhai' Sudra, Luke 'Lucas Piazón' Gray, Luke 'Scouse' Barnes, Sam Njau, Alfie 'ToTo' Chan, Tharshan 'T' Karunakaran, Roman 'Romano' Urban, Matt 'Matson' Batson and Charlie Bird, as well as Gosia Filipiak. Your friendship means the world to me.

Lastly, but most importantly, I am deeply grateful for all love and support from my extended family, which spans the Stewart, Threadgold, Defty, Brown, and Fortuny i Gómez clans, and includes the many cats and dogs. You all know who you are, and this thesis is long enough. Yes, I've finished writing, and no, I am not getting a 'real' job. Gràcies a Wizi, Cris i Ferran per acceptar-me com l'anglès pesat que sóc i per donar sempre suport a l'Anni i a mi. Most of all, I want to express my infinite love and appreciation to my Mum and partner in life, Anni - sempre junts, *Rinyo*. You two seem to be the only people that I listen to, and I am always amazed by your unconditional love and support, so thank you from the bottom of my heart.

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) and University of Birmingham funded Midlands Integrative Biosciences Training Partnership (MIBTP) BB/T00746X/1.

Declaration

I declare that this thesis is entirely my own work, with the exception of the following: (i) several sections of Chapter 3 are part of the following research article: Torres Esteban, M., **Stewart, M. J.**, Bragginton, E., Meroni, A., Pellizzari, A., Jeanrenaud, A., Smerdon, S. J., & Stucki, M. (2025). MDC1 mediates Pellino recruitment to sites of DNA double-strand breaks. *Life Science Alliance*, *8*(5), e202403074. https://doi.org/10.26508/lsa.202403074; (ii) the pGEX-6P-1-Peli1$_{13\text{-}273}$ and pGEX-6P-1-Peli2$_{15\text{-}275}$ plasmids were generated by Dr. Éilís Bragginton; (iii) the summary of ITC binding experiments conducted (Table 3.2) is collated from experiments performed by myself and Dr. Éilís Bragginton; (iv) the recombinant UBC9 and RanGAP 1$_{398\text{-}597}$ proteins were purified by Dr. Alex Lanz and Matthew Mackintosh.

All experimental work was conducted between April 2022 and June 2024 due to a change of supervision and research projects.

Table of Contents

Abstract ..... *ii*

Dedication ..... *iiv*

Acknowledgements ..... *iv*

Declaration ..... *vi*

List of Figures ..... *xiii*

List of Tables ..... *xvii*

List of Abbreviations ..... *xix*

Chapter 1 - General Introduction to Phospho-Regulation

1.1 Modular Cellular Signalling through Post-translational Modifications *2*

1.2 Protein Phosphorylation *3*

1.2.1 Protein Phosphorylation: The Original and Most Abundant Post-Translational Modification **3**

1.2.3 Protein Phosphorylation: Navigating Kinase Specificity and Noise **12**

1.2.4 Roles of Phosphorylation: True Signal in a World of Noise **13**

1.3 Phosphointeractomics: Insights and Challenges *20*

1.4 Aims and Objectives *22*

Chapter 2 - Materials and Methods

2.1 Molecular Biology Techniques *25*

2.1.1 Subculturing and Long-term Storage of *E.coli* Strains **25**

2.1.2 Bacterial Transformations **25**

2.1.2.1 Single-plasmid transformations using chemically competent bacterial strains **25**

2.1.2.2 Dual-plasmid transformations using non-competent bacterial strains **27**

2.1.3 Primer Design **28**

2.1.4 Standard PCR **29**

2.1.5 Site-directed Mutagenesis 30

2.1.6 Traditional Cloning 31

2.2 Protein Biochemistry Techniques 35

2.2.1 Recombinant Protein Overexpression and Purification 35

2.2.2 Purification of Fluorescent SUMO Proteins for Single Turnover *in vitro* SUMOylation Assays 39

2.2.3 Synthetic Peptides 41

2.2.4 1D SDS-PAGE Analysis 43

2.2.5 Phos-Tag SDS-PAGE Analysis 44

2.2.6 Western Blotting for Detection of Recombinant Phospho-proteins 45

2.2.7 Mass Spectrometry Analysis for Detection of Site-Directed Phosphoserine/ Threonine Incorporation into Recombinant Phospho-proteins 47

2.3 Interaction Studies 48

2.3.1 Analytical Size Exclusion Chromatography for Detection of Complex Formation 48

2.3.2 Amine-to-Amine Cross-Linking Analysis for Detection of Complex Formation 48

2.3.3 Microscale Thermophoresis Binding Analysis 50

2.3.4 Isothermal Titration Calorimetry Binding Analysis 53

2.4 Biochemical Assays 55

2.4.1 *In vitro* Casein Kinase II Assays and Purification of Casein Kinase II Phosphorylated Proteins 55

2.4.2 Fluorescent Single Turnover *In vitro* SUMOylation Assays 56

2.5 Structural Biology 58

2.5.1 X-ray Crystallographic Analysis of Pellino 1_{FHA}/$MDC1_{pT765}$ and PBD/$Rad50_{pT690}$ Complexes 58

2.5.1.1 Structure of the Human Pellino 1_{FHA}/$MDC1_{pT765}$ Phosphopeptide Complex 59

2.5.1.2 Structure of the Human PBD and Attempts at Crystallisation of a Potential Human PBD/Biotin-Rad50-pT690 Phosphopeptide Complex 60

2.5.2 Structural Model Predictions using AlphaFold 61

2.5.3 Visualisation and Analysis of Experimental and Predictive Structural Models 62

Chapter 3 - Structural and Biochemical Characterisation of MDC1 - pTQxF Mediated Recruitment of Pellino 1/2 to DNA Damage Sites

3.1 Introduction 64

3.1.1 Pellino E3 ligases 64

3.1.2 The Pellino FHA Domain and Substrate Specificity 66

3.1.3 Pellino Proteins in Innate Immunity Pathways 69

3.1.4 The Role of Pellino 1 in TLR3/4-specific NF-κB Proinflammatory Signalling 72

3.1.5 Pellino Proteins in IL-1R Signalling 74

3.1.6 DNA Damage Response signalling at double-stranded DNA breaks 76

3.1.7 MDC1 Acts as a Molecular Scafold in the DNA Damage Response 78

3.1.8 A Role for Peli Paralogues in the DDR 80

3.1.9 Limitations of the γH2AX-Dependent Pellino 1 Double-Stranded DNA Damage Recruitment Model 84

3.2 Preliminary Work Suggests an Alternative Pellino 1 DNA Double-Stranded Break Recruitment Model Involving the MDC1-TQxF Cluster 85

3.3 Aims and Objectives 90

3.4 Results 91

3.4.1 Recombinant Expression and Purification of Pellino 1/2-FHA and MDC1-tBRCT Domains 91

3.4.2 Pellino 1/2-FHA Domains Directly Bind to MDC1-pTQxF Phosphopeptides and Not a γH2AX Phosphopeptide 93

3.4.3 Crystal Structure of the Pellino 1-FHA:MDC1-pT765 Phosphopeptide Complex 98

3.4.4 AlphaFold Models Accurately Predict the Structure of the Pellino 1-FHA:MDC1-pT765 Complex 109

3.5 Discussion *119*

3.5.1 A revised model of Peli1/2 actvity in the DDR **119**

3.5.2 AlphaFold modelling of phospho-dependent Peli binding **122**

Chapter 4 - Structural and Biochemical Characterisation of a Potential Phosphorylation-Dependent Interaction Between PLK1 and Rad50

4.1 Introduction *125*

4.1.1 Rad50 and the Mre11/Rad50/Nbs1 (MRN) DNA repair complex **125**

4.1.2 Polo-like kinase **131**

4.1.3 PLK1: The Polo Box Domain **133**

4.1.4 PLK1: The Kinase Domain **136**

4.1.5 The PLK Family and PBD function **138**

4.1.6 PLK1 signalling the DDR **141**

4.2 Aims & Objectives *146*

4.3 Results *147*

4.3.1 PDB expression and purification **147**

4.3.2 Peptide synthesis **149**

4.3.3 Phosphopeptide Binding I – Microscale Thermophoresis **150**

4.3.4 Structural analysis of the PBD-Rad150 pT690 complex **153**

4.3.5 Interaction modelling **158**

4.3.6 Phosphopeptide Binding II – Isothermal Titration Calorimetry **162**

4.3.7 pT690 RAD50 Production by Amber Codon Suppression **165**

4.3.7.1 Detection 168

4.3.7.2 Rad50 pT690 Production 171

4.3.7.3 Validation/quantification of pT690 incorporation 175

4.3.8 ZCC Binding - MST **181**

4.3.9 ZCC Binding – Size-Exclusion Chromatography **186**

4.3.10 ZCC binding - Chemical cross-linking..... 190

4.3.11 ZCC Binding - Site-Specific Crosslinking Using Amber Codon Suppression..... 195

4.4 Discussion..... 198

4.4.1 Future experiments..... 201

Chapter 5 - Biochemical Characterisation of a SUMO E1 Phosphorylation Site Induced in Mitosis and the DNA Damage Response

5.1 Introduction..... 206

5.1.1 SUMO Isoforms..... 211

5.1.2 Selectivity Amongst Canonical SUMO1-3 Paralogues..... 213

5.1.3 Conserved Structural Features of the SUMO-Fold Drive SUMOylation Cascade Functionality while Unique SUMO Features Drive Paralogue Specificity..... 215

5.1.4 The Sae1:Sae2 SUMO Heterodimer is a Catalytic Hub for SUMO Activation and Transfer..... 219

5.1.5 The Structural Basis of Sae1:Sae2-mediated SUMO Processing..... 225

5.1.5.1 The Sae1 Inactive Adenylation Domain..... 226

5.1.5.2 Sae2..... 228

5.1.5.3 Sae2 Crossover and Re-entry Loops: Structural Hinges Required for SUMO Thioesterification234

5.1.5.4 The Sae2 Crossover Loop CxxC Motif..... 235

5.1.5.5 The Sae2 Catalytic Cysteine Domain..... 236

5.1.5.6 The Sae2 Ubiquitin Fold Domain Recruits Ubc9..... 237

5.2 Preliminary Work Suggests that Sae1-S185 Phosphorylation in the Unstructured loop is Induced by Multiple Stress-Inducing Stimuli..... 239

5.2.1 Sae1-S185 Phosphorylation is Induced in Response to Genotoxic Stress and Non-Phosphorylatable Sae1-S185A-Containing SUMO E1 Heterodimers Sensitises Cells to Genotoxic Stress..... 240

5.2.2 Sae1-S185 Phosphorylation is Induced by Mitotic Arrest, Enhanced by Genotoxic Stress in an Apparent DNA-PK, p38, ATM, and ATR-Dependent Manner, and is Associated with Mitotic SUMO 2/3 Chain Formation..... 242

5.2.3 *In vitro* SUMOylation Assays with Phosphomimetic Sae1-S185D-Containing SUMO E1 Heterodimers May Not Affect RanGAP1 SUMO Conjugation 243

5.2.4 Summary of Preliminary Data 244

5.2.4 Limitations of the Preliminary Data suggest a Lack of Importance of Sae1-S185 Phosphorylation in SUMO Activation 245

5.3 Aims and Objectives 246

5.4 Results 248

5.4.1 Recombinant Expression and Purification of Site-Specfically Phosphorylated Sae1-pS185 Protein 248

5.4.2 Purification of $Sae1_{pS185}$:Sae2 SUMO E1 Heterodimers 253

5.4.3 Recombinant Expression and Purification of Fluorescently-Labelled SUMO 1 and 2 Proteins 257

5.4.4 Fluorescent Single-Turnover *In Vitro* RanGAP1 SUMOylation Assays Suggest that Sae1-S185 Phosphorylation Does Not Impact SUMO E1 Activity 258

5.5 Discussion 263

5.5.1 Future Experiments 264

Chapter 6 - Closing Remarks

Closing Remarks 267

List of References 272

Appendix I – Primers 305

Appendix II – Expression Vectors 306

Appendix III – Cloned Plasmids 307

Appendix IV – Recombinant Proteins 309

Appendix V – PBD Crystal Structure Attempts 311

Appendix VI – AlphaFold Primary Sequence 312

List of Figures

Chapter 1

Figure 1.1 An overview of the modular protein phosphorylation signalling system. 4

Figure 1.2 Canonical protein phosphorylation. 6

Figure 1.3 Non-canonical protein phosphorylation. 8

Figure 1.4 The Hanks-type kinase bilobal fold. 10

Figure 1.5 The human kinome. 11

Figure 1.6 Protein phosphorylation priming drives allosteric protein kinase activation. 15

Figure 1.7 Phospho-binding module families recognise and bind to specific and distinct phospho-epitopes. 18

Figure 1.8 Phospho-binding modules are fundamental for multi-domain complex formation and PTM cross-talk within the DNA damage response. 20

Chapter 2

Figure 2.1 Skeletal Structures of BS^3 and BS(PEG)9 homobifunctional amine-reactive crosslinking agents. 49

Chapter 3

Figure 3.1 The human Pellino family domain architectures. 65

Figure 3.2 Primary sequence alignment of the Pellino family. 66

Figure 3.3 The 1.8Å crystal structure of the Pellino 2FHA domain. 67

Figure 3.4 Pellino 1/2 family members in TLR and IL-1R signalling cascades. 73

Figure 3.5 The domain architecture of MDC1 and phosphosite clusters. 78

Figure 3.6 Model of γH2AX-dependent Pellino 1 recruitment to DNA damage sites and Pellino 1-mediated DNA damage signalling. 81

Figure 3.7 Pellino and RNF8 E3 ligases share a common domain architecture. 83

Figure 3.8 MDC1-pTQxF peptide-pulldowns coupled with tandem mass spectrometry (LC/MS) proteomics identifies Pellino 1/2 as novel interactors of MDC1-pTQxF motifs. 86

Figure 3.9 Pellino 1 and 2 interact with MDC1-pTQxF phosphopeptides and not a γH2AX phosphopeptide. 87

Figure 3.10 Pellino 1/2 and MDC1 interact in cells in a DNA damage-dependent manner. 88

Figure 3.11 Pellino 1 Recruitment to DSB sites is dependent on the phosphorylation status of MDC1-TQxF sites. 89

Figure 3.12 Expression and purification of recombinant Pellino 1/2FHA and MDC1tBRCT domains. 92

Figure 3.13 ITC binding experiments demonstrate that Pellino 1/2FHA domains directly bind to MDC1-pTQxF phosphopeptides. 94

Figure 3.14 ITC binding experiments demonstrate that Pellino 1/2FHA domains do not directly bind to a γH2AX phosphopeptide. 95

Figure 3.15 ITC binding experiments demonstrate that MDC1tBRCT domains directly bind to the γH2AX phosphopeptide, but Pellino 1/2 domains do not. 96

Figure 3.16 Crystallisation trials of the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex. 99

Figure 3.17 The 2.55Å crystal structure of the Pellino 1FHA:MDC1pT765 phosphopeptide complex. 101

Figure 3.18 The $MDC1_{pT765}$ phosphopeptide binds to the Pellino 1 FHA core basic pocket. 103

Figure 3.19 Pellino 1_{FHA} residues R104, S135, and R136 directly interact with the pThr residue in the $MDC1_{pT765}$ phosphopeptide. 104

Figure 3.20 Pellino 1FHA pThr-contacting residues are structurally conserved throughout the Pellino and FHA families. 105

Figure 3.21 A schematic representation of all interactions between Pellino 1_{FHA} and the $MDC1_{pT765}$ phosphopeptide. 106

Figure 3.22 The Pellino family 'wing' insert contributes indirectly to phosphopeptide pThr +3 selectivity. 107

Figure 3.23 Structural alignments of Pellino 1_{FHA}:$MDC1_{pT765}$ and $RNF8_{FHA}$:RNF8tide complexes reveal differences in pThr +3 binding selectivity. 108

Figure 3.24 The Pellino 1FHA γ-methyl pocket involved in pThr selectivity. 109

Figure 3.25 AlphaFold 2 Multimer (AF-M) metric outputs for the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex structural models. 112

Figure 3.26 AlphaFold 2 Multimer (AF-M) and 3 (AF3) models accurately predict the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex structure. 114

Figure 3.27 LGA structural analysis of AF models compared to the crystal structure indicate that AF-M performs slightly better. 116

Figure 3.28 An updated model of ATM-dependent and MDC1-pTQxF-dependent Pellino 1 recruitment to DNA damage sites and Pellino 1-mediated DNA damage signalling. 120

Chapter 4

Figure 4.1 The MRN complex. 129

Figure 4.2 The Rad50 Zinc hook coiled-coil domain (ZCC) and T690 site are highly conserved. 131

Figure 4.3 The PLK1 Polo Box Domain. 135

Figure 4.4 PLK1-mediated phosphor-regulation of DNA damage signalling throughout the cell cycle. 145

Figure 4.5 Expression and purification of wild-type and phospho-binding deficient recombinant PLK1 PBD domains. 148

Figure 4.6 MST binding analysis indicates that the recombinant PBD protein functions as a phospho-binding module. 151

Figure 4.7	MST binding analysis suggests that the recombinant PBD protein binds to the biotin-Rad50-pT690 phosphopeptide.	153
Figure 4.8	The 2.58 Å crystal structure of unbound PBD.	155
Figure 4.9	AlphaFold 2 Multimer (AF-M) metric outputs for the PBD:$Rad50_{E690}$ phosphopeptide complex structural models.	160
Figure 4.10	AlphaFold 2 Multimer (AF-M) and 3 (AF3) models predict an interaction between PBD and the Rad50-pT/E690 peptide.	161
Figure 4.11	Structural alignment of AlphaFold 2 Multimer (AF-M) and 3 (AF3) models indicate alternate predicted Rad50 peptide binding configurations and differential PBD binding sites.	162
Figure 4.12	ITC binding experiments demonstrate that the PBD directly binds to the PBDtide control phosphopeptide, but no detectable binding was detected with biotin-$Rad50_{pT690}$ and $Rad50_{pT690}$-AYKK phosphopeptides.	164
Figure 4.13	pSer/Thr orthogonal translation systems (OTSs).	167
Figure 4.14	In vitro kinase assays suggest that the recombinant CK2 catalytic subunit (CK2α) phosphorylates the Rad50-ZCC fragment independently of the presence of an intact T690 site.	170
Figure 4.15	In vitro kinase assays coupled with mass spectrometry demonstrate that the anti-CK2α antibody is a suitable biomarker for recombinant Rad50-T690 phosphorylation.	171
Figure 4.16	Optimisation of OTS-dependent site-specific incorporation of pThr into recombinant Rad50-690 protein.	174
Figure 4.17	Tandem mass spectrometry (MS/MS) validation of site-specific phosphothreonine incorporation at Rad50-T690 in recombinant MBP-3C-Rad50-pT690-His6 protein.	176
Figure 4.18	Tandem mass spectrometry (MS/MS) validation of site-specific threonine misincorporation at Rad50-T690 in recombinant MBP-3C-Rad50-pT690-His6 protein.	177
Figure 4.19	Tandem mass spectrometry (MS/MS) validation of site-specific phosphoserine misincorporation at Rad50-T690 in recombinant MBP-3C-Rad50-pT690-His6 protein.	178
Figure 4.20	Tandem mass spectrometry (MS/MS) validation of site-specific glycine misincorporation at Rad50-T690 in recombinant MBP-3C-Rad50-pT690-His6 protein.	179
Figure 4.21	Phos-tag SDS-PAGE analysis suggests that ~48-67% of the recombinant MBP-3C-Rad50-pT690-His6 protein is phosphorylated.	180
Figure 4.22	MST binding analysis suggests that recombinant ZCC_{pT690} protein does not interact with either PBD proteins.	182
Figure 4.23	MST binding analysis suggests that recombinant ZCC protein interacts with both PBD proteins in a CK2-mediated phosphorylation-dependent and oligomeric state-dependent manner.	185
Figure 4.24	Analytical SEC complex analysis suggests that pZCC and PBD do not form a stable complex at a 1:1 stoichiometric molar ratio of ~90μM loading concentrations.	187
Figure 4.25	SDS-PAGE analysis of analytical SEC complex analysis suggests that pZCC and PBD do not form a stable complex at a 1:1 stoichiometric molar ratio of ~250μM loading concentrations.	188

Figure 4.26 SDS-PAGE analysis of analytical SEC complex analysis suggests that pZCC and PBD do not form a stable complex at a 1:13 stoichiometric molar ratio. 189
Figure 4.27 SDS-PAGE analysis of analytical SEC from BS^3 cross-linking experiments suggests that ZCC and PBD may interact phosphorylation-dependently, but independently of T690 phosphorylation status. 192
Figure 4.28 SDS-PAGE analysis of analytical SEC from BS(PEG)9 cross-linking experiments suggests that pZCC and PBD may cross-link independently. 194
Figure 4.29 Structure-guided design for site-specific incorporation of the AbK photoactivated cross-linker at the PBD phospho-ligand binding surface. 197

Chapter 5

Figure 5.1 The SUMOylation cascade. 210
Figure 5.2 Sequence conservation of SUMO family members. 212
Figure 5.3 Structural analysis of SUMO interaction surfaces. 218
Figure 5.4 Sae1:Sae2 structural conformational changes during ATP-binding (open 1), SUMO-binding (open 2), and SUMO adenylation (open 3), and thioesterification (closed). 222
Figure 5.5 The SUMO adenylate mimic (AMSN) compared to native SUMO adenylates and the SUMO tetrahedral adenylate-thioester bond exchange mimic compared to native SUMO bond exchange intermediates. 225
Figure 5.6 The Sae1 inactive adenylation domain and Sae2 active adenylation domain. 230
Figure 5.7 Remodelling of the SUMO adenylation active site drives adenylate-thioester bond exchange. 233
Figure 5.8 The Sae2 crossover loop and re-entry loop act as structural hinges during SUMO thioesterification. 235
Figure 5.9 The Sae2 ubiquitin fold domain (UFD) recruits Ubc9. 239
Figure 5.10 The Sae1-IDR pSer185 site is predicted to be located distally to the SUMO processing active site. 248
Figure 5.11 Purification and Western blot validation of site-specifically phosphorylated recombinant MBP-3C-Sae1-pS185-His6 protein. 250
Figure 5.12 Tandem mass spectrometry (MS/MS) validation of site-specific phosphoserine incorporation at Sae1-S185 in recombinant MBP-3C-Sae1-pS185-His6 protein. 252
Figure 5.13 SDS-PAGE analysis of recombinantly purified Sae1-S185 (A,C-D) and Sae1-pS185 (B,C,E-F)-containing SUMO E1 heterodimers. 256
Figure 5.14 SDS-PAGE analysis of recombinantly purified SUMO 11-97 S9C C52A and SUMO21-93 A2C C48A proteins used for fluorescently labelling with thiol-reactive AlexaFluor 488 and 647 maleimides. 258
Figure 5.15 Fluorescent single turnover in vitro RanGAP1 SUMO1ylation assays suggest that Sae1-S185 phosphorylation has no effect on SUMO E1 activity. 261
Figure 5.16 Fluorescent single turnover in vitro RanGAP1 SUMO2ylation assays suggest that Sae1-S185 phosphorylation has no effect on SUMO E1 activity. 262
Figure 5.17 Evolutionary conservation of Sae1-pS185, the mammalian-specific SQ-motif, and the potential animal-specific Sae1 bi-partite NLS. 266

List of Tables

Chapter 2

Table 2.1	*E.coli* strains used in this study.	26
Table 2.2	Antibiotics used in this study.	28
Table 2.3	Q5 High-Fidelity DNA Polymerase buffer components for standard PCR reactions.	29
Table 2.4	The Q5 High-Fidelity DNA Polymerase thermocycler programme parameters.	29
Table 2.5	QuikChange II buffer components for site-directed mutagenesis PCR reactions.	30
Table 2.6	QuikChange II thermocycle programme parameters for site-directed mutagenesis PCR reactions.	31
Table 2.7	Restriction endonucleases used in this study for traditional cloning.	32
Table 2.8	Vector restriction digest buffer components used for traditional cloning.	32
Table 2.9	Insert DNA restriction digest buffer components used for traditional cloning.	32
Table 2.10	Quick Ligase reaction buffer components used for traditional cloning.	34
Table 2.11	Synthetic peptides used in this study.	42
Table 2.12	Antibodies used to detect phosphorylated Rad50-pT690 and Sae1-pS185 incorporation this study.	47

Chapter 3

Table 3.1	Pellino family members pT +3 and TLR substrate binding specifities.	69
Table 3.2	Summary of ITC binding experiments conducted.	97
Table 3.3	Summary of crystallographic statistics for the Pellino 1_{FHA}:$MDC1_{pT765}$ complex structure.	102
Table 3.4	The human FHA domain family.	111
Table 3.5	Summary statistics from the Local-Global Alignment (LGA) analysis of AF-M *versus* AF3 predictive models compared to the crystal structure of the Pellino 1FHA:MDC1pT765 complex.	117

Chapter 4

Table 4.1	Summary of crystallographic statistics for the unbound PBD structure in the P2 2_1 2_1 space group.	156
Table 4.2	Summary of crystallographic statistics for the unbound PBD structure in the P1 2_1 1 space group.	157
Table 4.3	Summary of MST binding analyses conducted.	186
Table 4.4	Tandem Mass spectrometry analysis (MS/MS) of the ~170kDa cross-linked species band from pZCC:PBD BS^3-cross linking experiments.	193

Table 4.5 Tandem Mass spectrometry analysis (MS/MS) of the >170kDa cross-linked species band from pZCC:PBD BS^3-cross linking experiments. 194

List of Abbreviations

ΔG	Gibbs free energy change
ΔH	Enthalpy change
ΔS	Entropy change
AAD	Active adenylation domain
Abk	(S)-2-Amino-6-((2-(3-methyl-3H-diazirin-3-yl)ethoxy)carbonylamino)hexanoic acid hydrochloride
ABP	Activity-based activity probe
Ac	Acetylated
AIM2	Absent in melanoma-2
ALR	AIM2-like receptor
AMP	Adenosine 5'-monophosphate
AMSN	5'-(aminodeoxy)adenosine-5'-N-monosulfonamide
APS	Ammonium persulphate
AS2TS	Amino acid sequence into tertiary structure
ATM	Ataxia-telangiectasia mutated Ser/Thr kinase
ATP	Adenosine 5'-triphosphate
BCS	Basic Chemical Space Screen HT-96
BMP	Bone morphogenic protein
BRCT	BRCA1 COOH-terminal repeat
BS^3	Bis(sulfosuccinimydl)suberate
BV	Bed volume
C-terminal	Carboxyl-terminal
CC	Catalytic cysteine domain
Chk2	Checkpoint kinase 2
cIAP2	Cellular inhibitor of apoptosis protein 2
CIP	Calf intestinal phosphatase
CK2	Casein kinase 2
CLR	C-type lectin receptor

CPC	Chromosomal passenger complex
CPT	Camptothecin
DAMPs	Damage associated molecular patterns
DD	Death domain
DDR	DNA damage response
DMSO	Dimethyl sulfoxide
DNA-PK	DNA-protein kinase
DOL	Degree of labelling
DP	Differential power
DPBS	Dulbecco's phosphate-buffered saline
DSB	Double strand break
ECD	Ectodomain
FAT	FRAP ATM TRRAP
FCCH	First catalytic cysteine half domain
FHA	Forkhead-associated domain
GDC	Global distance calculation
H1	Histone 1
HCSM	Hydrophobic cluster SUMOylation motif
HR	Homologous repair
HRP	Horseradish peroxidase
HV	High volume
IAD	Inactive adenylation domain
IKKε	I-kappa-B epsilon
IL-1R	Interleukin-1-like receptor
IL-1RAcP	IL-1 receptor accessory protein
intein EPL	Intein-mediated expressed protein ligation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
iPTM	Interface predicted template modelling score

IR	Ionising radiation
IRAK1	Interleukin-1 receptor associated kinase 1
IRAK4	Interleukin-1 receptor associated kinase 4
IRF5	Interferon regulatory factor 5
IRIF	Ionising-radiation-induced foci
ISCM	Inverted SCM
ITC	Isothermal titration calorimetry
JNK	c-Jun NH2-terminal kinase
KAP1	KRAB-associated protein 1
K_D	Dissociation constant
LB	Luria-Bertani
LC-MS	Liquid chromatography-mass spectrometry
LGA	Local-Global Alignment
LGA_S	LGA scoring function
LPS	Lipopolysaccharide
LRM1	LR motif 1
LRR	Leucine-rich repeat
LUBAC	Linear ubiquitin chain assembly complex
MAPK	Mitogen-activated protein kinase
MDC1	Mediator of DNA damage checkpoint 1
MDC1-pTQxF	Phosphorylated MDC1-TQxF
$MDC1_{tBRCT}$	MDC1 C-terminal tandem BRCT repeats domain
MEK1/2	Mitogen-activated protein kinases 1/2
MH2	Mad homology 2
MIU1	Motif interacting with Ub 1
ML-792	[(1R,2S,4R)-4-[[5-[[1-[(3-bromophenyl)methyl]-1H-pyrazol-3-yl]carbonyl]-4-pyrimidinyl]amino]-2-hydroxycyclopentyl]methyl ester
MRN	Mre11/Rad50/NBS1 Complex
MST	Microscale thermophoresis

MyD88	Myeloid differentiation primary response protein 88
MYM	Myeloproliferative and mental retardation
n	Stoichiometry
N-terminal	Amino-terminal
N.D.B.	No detectable binding
NDSM	Negatively charged amino acid-dependent SUMOylation motif
NEB	New England Biolabs
NHEJ	Non-homologous end-joining
NHS	N-hydroxysuccinimide
NK	Natural killer
NLR	Nod-like receptor
NLS	Nuclear localisation signal
NOA	NEMO Optineurin Abin domain
Nod	Nucleotide-binding oligomerisation domain
NPC	Nuclear pore complex
PAC	Photoactivatable cross-linker
PAE	Predicted alignment error
PAMPs	Pathogen associated molecular patterns
PARP1/2	Polyadenosine 5’-diphosphoribose polymerase 1/2
PCNA	Proliferating cell nuclear antigen
PDB	Protein Data Bank
PDSM	Phosphorylation-dependent SUMOylation motif
Peli	Pellino
PES	Polyethersulfone
PIAS1	Protein inhibitor of activated STAT protein 1
PLA	Proximity ligation assay
pLDDT	Predicted local distance difference test
PML	Promyelocytic leukemia protein

PP_i	Pyrophosphate
prfA	Polypeptide release factor 1
ProST	Proline/serine/threonine-rich domain
PRR	Pattern recognition receptor
pSer	Phosphoserine
pSIM	Phosphorylated-SIM
pThr	Phosphothreonine
PTM	Post-translational modification
pTyr	Phosphotyrosine
RFU	Relative fluorescence units
RIG-1	Retinoic acid-inducible-protein 1
RIPK1	Receptor-interacting serine/threonine-protein kinase 1
RLR	RIG-1-like receptor
RMSD	Root-mean square deviation
RU	Relative units
SA	Sigma-Aldrich
SAB	SUMO Assay Buffer
Sae	SUMO-activating enzyme
SBM	SUMO-binding motif
SCM	SUMOylation consensus motif
SDS	Sodium dodecyl sulphate
SENP	Sentrin/SUMO-specific protease
SH2	Src Homology 2 domain
SIM	SUMO-interaction motif
SIRT1	Sirtuin 1
SIRT1/2	Sirtuin 1/2
STUbL	SUMO-targeted ubiquitin ligase
SUMO	Small ubiquitin-like modifier

T	Temperature
TBK1	TANK-binding kinase 1
TCR	T-cell receptor
TEMED	N,N,N',N'-tetramethylethane-1,2-diamine
TFA	Trifluoroacetic acid
TFS	ThermoFisher Scientific
TIF1b	Transcription intermediary factor 1b
TIM	TOPBP1-interaction motif
TIR	Toll/interleukin-1 receptor domain
TLR	Toll-like receptor
TNF	Tumour necrosis factor
TNFR	Tumour necrosis factor receptor
TOP2A	Topoisomerase II a
TRAF3	TNFR-associated factor 3
TRAF6	TNF receptor-associated factor 6
TRAM	TRIF-related adaptor molecule
TRIF	TIR domain-containing adaptor-inducing interferon-b
Ub	Ubiquitin
UBD	Ubiquitin-binding domain
UbL	Ubiquitin-like
UDM1	Ubiquitin-dependent DSB recruitment modules
UFD	Ubiquitin-fold domain
UIM	Ub-interacting motif
UMI	MIU-related UBD

Chapter 1

General Introduction to Phospho-Regulation

1.1 Modular Cellular Signalling through Post-translational Modifications

The intracellular and extracellular environments are in constant flux. As such, cells are dynamically presented with an array of stimuli that demand rapid and precise responses. If left unmet, these stimuli may result in the accumulation of unresolved genetic, mechanical, and chemical damage, which can threaten the survival of the host organism.

One crucial mechanism through which cells maintain this dynamic homeostatic adaptability is through the regulation and diversification of protein function by post-translational modifications (PTMs). To date, over 400 distinct PTMs that have been identified (*Khoury, Baliban, and Floudas, 2011*). While some PTMs, are irreversible (e.g. proteolysis), the vast majority involve the reversible covalent attachment of small chemical groups (e.g. phosphorylation, acetylation, and methylation), or proteinaceous modifiers (e.g. ubiquitination, SUMOylation, and other ubiquitin-like proteins). These molecular switches fundamentally alter the physicochemical properties of proteins, which in turn, have a profound impact on protein stability, localisation, functional activity, and interaction networks. The immense combinatorial potential of chemical and proteinaceous modifications, which are attached and removed in a highly regulated spatiotemporal manner, allows proteins to adopt a diverse array structural and functional states, often referred to as proteoforms. The wider implication of this molecular diversity is that hundreds of thousands- or even millions- of diverse proteoforms can be generated from the ~20,000 protein-coding genes encoded within the human genome.

At the heart of the PTM signalling is a central framework of molecular switches, formed by a three-part modular toolkit, termed modular signalling. Firstly, a group of enzymatic modules act as

"writers" that catalyse the covalent attachment of a PTM to an acceptor substrate. A second class of enzymatic modules, termed "erasers", counteract the activity of "writer" enzymatic modules, through the catalytic removal of the PTM from the modified substrate. Finally, "reader" modules recognise specific PTM-modified motifs, including the local sequence surrounding the modified site, and bind to the modified substrate (Figure 1.1). In turn, this tripartite system enables PTMs to function as molecular barcodes, providing cells with a versatile and sophisticated mechanisms to process and integrate diverse stimuli. Through modular signalling, cells achieve dynamic adaptability while maintaining the specificity required for precise and coordi nated signalling responses.

1.2 Protein Phosphorylation

1.2.1 Protein Phosphorylation: The Original and Most Abundant Post-Translational Modification

Protein phosphorylation is the most abundant protein PTM, accounting for as much as 90% of all identified PTMs (*Ramazi and Zahiri, 2021*). Phosphorylation is catalysed by the kinase superfamily of writer enzymes, commonly referred to as the human kinome (Figure 1.1). The human kinome consists of approximately 538 genes in humans, roughly corresponding to 2% of the entire human genome (*Hanks, Quinn, and Hunter, 1986; Manning et al., 2002; Zhang et al., 2021*). Protein phosphorylation is a reversible process, with the removal of phosphate groups facilitated by the phosphatase family of eraser enzymes (Figure 1.1). The dynamic interplay between kinases and phosphatases ensures precise temporal control of phosphorylation states across cellular proteins, thus enabling reversible modulation of protein activity, stability, and localisation.

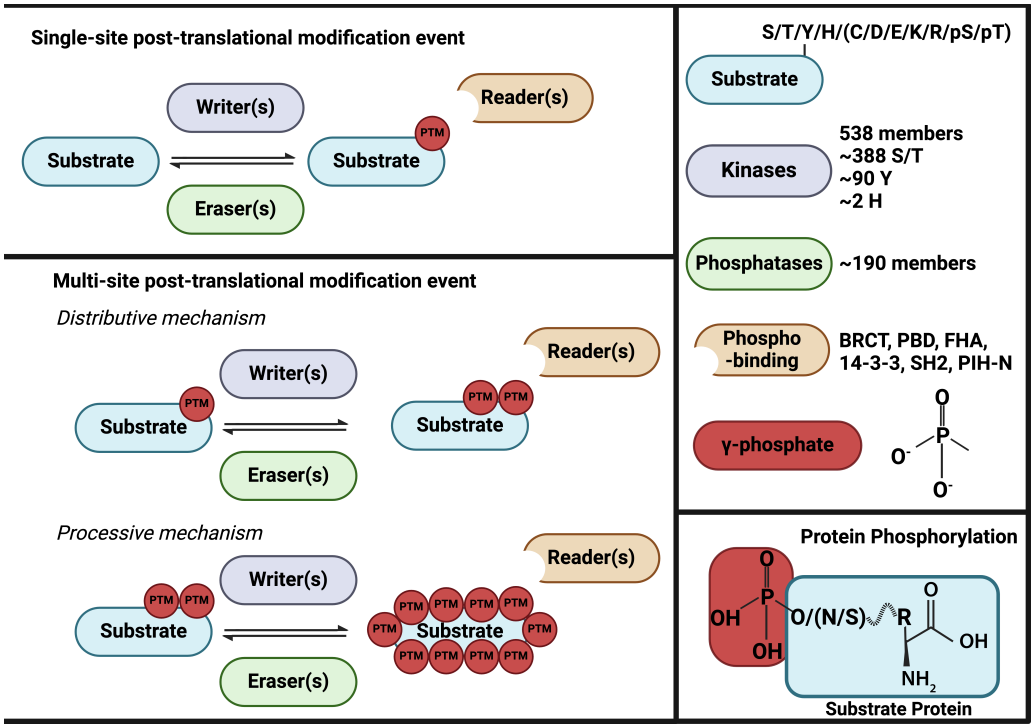


Figure 1.1 – An overview of the modular protein phosphorylation signalling system. Enzymatic "writer" modules (purple), termed kinases (purple), catalyse the reversible transfer of a γ-phosphate moiety (red) from adenosine 5'-triphosphate (ATP) onto specific residues within a substrate protein (light blue) in a process termed protein phosphorylation. A second class of "eraser" enzymatic modules (green), termed phosphatases (green), counteract kinase activity through catalysing hydrolytic removal of the γ-phosphate moiety from the substrate protein. Phosphorylation "reader" modules (orange), termed phospho-binding modules (orange), recognise the specific phosphorylated amino acid motif, including the local sequence surrounding the modified site, and bind to the modified substrate protein. A single protein substrate may be phosphorylated by a single kinase or multiple distinct kinases, including macromolecular complexes, with overlapping consensus sequences at a single site (top left panel), but proteins are often phosphorylated at multiple acceptor sites (bottom left panel). Multi-site phosphorylation can also occur through the diversity of kinase members with distributive, processive and mixed catalytic activity. Extensively investigated human kinase acceptor sites include canonical Ser, Thr, Tyr residues, and non-canonical His residues (top right panel). Additional, yet less characterised, non-canonical acceptor sites include Cys, Asp, Glu, Lys, Arg, phospho-Ser, and phospho-Thr residues (top right panel). The γ-phosphate group covalently attaches to the residue acceptor site through a phospho-ester (P-O), phosphoramidate (P-N), acyl-phosphate (P-O-C=O), thiophosphate (P-S), or diphospho-ester (P-O-P-O) bond to the side chain group of the amino acid (R and wavey dashed line) within a substrate protein (bottom right panel). Figure created with Biorender.com.

Protein phosphorylation involves the catalytic transfer of the γ-phosphate group from adenosine 5'-triphosphate (ATP) to specific residues on substrate proteins. Kinases catalyse the cleavage of the phosphoanhydride bond between the β- and γ-phosphate groups of ATP. The γ-phosphate is typically transferred to the hydroxyl group of serine (Ser), threonine (Thr), or tyrosine (Tyr) residues, resulting in the formation of a stable covalent phosphoester bond (P-O) between the phosphate group and the substrate (Figure 1.2). Serine phosphorylation (pSer) is the most frequent, followed by threonine (pThr) and tyrosine (pTyr) phosphorylation, which occur at an approximate relative frequency of 11.2:2.5:1 (*Bilbrough, Piemontese, and Seitz, 2022*).

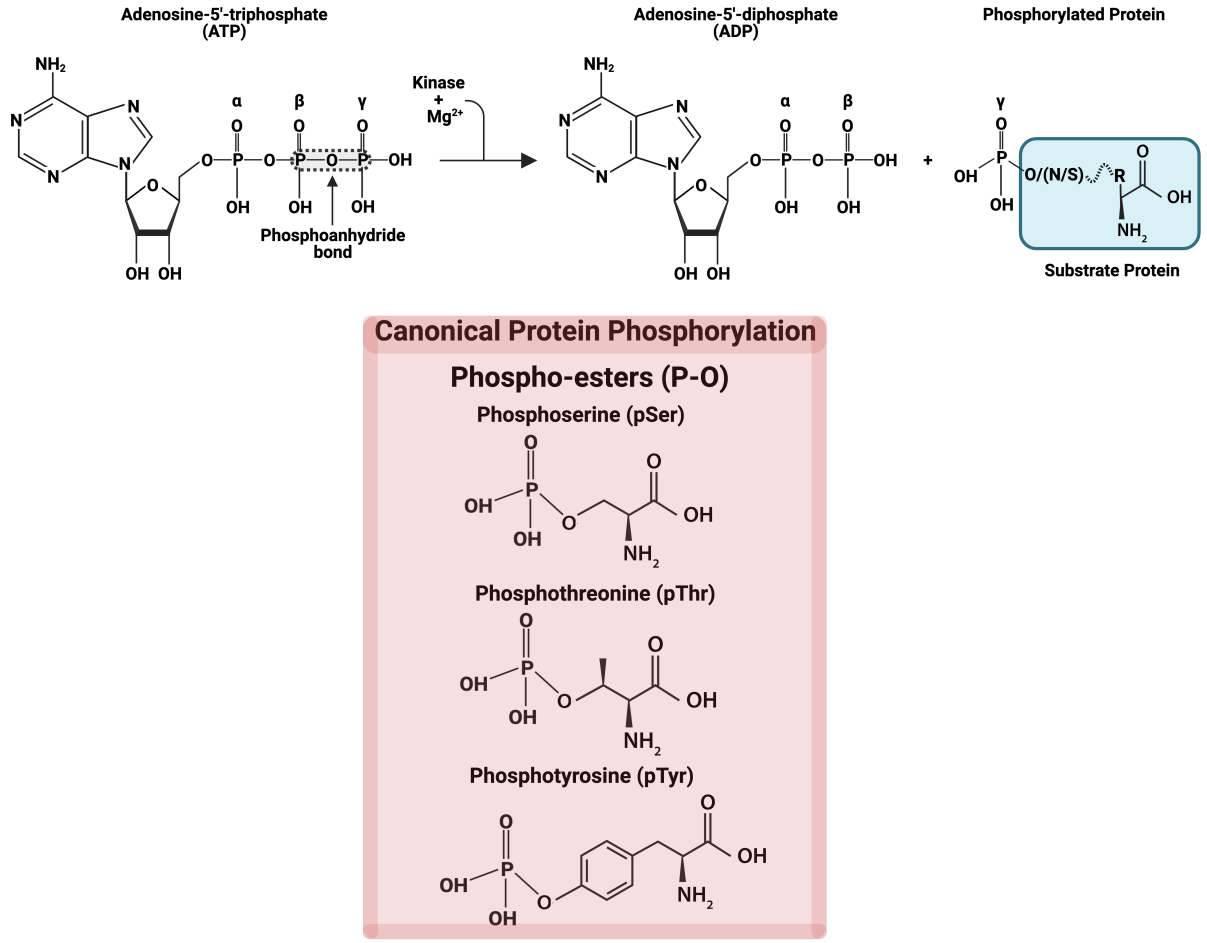


Figure 1.2 – Canonical protein phosphorylation. Kinases catalyse the cleavage of the 5'-adenosine triphosphate (ATP) phosphoanhydride bond, located between the β-phosphate and γ-phosphate group, and mediate transfer of the γ-phosphate group to a substrate protein acceptor site (top panel). Kinases require Mg^{2+} cofactors for efficient catalysis (top panel). In canonical protein phosphorylation, the γ-phosphate group is transferred to the hydroxyl side chain group of substrate Ser, Thr, and Tyr acceptor residues, forming a covalent phospho-ester bond (bottom panel). All chemical structures are depicted in the skeletal formula format. Figures created with Biorender.com.

Although canonical pSer/pThr/pTyr dominate total phosphorylation events, non-canonical protein phosphorylation events have also been identified and remain an area of active research and debate. These include phosphorylation of histidine (pHis), arginine (pArg), lysine (pLys), aspartic acid (pAsp), glutamic acid (pGlu), and cysteine (pCys) residues (Figure 1.3; *Boyer et al., 1962; Hardman et al., 2019*). Among non-canonical phosphorylation events, histidine phosphorylation (pHis) is the

most characterised and is supported by substantial experimental evidence. Histidine kinases, including non-metastatic enzymes 1 and 2 (NME1 and NME2), catalyse the formation of heat- and acid-labile phosphoramidate bonds (P-N) at the N1 and N3 positions of the imidazole side chain of substrate histidine residues, which are reliably detected using monoclonal antibodies and in phosphoproteomics screens (*Fuhs et al., 2016; Hardman et al., 2019; Gary et al., 2024*). Another distinct category of non-canonical phosphorylation is pyrophosphorylation (Figure 1.3), which involves Mg^{2+}-dependent and enzyme-independent transfer of a high-energy β-phosphate moiety from inositol pyrophosphates to pre-phosphorylated serine and threonine substrates (*Penkert et al., 2017; Ganguli et al., 2020; Morgan et al., 2024*).

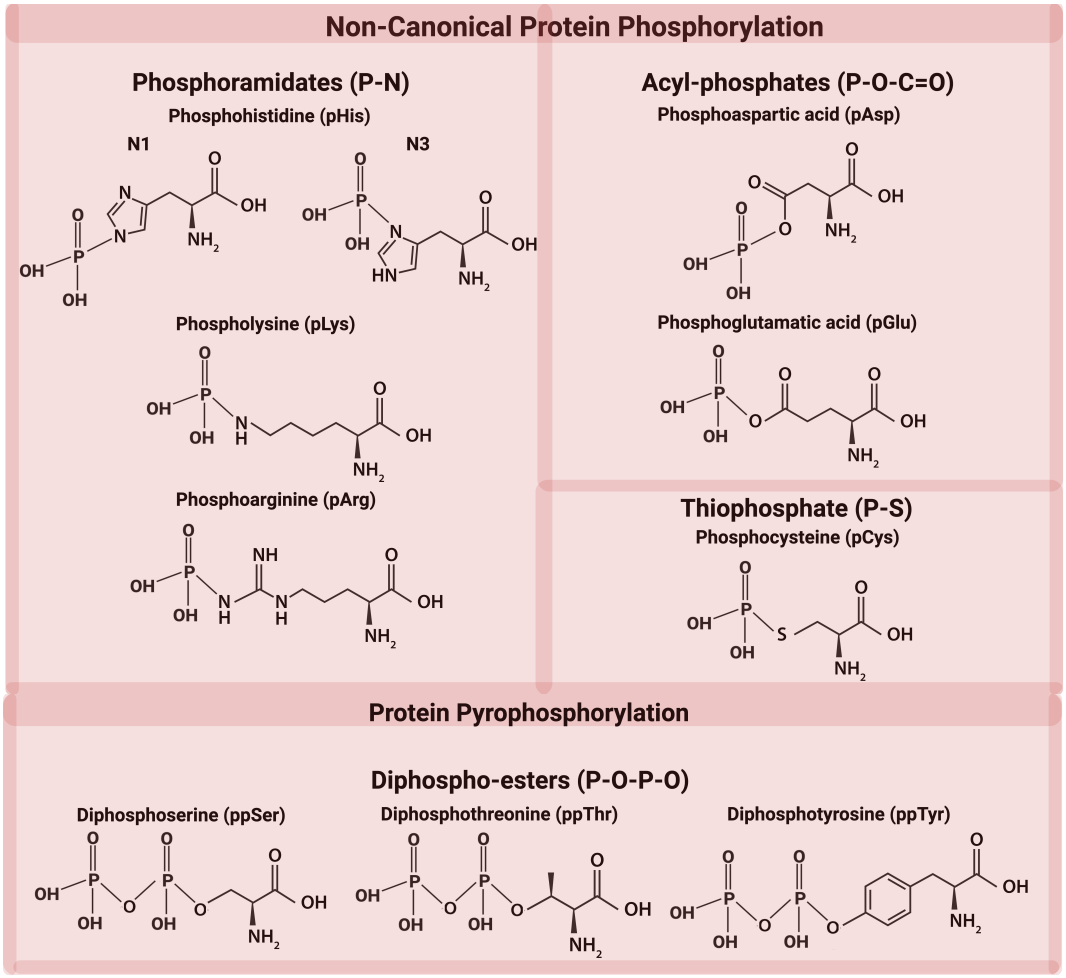


Figure 1.3 – Non-canonical protein phosphorylation. In non-canonical protein phosphorylation, transfer of the γ-phosphate group to a substrate protein may occur through covalent attachment to a nitrogen atom side chain atom (phosphoramidates) in His, Lys, and Arg residues. In histidine phosphorylation, the γ-phosphate group covalently attaches to either the N1 or N3 atom within the imidazole side chain group. Lys phosphorylation may occur through covalent attachment to the ε-amino group nitrogen atom, while Arg phosphorylation may involve biased attachment to the guanidinium side chain group η nitrogen atom. Additionally, γ-phosphate may covalently attach to the carboxylic acid side chain group oxygen atoms of Asp and Glu residues (acyl-phosphates) and the thiol side chain group sulphur atom of Cys residues (thiophosphates). In protein pyrophosphorylation, a β-phosphate group is instead transferred enzyme-independently from inositol pyrophosphates to the γ-phosphate group of a pre-phosphorylated pSer, pThr, and pTyr acceptor site (diphospho-esters). All chemical structures are presented using the skeletal formula. Figure created with Biorender.com.

Eukaryotic kinase superfamily members share a conserved catalytic domain of 250-300 amino acids, which adopt a bilobal structural configuration with distinct N-terminal and C-terminal lobes

(Figure 1.4). The sequence and structural conservation of the bilobal fold highlights the remarkable functional efficiency of kinases. Key conserved motifs within the fold, including GxGxxG, AxK, DFG, and HRD motifs, are essential for ATP binding, Mg^{2+} coordination, substrate positioning, and stabilisation of active-state conformations during catalysis (Figure 1.4; *Taylor and Kornev, 2011*). The P+1 loop is a pivotal feature that influences recognition of distinct substrate sequence consensus motifs through interactions with residues adjacent to the phosphorylation site (Figure 1.4; *Taylor and Kornev, 2011*). Thus, the P+1 loop in conjunction with other structural features ensures substrate specificity, which varies across eukaryotic subfamilies. The activation- or T loop- often contains a phosphorylatable Ser/Thr residue, which, upon phosphorylation by upstream kinases, stabilises active-state structural configurations and is thus able to modulate kinase activity (Figure 1.4). The sequence conservation and substrate specificity of kinases enabled the identification of 538 members, and their classification of kinases into eight major groups and 189 subfamilies (Figure 1.5; *Hanks, Quinn, and Hunter, 1986; Manning et al., 2002*). Furthermore, Kinases are classified into three broad categories based on substrate specificity: serine/threonine-specific kinases, tyrosine-specific kinases, and dual-specificity kinases that phosphorylate both serine/threonine and tyrosine residues.

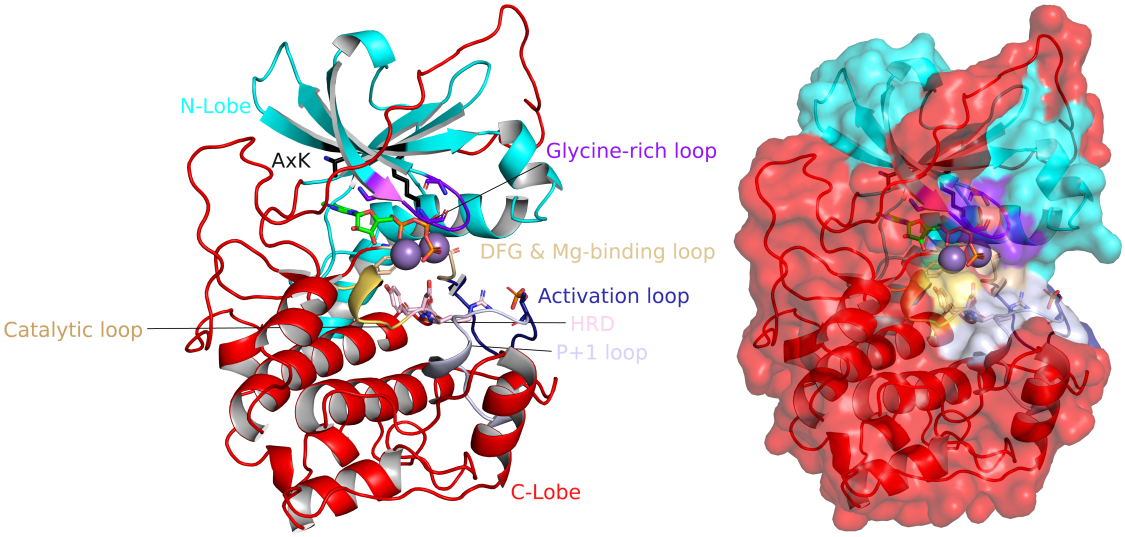


Figure 1.4 – The Hanks-type kinase bilobal fold. The secondary structure of murine protein kinase A (PKA)/cAMP-dependent protein kinase in complex with the non-hydrolysable ATP analogue, adenylyl-imidodiphosphate (AMP-PNP), and two Mg^{2+} (light purple spheres) is provided (left panel; accession code: 1ATP; *Zheng et al., 1993*). The AMP-PNP molecule is depicted in the stick format and coloured by atom. The AMP-PNP molecule binds in a central cleft between the N-lobe (cyan) and C-lobe (red) as shown by the PKA surface representation (right panel). Key canonical kinase features within the N-lobe, such as the A-x-K (black) motif and the G-x-G-x-x-G motif residues within the glycine-rich loop (purple blue) are highlighted. Both motifs are involved in coupling and positioning of the γ-phosphate group from ATP and the N-lobe C helix for catalysis. Conserved C-lobe elements, including the H-R-D motif (Y-R-D in PKA; light pink), located within the catalytic loop (yellow orange), and the D-F-G motif-containing Mg-binding loop (wheat) are indicated. The catalytic loop H-R-D motif Arg residue functions to stabilise active-state kinase structural conformational states through an electrostatic interaction with the activation loop (density) phospho-regulatory site, while the Asp residue acts as a catalytic base during phospho-transfer. The active-state stabilisation facilitates the interaction between the D-F-G motif Asp residue and an activating Mg^{2+} ion cofactor during catalysis. This region precedes the indicated P+1 loop (blue white), which is involved in binding to the kinase substrate +1 residue. All key canonical kinase motifs are represented in the stick format. Adapted from figure published by *Taylor and Kornev, 2011*. Figure created using the PyMOL Molecular Graphics System v3.1.1 (Schrödinger, LLC).

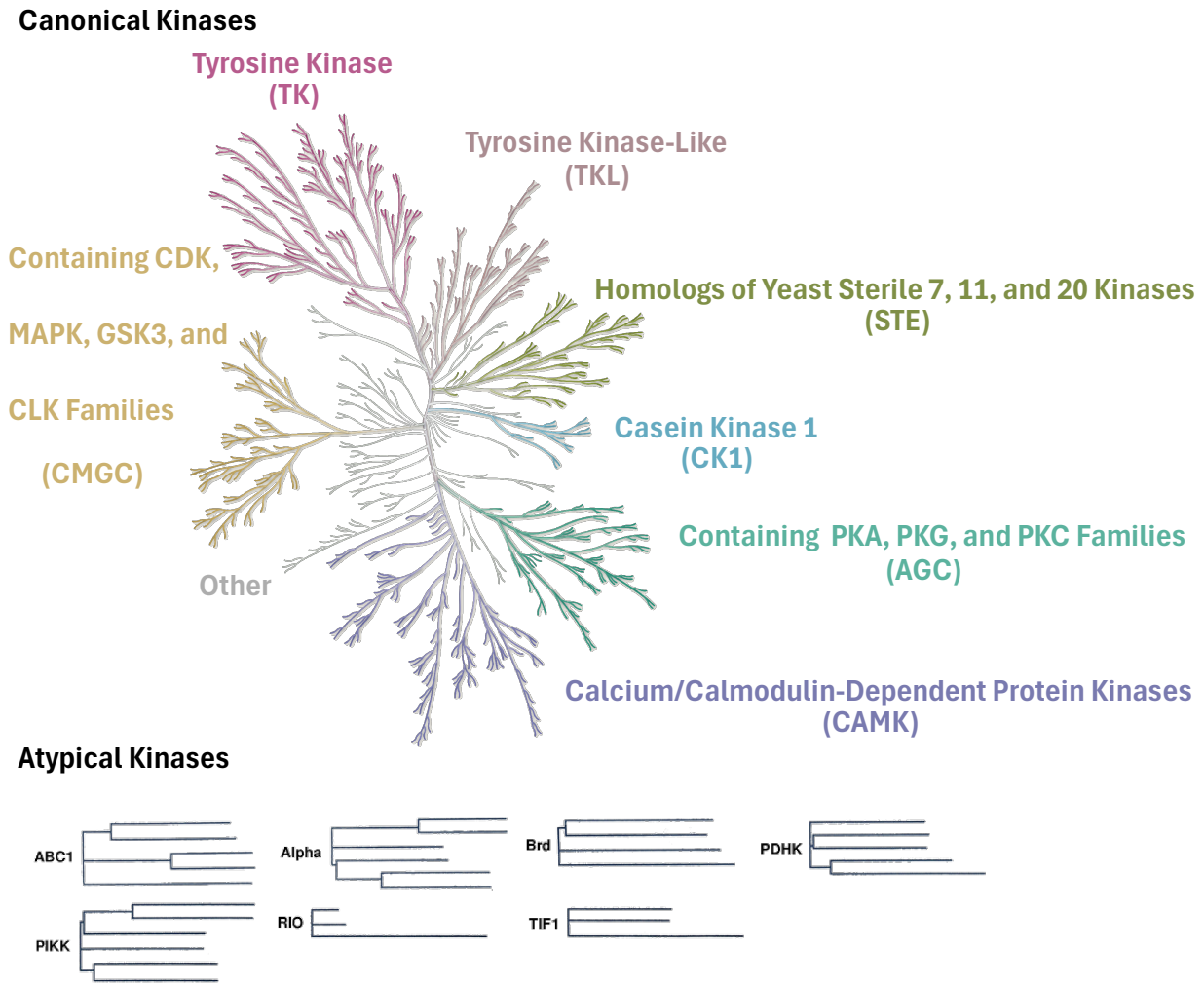


Figure 1.5 - The human kinome. Each dendrogram of typical (top panel) and atypical (bottom panel) human protein kinases illustrates the sequence-based classification of human protein kinases into 8 major groups according to catalytic domain similarity. Each major branch corresponds to a distinct kinase major group, while internal branching points represent kinase subfamilies and each terminal branching point represents a single protein kinase member *(Manning et al., 2002)*. Figure created using KinMap *(Eid et al., 2017)*. Permission of use granted by Cell Signalling Technology, inc.

Protein phosphorylation was first identified in studies of tumour cells, and kinase activity was demonstrated in early work on casein phosphorylation by casein kinases (*Williams-Ashman and Kennedy, 1952; Burnett and Kennedy, 1954*). While these initial studies were pivotal for recognising protein phosphorylation, the direct phospho-regulation of a substrate protein was first identified in research on glycogen metabolism in rabbit skeletal muscle (*Fischer and Krebs, 1955; Krebs, Kent,*

and Fischer, 1958; Fischer et al., 1959). This seminal work identified two forms of glycogen phosphorylase: a minimally active dimer requiring high concentrations of ADP (phosphorylase B) and an ADP-independent active tetramer (phosphorylase A) that catalyses phosphorolysis of the α-1,4-glycosidic bonds of terminal glucose residues (*Fischer and Krebs, 1955*). The conversion of phosphorylase B to phosphorylase A was shown to depend on phosphorylation of glycogen phosphorylase B-S15 by phosphorylase B kinase in the presence of ATP and Mg^{2+} (*Krebs, Fischer, and Kent, 1958; Fischer et al., 1959*). Earlier, it had been shown that phosphorylase A could be converted back to phosphorylase B by a phosphatase that removed the prosthetic phosphate group, later identified as protein phosphatase 1 (*Cori and Green, 1943*).

1.2.3 Protein Phosphorylation: Navigating Kinase Specificity and Noise

The initial discovery of a single phosphorylation-dephosphorylation relationship by a single pair of kinase-phosphatases as a means of regulation of glycogen phosphorylase activity was later shown to be the exception to the rule. Phospho-regulation is now recognised to be far more complicated than first anticipated, relying on complex signalling cascades, wherein multiple kinase-phosphatase relationships catalyse phosphorylation of single proteins and protein complexes at multiple sites. This complexity is further exemplified by extensive phosphorylation crosstalk with other PTMs to coordinate precise spatiotemporal of protein regulation and cellular signalling. Advances in phosphoproteomics have revealed the immense scale of the complexity of phosphorylation, with as many as 90,000 pSer, pThr, pTyr high confidence phosphosites identified across 12,000 proteins in the human phosphoproteome (*Ochoa et al., 2020*).

The vast majority of these phosphorylation sites have no known kinase or cellular function. In fact, many phosphorylation sites and kinase-phosphorylation relationships are poorly conserved through evolution, and appear to have no structural or functional consequences, leading to the classification of these phosphorylation events as functionally silent and non-specific (*McDonald, Trost, and Napper, 2018*). Non-specific phosphorylation events represent background noise within the milieu of true functional phosphorylation events, which are fundamental for precise control of protein regulation. These non-specific phosphorylation events are likely a consequence of the limited specificity provided by kinases, which often share overlapping/degenerate consensus phosphorylation sequences and exhibit similar localisation patterns. As a result, non-specific phosphorylation may arise as a result of random encounters in the crowded cellular environment. In turn, the vast scope of non-specific phosphorylation appears a very wasteful use of metabolic ATP production. However, estimations suggest that non-specific phosphorylation consumes only ~1% of total cellular ATP production, which is a negligible metabolic cost to pay in the context of cellular energy output (*Lienhard, 2008*). As a result, there is minimal selective evolutionary pressure to eliminate non-functional phosphorylation events, provided it has no major negative impact on protein stability and activity.

1.2.4 Roles of Phosphorylation: True Signal in a World of Noise

Non-specific phosphorylation events may indeed be an evolutionarily tolerable by-product of degenerate kinase specificity and the crowded cellular environment, however there is a significant proportion of phosphorylation events that are vital for protein regulation in cellular signalling.

Phosphorylation-dependent protein regulation is highly effective as a molecular switch, because of the inherent biochemical properties of the γ-phosphate. Each phosphorylation event introduces a bulky, dianionic, and tetrahedrally coordinate γ-phosphate group into the otherwise polar and bent hydroxyl groups of Ser/Thr/Tyr side chains (*Bilbrough, Piemontese, and Seitz, 2022*). As a result, phosphorylation can alter the local electrostatic environment surrounding the phosphosite, which can destabilise existing intramolecular and intermolecular interactions within an unmodified protein through electrostatic repulsion, and also facilitate the formation of new interactions through electrostatic attraction. The geometric rearrangement induced by phosphorylation can also impact the local structural conformation of the phosphorylation site, while the four oxygen atoms in γ-phosphate provide at least three hydrogen-bond acceptor sites, thus enabling specific interactions with hydrogen-bond donors.

One of the broader implications of the biochemical properties of the γ-phosphate group in driving protein regulation is in its ability to induce structural conformational changes, which can bias or stabilise a protein towards a specific conformational state by reducing the number of dynamic conformational ensembles a protein adopts. This allosteric phospho-regulatory mechanism is particularly evident in the control of phosphotransferase activity of most eukaryotic kinases (Figure 1.6; *Reinhardt and Leonard, 2023*). In this process, phosphorylation of specific Ser/Thr sites within the T-loop induces a conformational shift that disrupt autoinhibitory interactions, while also facilitating critical interactions that optimally position active site residues primed for efficient substrate phosphorylation (Figure 1.6). This ability of kinase-catalysed phosphorylation events leading to activation of downstream kinases results in signal amplification and is a central property of signalling cascades.

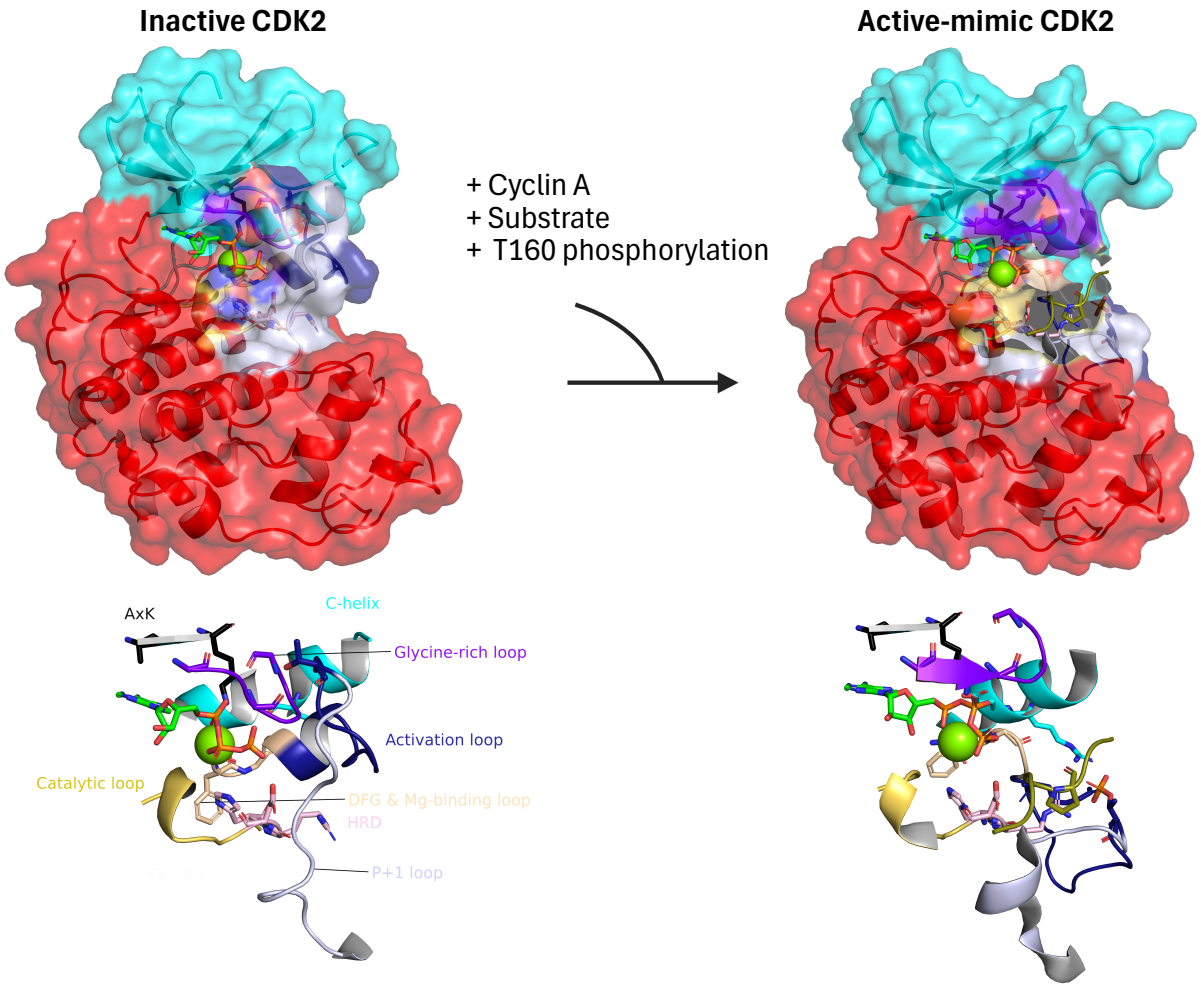


Figure 1.6 – Protein phosphorylation priming drives allosteric protein kinase activation. The surface representation of inactive human cyclin-dependent kinase 2 (CDK2) bound to ATP and a single Mg ion (green sphere) are shown (top left panel; accession code: 1HCK; *Schulze-Gahmen, Bondt, and Kim, 1996*), alongside the surface representation of active CDK2 in complex with AMPNP, cyclin A (not shown), substrate peptide (olive), and a single Mg ion (green sphere) are provided (top right panel; accession code: 1QMZ; *Brown et al., 1999*). All key features are coloured according to the same scheme as in Figure 1.4 and canonical kinase motifs are represented in the stick format. The conformational change induced by phosphorylation priming, substrate binding and cyclin A binding-mediated CDK2 activation is demonstrated through the secondary structures of canonical Hanks-type kinase features within CDK2 (bottom panels). An electrostatic interaction between the A-x-K motif Lys residue and C-helix (cyan) Glu residue stabilises the C-helix in an optimal position for catalysis. Additionally, the activation loop is repositioned out of the active site through hydrogen-bonding contacts between the pThr γ-phosphate group and the H-R-D motif Arg residue, C-helix, and cyclin A (not shown), which stabilises the active kinase structural conformation. Figure created using the PyMOL Molecular Graphics System v3.1.1 (Schrödinger, LLC).

For example, in the DNA damage response (DDR) pathway, double-stranded DNA breaks trigger recruitment of the MRE11:Rad50:Nbs1 (MRN) sensor complex, which triggers activation of the apical ataxia-telangiectasia mutated Ser/Thr kinase (ATM). ATM activation involves dissociation of inactive dimers into active monomers and is associated with autophosphorylation at Ser1981 (*Warren and Pavletich, 2022*). Amongst over 700 substrates, activated ATM phosphorylates the transducer kinase, checkpoint kinase 2 (Chk2) at Thr68, thereby initiating its initial kinase activity (*Ahn et al., 2000; Matsuoka et al., 2007*). Chk2 subsequently dimerises and undergoes autophosphorylation of Thr383 and Thr387, located in the T loop, resulting in full Chk2 activation (*Schwarz et al., 2003; Cai, Chehab, and Pavletich, 2009*). ATM-mediated Chk2 activation increases the functional repertoire of active kinases, which collectively exhibit shared and distinct substrate preferences. Consequently, these initial events contribute to both phosphorylation diversification and amplication within the DDR. A canonical example of ATM and Chk2-mediated phosphorylation as a means to reinforce DDR signalling is the p53-p21 signalling axis. ATM-mediates phosphorylation of the p53 transactivation domain at S15, as well as the negative regulator of p53, MDM2, at S395, while Chk2 phosphorylates p53 at S20 *(Canman et al., 1998; Hirao et al., 2000; Maya et al., 2001)*. In turn, these phosphorylation events disrupt the inhibitory p53-MDM2 interaction, thereby stabilising p53 levels and promoting p53-mediated transcriptional upregulation of p21. Ultimately, this cascade culminates in S-phase checkpoint arrest, enabling time for efficient DNA repair (*Abuetabh et al., 2022*).

A second major implication of protein phosphorylation is the formation of unique and specific binding surfaces, which are recognised by phospho-binding ‘reader’ modules. Phospho-binding modules provide specificity to the largely non-specific nature of the total phosphorylation pool by

mediating highly specific and often high affinity protein-protein interactions (Figure 1.7). In turn, these interactions are facilitated through salt bridge and hydrogen-bonding contacts with the pSer/pThr/pTyr γ-phosphate group(s) through complementary interaction surfaces containing basic amino acids (His/Lys/Arg), peptide backbone atoms and free hydroxyl group-containing amino acids, as well as through recognition of the local sequence surrounding the phosphosite(s). The latter feature is termed the phosphorylation interaction consensus sequence and is the driving force of true phospho-specificity.

Phospho-binding modules are non-catalytic domains that play a crucial role in the recruitment of proteins to phosphorylated target sites, often influencing the subcellular localisation of proteins containing these domains. This functional feature is exemplified by studies of the Polo-like kinase 1 (PLK1) non-catalytic C-terminal Polo-Box domain (PBD) and Forkhead-associated (FHA) domains (Figure 1.7; *Durocher et al., 2000; Elia et al., 2003*). The PBD specifically recognises and binds to [P/F]- [Φ/P]-[Φ]-[T/Q/H/M]-S-[pSer/pThr]-[P/X]-containing phosphoproteins, which are phosphorylated throughout every single stage of the cell cycle. Consequently, the PBD directs the subcellular recruitment of PLK1 throughout every dynamic cell cycle transition, thereby enabling its kinase domain to regulate key mitotic and DNA damage events (*Lowery et al., 2007; Macůrek et al., 2008; Yim and Erikson et al., 2009; Shen et al., 2013*). Conversely, FHA domains recognise and bind to [pThr]-x-x-D/I/L/Y/M/S/A motifs in phosphoproteins (Figure 1.7), where phosphorylation is induced in response to DNA damage (*Durocher et al., 2000; Huen et al., 2007; Huoh and Ferguson 2014*). This specificity is vital for the recruitment of FHA domain-containing proteins, such as RNF8, to DNA damage sites during DDR signalling-mediated maintenance of genome stability (*Huen et al., 2007*).

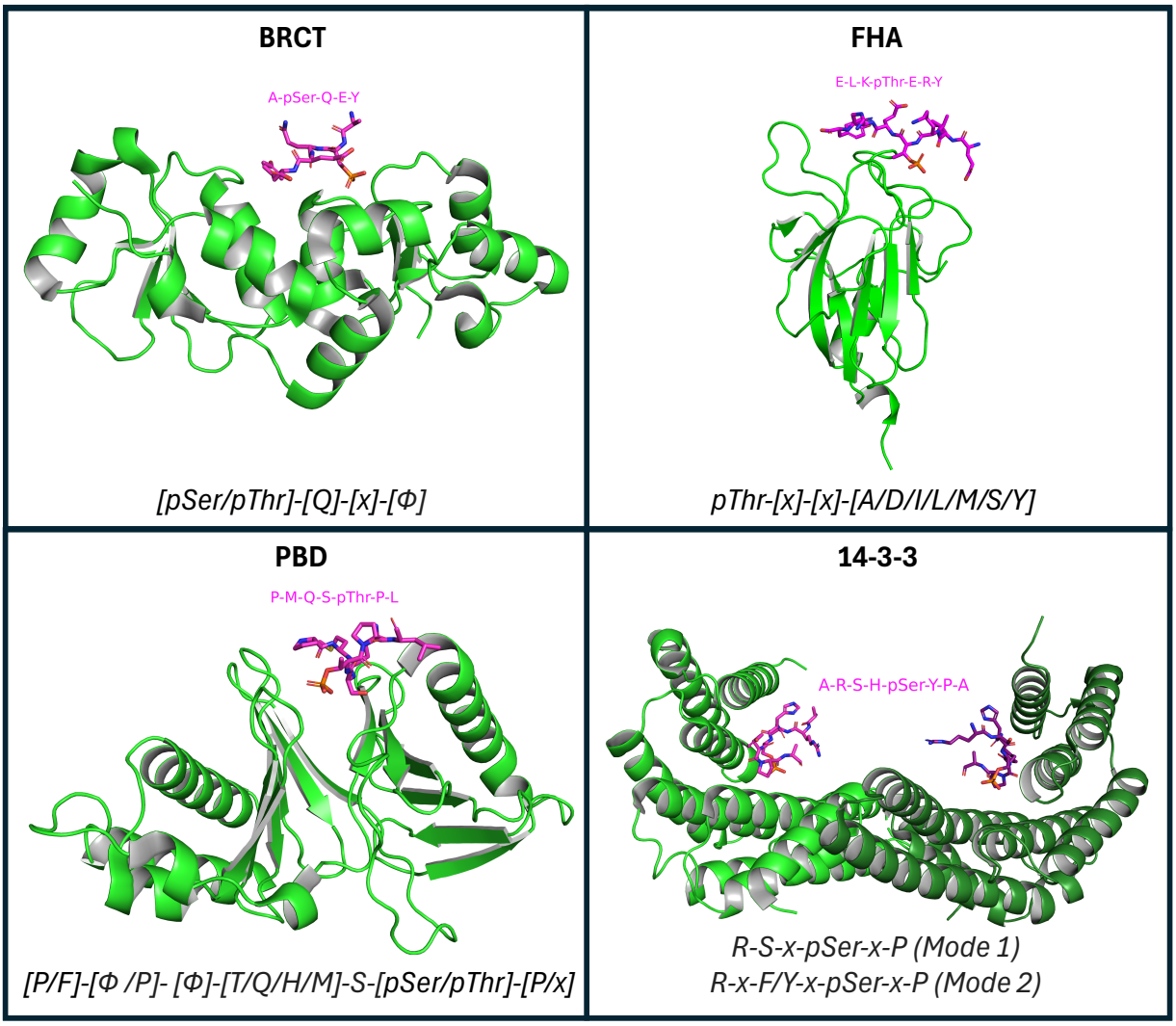


Figure 1.7 – Phospho-binding module families recognise and bind to specific and distinct phospho-epitopes. The crystal structures of the human MDC1 tandem BRCT domain bound to a γ-H2AX phosphopeptide (top left panel; accession code: 2AZM; *Stucki et al., 2005*), RNF8 FHA domain bound to an optimal phosphopeptide (top right panel; accession code: 2PIE; *Huen et al., 2007)*, PLK1 PBD bound to an optimal phosphopeptide (bottom left panel; accession code: 1UMW; *Elia et al., 2003)*, and 14-3-3ζ homodimers bound to two optimal phosphopeptides containing a mode 1 consensus motif (bottom right panel; accession code: 1QJB; *Rittinger et al., 1999)* are provided. The secondary structure of each phospho-binding domain is presented in the cartoon format (green or forest green) with the phosphopeptide in the stick format (magenta or violet purple) and coloured by atom. Each phosphopeptide amino acid sequence is highlighted at the top of each structure. The phospho-binding interaction consensus sequence for each pSer/pThr-specific phospho-binding domain is detailed at the bottom of each panel. Figure created using the PyMOL Molecular Graphics System v3.1.1 (Schrödinger, LLC).

Similarly, phospho-binding modules serve a critical role in the formation of multi-domain proteincomplexes and provide direct links between phosphorylation and other PTM signalling cascades (Figure 1.8), such as the ubiqutination and ubiquitin-like pathways. The role of phospho-binding modules in protein complex assembly is highlighted by the scaffold properties of mediator of DNA damage checkpoint 1 (MDC1), which is dependent on its DNA damage-induced phosphorylation (Figure 1.8; *Huen et al., 2007; Chapman and Jackson, 2008*). For instance, during ionising-radiation induced foci (IRIF) formation, phosphorylation-dependent recruitment to MDC1 involves the Nbs1 component of the MRN complex via its FHA-BRCT super module, as well as RNF8 through its respective FHA domains (Figure 1.8; *Huen et al., 2007; Chapman and Jackson, 2008*). This recruitment facilitates PTM crosstalk during DDR signalling, because RNF8 contains a C-terminal E3 RING domain, which catalyses Lys63-linked polyubiquitination of histone 1 (Figure 1.8; *Huen et al., 2007; Thorslund et al., 2015*).

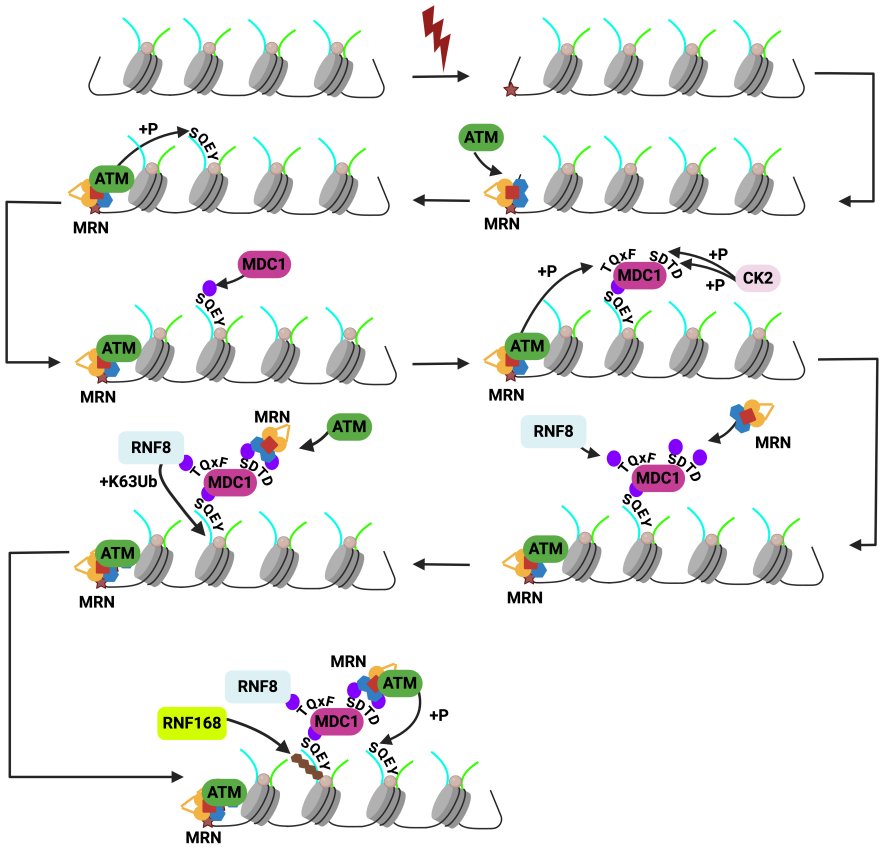


Figure 1.8 – Phospho-binding modules are fundamental for multi-domain complex formation and PTM cross-talk within the DNA damage response. Chromatin structures are represented with DNA (black line), nucleosomes (grey barrels) with exposed H2AX (cyan) and H2B (light green) tails, and H1-linker histones (beige balls). Ionising radiation (red lightning bolt) induces double-stranded DNA break formation (red star), which leads to recruitment of the heteropentameric/hexameric MRE11-Rad50-NBS1 (MRN; blue hexagons, yellow circles and red square) complex. Inactive ATM Ser/Thr kinase (green) is recruited to the MRN via an interaction with NBS1, whereupon it dissociates into active monomers and catalyses phosphorylation (+P; purple) of H2AX at S139 (SQEY). Phosphorylated H2AX (γH2AX) generates a specific MDC1 (magenta) tandem BRCT repeat domain interaction surface, which mediates MDC1 recruitment. ATM catalyses phosphorylation of the recruited MDC1 at T-Q-x-F motifs, while CK2 (pink) phosphorylates MDC1 at S-D-T-D motifs. Both phosphorylation events create phosphorylation-dependent interaction surfaces for the RNF8 (light blue) FHA domain and NBS1 FHA-BRCT repeat supermodular domain, leading to recruitment of RNF8 and further MRN complexes to DSB sites. MRN complex retention leads to further recruitment and activation of ATM kinase as a positive DDR signalling feedback loop. Recruited RNF8 catalyses K63-linked polyubiquitination (+K63-Ub; brown hexagons) of H1, which triggers recruitment of RNF168 (lime green). Figure created with Biorender.com.

1.3 Phosphointeractomics: Insights and Challenges

Phospho-binding modules have been extensively identified and characterised through high-throughput phosphointeractomics and oriented peptide library screens, which, in turn has provided insight into the specificity of their consensus interaction motifs (*Songyang et al., 1993; Elia, Cantley, and Yaffe, 2003*). Together, these methods are power tools that were fundamental for characterisation of key families of phospho-binding modules, including the Src Homology 2 (SH2) domains that recognise pTyr, 14-3-3 with specificity for Arg-Ser-x-pSer-x-Pro and Arg-x-Tyr/Phe-x-pSer-x-Pro motifs, FHA-pThr interaction motifs, and the PLK1-PBD (*Songyang et al., 1993; Durocher et al., 2000; Elia, Cantley, and Yaffe, 2003; Elia et al., 2003*).

High-throughput phosphointeractomics often employs the use of peptide versus phosphopeptide baits from known phosphosites in affinity pull-down experiments from cell extracts, followed by mass spectrometry to identify phosphorylation-dependent interaction partners. Despite its utility, this approach has several major limitations. Selecting an appropriate cellular stimulus is difficult, as it must maximise the abundance of binding partners, while minimising non-specific binders. Use of phosphopeptides removes the structural and functional context of full-length proteins, which can have the unintended consequence of altering the true affinity between phosphoproteins and their binding partners. This approach may not detect transient interactions, which are of substantial importance in the context of dynamic cellular signalling. Finally, detected binding partners may not directly interact with the phosphopeptide, but instead bind indirectly via a *bona fide* interactor, thereby complicating the data interpretation.

1.4 Aims and Objectives

Phosphoproteomics has highlighted the vast complexity of the human phosphoproteome, much of which remains functionally uncharacterised. The true challenge of the scale of the phosphoproteome lies in the discernment of functionally significant phosphorylation sites from the extensive background noise provided by non-specific phosphorylation events. Consequently, addressing this issue requires direct experimental approaches that validate and characterise phosphorylation events at the molecular level, while closely mimicking the true biochemical and structural physiological state of phosphoproteins. In addition, phosphointeractomics has been instrumental for the characterisation of phosphorylation-dependent interactions, however the numerous uncharacterised phosphosites suggests that many phospho-binding modules remain to be found.

The DDR cascade represents an ideal model system to study the modularity of phosphorylation. It is a signal transduction pathway that is underpinned by phosphorylation-dependent interactions, phosphorylation-dependent allosteric protein regulation, and extensive PTM crosstalk between phosphorylation, ubiqutination, and SUMOylation. Despite advances in identification of phosphorylation sites in DDR proteins through phosphoproteomics screens, the precise molecular roles of many of these phosphorylation events remain poorly understood.

The aims of this PhD were to decipher the potential phospho-regulation of 3 key DDR proteins- MDC1, Rad50 and Sae1- which have been identified as phosphorylation targets in response to DNA damage through phosphoproteomics screens. Specifically, this work focused on whether phosphorylation of MDC1-TQxF motifs facilitates the recruitment of the Pellino 1 FHA domain,

phosphorylation of Rad50 at T690 induces a phosphorylation-dependent interaction with the PLK1-PBD, and phosphorylation of Sae1 at S185 allosterically regulates SUMO E1 activity. This work combined a multi-disciplinary approach that sought to uncover the molecular details of phospho-regulation of these proteins in the DDR. This approach included biophysical binding experiments, genetic code expansion protein engineering, protein biochemistry, and both experimental and AI-driven predictive structural biology.

Chapter 2

Materials and Methods

2.1 Molecular Biology Techniques

2.1.1 Subculturing and Long-term Storage of *E.coli* Strains

Non-competent bacterial cells (Table 2.1) were streaked onto LB-agar solid media (1% (w/v) NaCl, 1% (w/v) tryptone, 0.5% (w/v) yeast extract, and 1.5% (w/v) agar) using the quadrant streak procedure and incubated at 37°C to isolate single colonies. Single colonies were inoculated into 10mL LB liquid media (1% (w/v) NaCl, 1% (w/v) tryptone, 0.5% (w/v) yeast extract) cultures grown or 16-18 hours at 37°C, 190rpm. 600µL of overnight culture was mixed with 400µL of filter-sterilised 50% (v/v) glycerol and stored at -80°C in a 2mL cryogenic vial (Nalgene).

2.1.2 Bacterial Transformations

Bacterial transformations were routinely conducted to propagate plasmids (NEB®Stable), generate transgenic plasmids and single point mutations and for either standard recombinant protein expression (BL21 (DE3)) or site-directed unnatural amino acid incorporation.

2.1.2.1 Single-plasmid transformations using chemically competent bacterial strains

25µL of chemically competent cells were thawed on ice and 100ng of plasmid DNA or 2µL of ligation mixture or 1µL of mutagenesis PCR product was added to the cells. The DNA:bacteria mixture was incubated on ice for 30 minutes and then the cells were heat-shocked in a water bath at 42°C for either 10s (BL21(DE3)), 30s (NEB Stable), or 45s (XL1-Blue). 200µL of SOC outgrowth media (New England Biolabs (NEB) #B9020S: 2% (w/v) vegetable peptone, 0.5% (w/v) yeast extract, 10mM NaCl,

2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, and 20mM D-glucose) were added to the cells and the transformation mixture was grown at 37°C, 190rpm for 1 hour. The transformation mixture was then plated onto selective LB-agar solid media supplemented with antibiotic (Table 2.2) and spread using ColiRoller™ plating beads (Novagen). Finally, the plates were incubated upside down for 16-18 hours at 37°C.

***E. coli* Strain**	**Genotype**	**Source**	**Use**
NEB® Stable (Competent)	F' proA+B+ lacIq Δ(lacZ)M15 zzf::Tn10 (TetR)/Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- Φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)	NEB #C3040I	Propagation of plasmid DNA and traditional cloning.
XL1-Blue (Competent)	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´proAB lacIqZΔM15 Tn10 (Tetr)]	Agilent Technologies #200521	Site-directed mutagenesis.
BL21 (DE3) (Competent)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	NEB #C2527I	Recombinant protein expression.
BL21 ΔserB*	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS ΔserB λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	A gift from Jesse Rinehart & Dieter Söll (Addgene plasmid #34929)	Recombinant protein expression with site-directed incorporation of phosphoserine.
rEcoli XpS*	MG1655-C321 mutS+, λ-, Δ(ybhB-bioAB)::zeoR, ΔprfA, ΔserB	A gift from Jesse Rinehart (Addgene plasmid #192872)	Recombinant protein expression with site-directed incorporation of phosphoserine.
BL21 ΔserC ΔycdX*	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS ΔserC ΔycdX [λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5]	A gift from Jason Chin (University of Cambridge)	Recombinant protein expression with site-directed incorporation of phosphothreonine.
BL21(DE3) ΔycdX ΔserC ΔpphA*	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 ΔycdX ΔserC ΔpphA	A gift from Jesse Rinehart (Yale University)	Recombinant protein expression with site-directed incorporation of phosphothreonine.
C321 . ΔA.exp ΔycdX ΔserC ΔpphA*	C321.ΔA.exp (MG1655 mutS+, λ-, Δ(ybhB-bioAB)::zeoR, ΔprfA; all 321 UAG codons changed to UAA) DycdX DserC DpphA	A gift from Jesse Rinehart (Yale University)	Recombinant protein expression with site-directed incorporation of phosphothreonine.

Table 2.1 - *E.coli* strains used in this study. The genotype, source, and purpose of each strain are indicated. All strains were stored at -80°C in 20% (v/v) glycerol. *see 4.3.7 and 5.4.1 for more details.

2.1.2.2 Dual-plasmid transformations using non-competent bacterial strains

This protocol is adapted from the procedure detailed in *Mohler et al., 2023*, but was employed for both site-directed unnatural amino acid incorporation and for co-expression of casein kinase 2 with the Rad50 zinc-coiled-coil fragment. Bacteria from glycerol stocks were transferred aseptically into 5mL of LB liquid media using a pipette tip, and cultures were grown for 16-18 hours at 37°C, 250rpm. The following day, 100µL of overnight culture was inoculated into 20mL of LB liquid media, and was grown at 37°C, 250rpm until cultures reached an OD_{600nm} of ~0.4-0.5. All further steps were conducted on ice unless stated otherwise. Bacterial cultures were pelleted by centrifugation at 13,000xg for 1 minutes, and the supernatant was discarded, while the bacterial pellet was washed using 20mL of ice-cold sterile 10% (v/v) glycerol by gentle pipetting. The pellet- discard-glycerol wash procedure was conducted thrice. The bacterial pellets were then resuspended gently with 100µL ice-cold sterile 10% (v/v) glycerol and 50µL of cells were transferred aseptically to each sterile 1.5mL Eppendorf tubes. 100ng of orthogonal translational system component plasmid and 100ng of expression plasmid containing a site-specific amber stop codon were added aseptically to the bacterial cells, and the transformation mixture was then transferred to a pre-chilled 1.8mL, 0.1cm gap Gene Pulser electroporation cuvette (Bio-Rad #1652089). The transformation mixtures were electroporated (1800 V, 200Ω, 25mF) using an Eppendorf Eporator®, and were then immediately transferred to 1.5mL Eppendorf tubes containing 600µL SOC media. The mixtures were incubated at 37°C, 190rpm for 1 hour, whereupon the total transformation mixture was plated onto LB-agar solid media supplemented with 1x antibiotics (Table 2.2). The transformations were finally spread and incubated, as previously described.

Antibiotic	1000x Stock Concentration	Supplier
Ampicillin sodium salt	100mg/mL in ddH_2O	SA #A0166
Kanamycin sulphate	50mg/mL in ddH_2O	SA #60615
Chloramphenicol	25mg/mL in ≥99.5% ethanol	SA #C0378
Spectinomycin dihydrochloride pentahydrate	50mg/mL in ddH_2O	SA #S4014

Table 2.2 - Antibiotics used in this study. The 1000x stock solution and solvent used for solubilisation and stability for each antibiotic are indicated. All antibiotics were stored at 4°C for a maximum of 1 month.

2.1.3 Primer Design

Oligonucleotide DNA primers were manually designed to amplify and mutate genes encoding proteins of interest (Appendix I). Forward primers were designed using a random TAATA or ATTAT sequence preceding either a 5' restriction site or region of interest, followed by at least 18 nucleotides corresponding to the cDNA of interest, and terminating with a 3' GC clamp. Reverse primers were designed similarly, except the reverse complement of at least 18 nucleotides at the 3' end of the cDNA of interest was used. For traditional cloning, the sequence of both primers was manually adjusted until the theoretical melting temperature were within 5°C of one another. The theoretical melting temperature (T_m), secondary structure, and likelihood of hairpin and primer dimer formation was then analysed using the OligoCalc: Oligonucleotide Properties Calculator *(Kibbe WA, 2007)*, NEB Tm calculator (NEB: https://tmcalculator.neb.com/#!/main), and DNA Oligos In Tubes webserver (Sigma-Aldrich (SA): https://www.sigmaaldrich.com/GB/en/configurators/tube?product=standard). Upon arrival, oligonucleotides were pelleted by centrifugation at 13,000 x g for 1 minute and were resuspended

to 100μM using the indicated volume of ddH_2O for stocks. 100μM primer stocks were then stored at -20°C until required.

2.1.4 Standard PCR

Standard Polymerase Chain Reaction (PCR) was performed to amplify genes of interest with engineered 5' and 3' restriction sites for traditional cloning. 25.5μL PCR reactions were performed using the Q5® High-Fidelity DNA Polymerase (NEB) using the volumes and parameters indicated in Tables 2.3 and 2.4.

Component	Volume (μL)
5x Q5 Reaction Buffer	5
10mM dNTPs	0.5
10μM F Primer	1.25
10μM R Primer	1.25
10ng/μL DNA Template	0.5
Nuclease-Free ddH_2O	16.25
Q5 High-Fidelity DNA polymerase	0.25

Table 2.3 - Q5 High-Fidelity DNA Polymerase buffer components for standard PCR reactions. The volume (μL) for each buffer component in a 25.5μL PCR reaction is indicated. All buffer components were stored at -20°C.

Cycles	Temperature	Time
1	98°C	30s
30	98°C	10s
	T_m	30s
	72°C	20s/insert of interest lenght in kb
1	72°C	2 minutes
∞	10°C	∞

Table 2.4 - The Q5 High-Fidelity DNA Polymerase thermocycler programme parameters. The number of cycles applied for a specific temperature and duration is indicated.

Successful PCR product formation was validated by agarose gel electrophoresis. Each sample was prepared by mixing 5μL of PCR product with 1μL 6x Loading Dye (NEB #B7024S). Samples were separated by molecular weight (MW), along with a 1kb on a hand-cast 1% (w/v) agarose gel in 1x TBE (130mM Tris, 45mM boric acid, 2.5mM EDTA, pH 7.6) supplemented with 1x SYBR™ Safe DNA Gel Stain (Invitrogen™ #S33102), with 60mL of 1x TBE running buffer at 60V for 50 minutes. A separate 3μL GeneRuler 1kb Plus DNA ladder (Invitrogen™ #SM1333) sample was included as a reference for DNA molecular weights. Agarose gels were imaged by UV fluorescence using a Vilber Fusion FX Spectra Chemiluminescence/Fluorescence Gel Doc System.

2.1.5 Site-directed Mutagenesis

Site-directed mutagenesis was employed to introduce site-specific mutations into genes encoding proteins of interest and was conducted using the QuikChange II kit (Agilent Technologies #200523). 25.5μL PCR reactions were prepared using the indicated volumes and parameters (Tables 2.5 and 2.6):

Component	Volume (μL)
10x Reaction Buffer	2.5
dNTP Mix	0.5
5μM F Primer	1.5
5μM R Primer	1.5
10ng/μL DNA Template	0.5
Nuclease-Free ddH_2O	18.5
Pfu Ultra HF DNA Polymerase (2.5U/μL)	0.5

Table 2.5 - QuikChange II buffer components for site-directed mutagenesis PCR reactions. The volume (μL) for each buffer component in a 25.5μL PCR reaction is indicated. All buffer components were stored at -20°C.

Cycles	Temperature	Time
1	95°C	30s
16	95°C	30s
	55°C	1 minute
	68°C	1 minute/template in kb
∞	10°C	∞

Table 2.6 - QuikChange II thermocycle programme parameters for site-directed mutagenesis PCR reactions. The number of cycles applied for a specific temperature and duration is indicated.

0.5μL of *DpnI* restriction enzyme was then added to each PCR product to digest the hemi-methylated template DNA. Digestions were conducted in a water bath for 1 hour at 37°C and were then transformed as previously described. Single colonies were isolated and inoculated into 6mL LB liquid media supplemented with the appropriate antibiotic, and cultures were grown overnight at 37°C, 190rpm. The following day, glycerol stocks of the overnight culture was prepared, as previously described, and the plasmid DNA was extracted and purified from the cultures using the Qiaprep® Spin Miniprep Kit (Qiagen #27106). Successful mutants were validated by DNA sequencing. To that end, 5μL of plasmid DNA was sent to Source Bioscience UK and was sequenced using either in-house vector backbone-derived primers or 5μL of 10μM pre-designed sequencing primer. The resulting DNA sequence was aligned with the pre-designed sequence using the freely available ApE software (*Davis and Jorgensen, 2022*). Multiple rounds of DNA sequencing were performed until 100% sequence coverage of the entire cDNA fragment of interest was achieved.

2.1.6 Traditional Cloning

Restriction enzyme-based cloning was employed for routine shuttling of cDNAs, encoding proteins or protein fragments of interest, between expression vectors (Appendices II and III). This procedure

followed the classic six key steps, involving the i) preparation of compatible sticky DNA ends, ii) ligation of compatible DNA fragments, iii) bacterial transformation and antibiotic selection of ligation products, iv) mini-prep based DNA extraction and purification of single clone plasmid DNA, v) restriction digest analysis of clonal plasmid DNA, and vi) DNA sequencing-mediated validation of positive clones.

To prepare the compatible vector DNA fragment, ~4μg of vector DNA was digested for 2 hours at 37°C in a water bath with 20U of appropriate restriction enzyme(s) (Table 2.7) in 50μL reactions containing 1x rCutSmart Buffer (NEB) (Table 2.8). After the 1st hour had passed, 1μL of Quick Calf Intestinal Phosphatase (CIP; NEB #M0525S) was added to the restriction digest mixture to prevent phosphorylation-dependent vector DNA re-ligation. Following the 2 hours digestion, 10μL of 6x Loading Dye was added to the mixture and 2x 30μL samples were loaded onto a 1% (w/v) agarose gel for electrophoresis-based molecular weight separation. Gels were visualised using a FastGene® Blue Light LED Transilluminator 470nm (GENEFLOW #G5-0718) and successfully cut vector DNA was excised and transferred to a pre-weighed 1.5mL Eppendorf tube. The cut vector DNA was then extracted and purified using the QIAquick® Gel Extraction Kit (Qiagen #28704). To prepare the compatible insert DNA fragment, 25μL PCR reactions were conducted, as previously described. The remaining 20μL of gel electrophoresis-validated PCR products were digested for 1 hour at 37°C in a water bath with 20U of appropriate restriction enzyme(s) in 50μL reactions containing 1x rCutSmart Buffer (NEB) (Table 2.9). The digested PCR products were then purified using the Qiaquick® PCR Purification Kit (Qiagen #28104). 2μL of compatible sticky-end DNA fragments were each quantified using a NanoDrop 2000 Spectrophotometer (ThermoFisher Scientific, TFS) by measuring the A_{260} multiplied by 50, and DNA purity was determined by a A_{260}/A_{280} measurement of between 1.8-2.0.

Restriction Endonuclease	Restriction Enzyme Cut Site (5'-3')	Supplier
BamHI-HF®	G↓GATCC	NEB #R3136S
EcoRI-HF®	G↓AATTC	NEB #R3101S
HindIII-HF®	A↓AGCTT	NEB #R3104S
KpnI-HF®	GGTAC↓C	NEB #R3142S
NdeI-HF®	CA↓TATG	NEB #R0111S
NcoI-HF®	C↓CATGG	NEB #R3193S
XhoI	C↓TCGAG	NEB #R0146S

Table 2.7 - Restriction endonucleases used in this study for traditional cloning. The restriction enzyme target cut site (5'-3') is indicated by an arrow. All restriction enzymes were stored at -20°C.

Component	Volume (μL)
Vector DNA (4μg)	x
Restriction Enzyme I (20U)	1
Restriction Enzyme II (20U)	1
10x rCutSmart Buffer	5
Nuclease-Free ddH_2O	up to a final volume of 50
Quick CIP (5U)	1 (after 1 hour incubation at 37°C)

Table 2.8 - Vector restriction digest buffer components used for traditional cloning. The volume (μL) for each buffer component in a 50μL vector DNA restriction digest reaction is indicated. Typically, the vector DNA volume (x) was 40μL, based on a concentration of 100ng/μL.

Component	Volume (μL)
Insert DNA	20
Restriction Enzyme I (20U)	1
Restriction Enzyme II (20U)	1
10x rCutSmart Buffer	5
Nuclease-Free ddH_2O	23

Table 2.9 - Insert DNA restriction digest buffer components used for traditional cloning. The volume (μL) for each buffer component in a 50μL insert DNA restriction digest reaction is indicated.

DNA ligations (Table 2.10) were performed by mixing 50ng of pure digested vector DNA with a 3-molar excess of pure insert DNA in 20μL reactions containing 1μL of Quick Ligase (NEB #M2200S), which were incubated at room temperature for 5 minutes. A separate non-insert ligation control was also conducted to semi-quantitatively monitor vector re-ligation events by the number of antibiotic-resistant colonies after bacterial transformation. Both ligation samples were transformed into NEB Stable, as previously described, and single potential positive colonies from vector:insert ligation samples were innoculated into 6mL LB liquid media supplemented with the appropriate antibiotic, and were grown for 16-18 hours at 37°C, 190rpm. Glycerol stocks of potential clones were made, as described previously, and plasmid DNA was extracted and purified from the overnight cultures using the Qiaprep® Spin Miniprep Kit. Analytical digests were conducted by digesting 5μL of plasmid DNA with the appropriate restriction enzyme(s) in 50μL reactions containing 1x rCutSmart Buffer for 1 hour at 37°C and separating the digested plasmid DNA by molecular weight-based gel electrophoresis. Positive clones were validated by identification of the expected DNA fragments molecular weights and were then further validated by DNA sequencing.

Component	Volume (μL)
2x Quick Ligase Buffer	20
Vector DNA (50ng, nmols)	x
Insert DNA (3nmols)	y
Nuclease-Free ddH_2O	up to a final volume of 20
Quick Ligase	1

Table 2.10 - Quick Ligase reaction buffer components used for traditional cloning. The volume (μL) for each buffer component in a 20μL ligation reaction is indicated.

2.2 Protein Biochemistry Techniques

2.2.1 Recombinant Protein Overexpression and Purification

Several colonies were isolated from transformation plates and were transferred aseptically either to a 150mL conical flask containing 50mL LB liquid media supplemented with the appropriate antibiotic (standard recombinant protein overexpression) or to three 500mL conical flasks each containing 150mL LB liquid media supplemented with the appropriate 1x antibiotics (overexpression of site-directed phosphoserine/threonine and Abk incorporated proteins). All cultures were incubated at 37°C, 250rpm for 16-18 hours in a New Brunswick Innova® 44 stackable incubator shaker (Eppendorf #M1282-0002). The following day, 1mL of overnight culture was transferred to a 1.6mL cuvette and the relative cell density was quantified at A_{600nm} through comparison with a 1mL LB liquid-media only control. For standard protein expressions, the overnight culture was transferred to 3x 2.5L TunAir™ full-baffle shake flasks (SA #Z710822) each containing LB liquid media, unless specified otherwise, and supplemented with the appropriate antibiotic to an inoculation A_{600nm} of ~0.05. For site-directed phosphoserine/threonine incorporated protein overexpressions, the overnight cultures were transferred to 10x 2.5L TunAir™ full-baffle shake flasks, each containing Terrific Broth liquid media (TB; Invitrogen ™ #11538916) and supplemented with the appropriate 1x antibiotics to an inoculation A_{600nm} of ~0.1. All flasks were incubated at 37°C, 250rpm until an A_{600nm} of either ~0.4 (site-directed phosphoserine/threonine and Abk incorporated protein expressions) or ~0.6-0.8 (standard protein overexpression) was reached, whereupon recombinant protein overexpression was induced by supplementation with either 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG; BioServ UK Ltd #BS-2102C; standard Tac/T7 promoter protein overexpression) or 0.2% (w/v) L-(+)-arabinose (standard ara promoter protein

overexpression; SA #A3256) or 1mM IPTG and 0.2% (w/v) L-(+)-arabinose (overexpression of site-directed phosphothreonine incorporated proteins) or 1mM IPTG, 0.2% (w/v) L-(+)-arabinose and 2mM O-phospho-L-serine (SA #P0878; overexpression of site-directed phosphoserine incorporated proteins) or 1mM IPTG, 0.2% (w/v) L-(+)-arabinose and 250μM (*S*)-2-Amino-6-((2-(3-methyl-3*H*-diazirin-3-yl)ethoxy)carbonylamino)hexanoic acid hydrochloride (Abk; Iris Biotech #HAA3110; overexpression of site-directed Abk incorporated proteins). Following inducer addition, flasks were incubated at either 18°C, 250rpm for 16-18 hours (standard protein overexpression) or at 37°C, 250rpm for 4 hours (phosphoserine/threonine and Abk incorporated protein overexpressions).

The post-induction bacterial cells were harvested by centrifugation for 20 minutes at 4000 x g, 4°C using an Avanti J-20 Floor Super Speed Centrifuge (Beckman Coulter) and J-Lite®JLA-10.500 fixed-angle rotor (Beckman Coulter), and bacterial pellets were resuspended on ice with 20mL/L culture Lysis Buffer supplemented with 0.5mg/mL lysozyme (SA #L-6876), 0.1mg/mL DNase I (SA #DN25), 1x cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail tablet (Roche #11836170001) per 10mL Lysis Buffer, and a further 1x PhosSTOP™ tablet (Roche #490684500) per 10mL Lysis Buffer for recombinant site-directed phosphoserine/threonine incorporated protein purifications. The resuspended cells were mixed together in a 1:1 volumetric ratio at this stage during purification of the Sae1:Sae2 heterodimers. Bacterial cells were lysed by disruption of the cell wall and plasma membrane by applying a full amplitude 23kHz sound waves in 10s on and 10s off cycles for 5 minutes using an Soniprep 150 Ultrasonic Disintegrator (MSE). Large cellular debris was separated from the cell lysate by centrifugation at 21000 x g, 4°C for 20 minutes using a JA25.50 fixed-angle rotor (Beckman Coulter). Meanwhile, 2mL of 50% (v/v) affinity bead 1 slurry were equilibrated by a 1 minute 13,000 x g centrifugation spin, removal of the 1mL supernatant and addition of 1mL

appropriate Lysis Buffer. The supernatant containing the cellular lysate was then transferred to a 50mL Falcon tube, wherein 2mL of pre-equilibrated affinity beads 1 were added per 3L of original culture and the Falcon tube was placed on a rotator at 4°C for at least 1 hour. The lysate:bead mixture was transferred to a 27mL Econo-Column® Chromatography Column (Bio-Rad #7371517), which was stoppered for 5 minutes in order to allow for the beads to settle at the bottom. The column stopper was removed and the column was then attached to a single-channel peristaltic pump (Cytiva #11340142) that was set at a flow rate of ~300mL/hr to allow for the flow through of non-bead attached proteins. Once the total lysate had flown through, the beads were washed twice with 20 bed volumes (BVs) of Lysis Buffer, which was supplemented with 10mM imidazole (SA #I2399) for recombinant His_6-tagged proteins. Recombinant proteins were eluted by either 5x 1mL Lysis Buffer supplemented with 25mM reduced L-Glutathione (SA #G4251; GST-tagged protein elutions) or 250mM imidazole (His_6-tagged protein elutions) or 10mM D-(+)-Maltose monohydrate (SA #63418; MBP-tagged protein elutions). Alternatively, the beads were resupended with 5mL of Lysis Buffer and transferred to a 50mL Falcon tube and the N-terminal affinity tag was cleaved through supplementation with 1% (v/v) of either recombinant GST-HRV-3C protease or ProTEV Plus (Promega #V6101). Cleavage reactions were left on a rotator for 16-18 hours at 4°C, whereupon the protein:bead mixture was reapplied to the 27mL Econo-Column and left to settle again, before collecting the pure untagged recombinant protein flow through. For dual-tagged purifications, 2mL 50% slurry of pre-equilibrated affinity beads 2 was added to the eluent, which was transferred to a 15mL Falcon tube and left to incubate on a rotator at 4°C for at least 1 hour. The protein:bead mixture was then transferred to a 10mL Poly-Prep® Chromatography Column (Bio-Rad #7311550) that was connected to the single-channel peristaltic pump set to a flow rate of 100mL/hr. The beads were washed with 10BVs of Lysis Buffer before eluting with 5x 1mL of Lysis Buffer supplemented

with 250mM imidazole. All imidazole eluted proteins were immediately dialysed into 1L of Lysis Buffer that absented imidazole using a 3mL 10kDa molecular weight cutoff Slide-A-Lyzer™ Dialysis Cassette (TFS #66382), while GST-tagged proteins that were eluted with reduced L-glutathione were desalted into Lysis Buffer using a 1.0-2.5mL PD-10 Desalting Column with Sephadex G-25 resin (Cytiva #17085101). Briefly, the latter desalting step involved removing the top cap, pouring off the storage solution, and removal of the top filter, before placing the desalting column in a 50mL Falcon tube. 5mL of Lysis Buffer was added 5 times to the desalting column and left to enter the resin, before discarding the flow through. On the fifth fill the desalting column was centrifuged at 1000 x g, 4°C for 2 minutes and the flow through was discarded before transferring the desalting column to a fresh 50mL Falcon tube. 2.5mL of GST-tagged recombinant protein was added to the column and allowed to enter the resin before the desalted protein was collected by centrifugation at 1000 x g, 4°C for 2 minutes.

Apart from site-directed phosphothreonine incorporated proteins, which were of low yield, all other purified recombinant proteins were concentrated to approximately 1mL using either a 5kDa (recombinant SUMO proteins) or 10kDa molecular weight cutoff Vivaspin® 20 Polyethersulfone (PES) Centrifugal Concentrator (Sartorius #VS2011 and #VS2001), before samples were separated by molecular weight/hydrodynamic size using a 5mL sample application loop and Superdex 200 10/300 GL gel filtration column (Cytiva #GE17-5175-01) connected to an ÄKTA go™ chromatography system (Cytiva #29383015), which was set at a flow rate of 0.3mL/min, incubated at 4°C, and pre-equilibrated with 50mL of the appropriate degassed Lysis Buffer. 1mL fractions were collected for 40mL elution programme, and fractions that corresponded to A_{280nm} peaks were collected for SDS-PAGE analysis. SDS-PAGE validated fractions containing pure protein were

concentrated again using either 5kDa or 10kDa molecular weight protein concentrators. Pure proteins were then quantified by UV spectroscopy using a NanoDrop 2000 Spectrophotometer by A_{280nm} measurement of 2µL protein sample relative to a Lysis Buffer control and calculation using the rearranged Beer-Lambert Law equation:

$$c = \frac{A}{\varepsilon l}$$

where *c* is the molar protein concentration (M), *A* is the absorbance at 280nm, ε is the molar absorption/extinction coefficient ($M^{-1}cm^{-1}$), and *l* is a path length of 1cm. The protein concentration was also converted to mg/mL, where necessary, by multiplication of the molar protein concentration by the predicted protein molecular weight in Daltons. Protein molecular weight and extinction coefficient values were predicted by inputting the primary sequence of a given protein into the Expasy ProtParam webserver (https://web.expasy.org/protparam/). Recombinant phosphoserine/threonine incorporated proteins were validated by Western blotting using either an anti-phospho-CK2 substrate (MBP-3C-Rad50 $_{585\text{-}766\ pS/T690}$-$His_6$), anti-$Sae1_{pS185}$ ($Sae1_{pS185}$), anti-$Chk2_{pT68}$ (MBP-3C-$Chk2_{pT68pT383}$-His_6) or an anti-$Chk2_{pT383}$ (MBP-3C-$Chk2_{pT68pT383}$-His_6 and MBP-3C-$Chk2_{210\text{-}531\ pT383}$-$His_6$) antibody, and mass spectrometry. All purified recombinant proteins (Appendix IV) were then snap-frozen in liquid nitrogen and stored at -80°C until required.

2.2.2 Purification of Fluorescent SUMO Proteins for Single Turnover *in vitro* SUMOylation Assays

Recombinant $SUMO1_{1\text{-}97\ S9C\ C52A}$ and $SUMO2_{1\text{-}93\ A2C\ C48A}$ proteins were purified, as described previously, to contain an N-terminally exposed thiol groups to label with thiol-reactive Invitrogen™ Alexa Fluor ™ 488 C_5 maleimide (TFS #A10254) and Invitrogen™ Alexa Fluor ™ 647 C_2 maleimide

(TFS #A20347) dyes, respectively. Each dye was solubilised to a final concentration of 25mM with dimethyl sulfoxide (DMSO; SA #472301) and was added to 100μM SUMO1/2 protein solution to a final concentration of 1mM. The dye:protein solutions were incubated in the dark at room temperature for 2 hours, whereupon the maleimide-thiol reaction was quenched upon addition of 5mM reduced L-glutathione. The excess DMSO, dye, and reduced L-glutathione was removed from the fluorescently labelled SUMO proteins using a 5kDa molecular weight cutoff PD SpinTrap™ G-25 desalting column (Cytiva #28918004). Briefly, this involved a short vortex of the desalting column followed by placing into a 1.5mL Eppendorf tube and centrifugation at 800 x g, 4°C for 1 minute to remove the storage solution. The storage solution was discarded and 400μL of SUMO Assay Buffer (20mM HEPES pH 7.5, 50mM NaCl, and 5mM $MgCl_2$) was added to the resin and the desalting column was spun at 800 x g, 4°C for 1 minute. This discard, equilibrate, and centrifugation cycle was repeated 4 further times, until 140μL of either fluorescently labelled $SUMO1_{1\text{-}97\ S9C\ C52A}$ or $SUMO2_{1\text{-}93\ A2C\ C48A}$ was added to the resin, and the desalting column was transferred to a fresh 1.5mL Eppendorf tube. The desalting column was centrifuged at 800 x g, 4°C for 2 minutes to obtain each fluorescent conjugate SUMO protein eluent. The degree of labelling (DOL) for each fluorescent SUMO conjugate was quantified using a NanoDrop 2000 Spectrophotometer by A_{280nm} and A_{493nm} or A_{651nm} measurement of 2μL protein sample relative to a SUMO Assay Buffer control and calculation using the following equations:

$$[SUMO] = \frac{A_{280nm} - (A_{max} \times CF)}{\varepsilon}$$

$$DOL = \frac{A_{max}}{\varepsilon' \times [SUMO]}$$

where A_{280nm} is the absorption value at 280nm, A_{max} is the absorption value at the indicated maximum wavelength, *CF* is the correction factor for each dye at A_{280} (Alexa Fluor 488 C_5-maleimide: 0.11 and

Alexa Fluor 647 C_5-maleimide: 0.03), ε is the predicted extinction coefficient for each SUMO protein (SUMO1: 4470 $cm^{-1}M^{-1}$ and SUMO2: 1490 $cm^{-1}M^{-1}$), ε' is the extinction coefficient for each dye (Alexa Fluor 488 C_5-maleimide: 72,000 $cm^{-1}M^{-1}$ and Alexa Fluor 647 C_5-maleimide: 265,000 $cm^{-1}M^{-1}$), and *[SUMO]* is the molar concentration of each SUMO protein. A DOL value between 2 and 10 was considered optimal, however fluorescent SUMO conjugates were further validated by SDS-PAGE analysis, and through the ability to be adenylated, thioesterified, transthioesterified to Ubc9 in a Sae1:Sae2-dependent manner, and conjugated to $RanGAP1_{398-587}$ during *in vitro* SUMOylation assays.

2.2.3 Synthetic Peptides

All synthetic peptides used in this study were purchased from external biotechnology companies, namely Biomatik (Ontario, Canada), Synpeptide (Shanghai, China), and Genscript (Oxford, UK), however the PBDtide was kindly gifted from the Yaffe laboratory (MIT, USA) (Table 2.11). Most of the synthetic non-phosphorylated and phosphorylated peptides were designed with N-terminal acetylation and C-terminal amidation, however the biotin-Rad50-pT690 peptide included an N-terminal biotin tag connected to an ε-aminocaproic acid linker (eahx) to validate published peptide-pulldown interactomics experiments (see section 4.1 and 4.3.2), and the γH2AX phosphopeptide absented C-terminal amidation as observed in the MDC1/γH2AX complex crystal structure (*Stucki et al., 2005*). Moreover, all purchased synthetic peptides were designed to meet the following criteria: ≥95% purity, standard trifluoroacetic acid (TFA) removal, peptide nitrogen content analysis, liquid chromatography-mass spectrometry (LC-MS) molecular weight analysis, and qualitative solubility analysis in ddH_2O, Dulbecco's phosphate-buffered saline (DPBS), and DMSO. All synthetic peptides were aliquoted in 1mg lyophilised batches and stored at -20°C upon arrival. For binding

experiments, each synthetic peptide was solubilised in the appropriate buffer and quantified by measuring the A_{280nm} of 2µL peptide solution relative to a buffer only control using a NanoDrop 2000 Spectrophotometer and calculating the molar concentration using the Beer-Lambert Law equation, as described previously. In contrast, 1mg of Biotin-Rad50-pT690 phosphopeptide was solubilised in 30µL 0.1M $(NH_4)HCO_3$ and then diluted to the desired concentration with PLK1 Lysis Buffer. The molar concentration of Biotin-Rad50-pT690 phosphopeptide solution was then estimated by the mass of peptide in the total peptide solution volume (mg/mL) divided by the molecular weight. Following binding experiments, any residual peptide solution was snap-frozen in liquid nitrogen and stored at -80°C.

Peptide	Sequence (N-C)	MW (Da)	ε ($M^{-1}cm^{-1}$)	Supplier
PBDtide	Ac-MAGPMQSpTPLNGAYKK-Am	1815.05	1490	Koch Institute
Rad50-pT690 #1	Ac-RVFQpTEAELAYKK-Am	1661.4	1490	Biomatik
Biotin-Rad50-pT690	Biotin-eahx-RVFQpTEAELQEVISD-Am	2182.4	-	Genscript
Rad50-pT690 #2	Ac-RVFQpTEAELAYKK-Am	1661.4	1490	Genscript
Rad50-T690	Ac-RVFQTEAELAYKK-Am	1581.4	1490	Genscript
MDC1-pT699	Ac-AYDLQApTQCFA-Am	1351.41	1490	Synpeptide
MDC1-pT719	Ac-AYEDEPpTQAFA-Am	1362.33	1490	Synpeptide
MDC1-pT752	Ac-AYEVLApTQPFC-Am	1362.48	1490	Synpeptide
MDC1-pT765	Ac-AYEDSEpTQPFD-Am	1422.34	1490	Synpeptide
Irak1-pT141+3Y	Ac-AYSSASpTFLYP-Am	1327.37	2980	Synpeptide
H2AX-pS139	Ac-GGKATQApSQEY	1332	1490	Synpeptide

Table 2.11 - Synthetic peptides used in this study. The amino acid composition (N-C) for each synthetic peptide is given, where a lower-case p indicates phosphorylation, Ac is N-terminal acetylation, Biotin-eahx is an N-terminal biotin tag connected to an ε-aminocaproic acid linker and Am is C-terminal amidation. The binding solution is provided, as well as the experimentally derived molar mass and molar extinction coefficient values that were used to estimate the peptide concentration in solution using the Beer-Lamber Law equation from A_{280nm} values that were observed by UV spectroscopy. The manufacturer for each individual peptide is provided. All lyophilised peptides were stored at -20°C, whereas peptide solutions were snap-frozen and stored at -80°C.

2.2.4 1D SDS-PAGE Analysis

10μL of all purified recombinant proteins were mixed in a 1:1 volumetric ratio with 2x Laemmli Buffer (125mM Tris-HCl pH 6.8, 4% (w/v) SDS, 20% (w/v) glycerol, 10% (v/v) 2-mercaptoethanol; TFS #125472500, and 0.004% (w/v) bromophenol blue; SA #B0126) in a 1.5mL Eppendorf tube. 20μL *In vitro* SUMOylation assay samples were mixed in a 1:1 volumetric ratio with 2x Non-reducing Laemmli Buffer (125mM Tris-HCl pH 6.8, 4% (w/v) SDS, 20% (w/v) glycerol, and 0.004% (w/v) bromophenol blue) in a 1.5mL Eppendorf tube to minimise loss of thioester-linked intermediates. All sample Eppendorf tube lids were pierced and then boiled by incubation at 95°C for 5 minutes. All samples, except *in vitro* SUMOylation assay samples, were run on mPAGE™ 4-12% Bis-Tris Precast gels (Millipore #MP42G12). *In vitro* SUMOylation assay samples were run on NuPAGE™ 4-12% Bis-Tris Mini Protein Gels. All samples were run with 3μL of either PageRuler™ Prestained Protein Ladder (TFS #26616) or Invitrogen™ SeeBlue™ Plus2 Pre-Stained Protein Standard (TFS #10174502) as a protein molecular weight standard. Loaded SDS-PAGE gels were inserted into either an Invitrogen™ XCell *SureLock*™ Mini-Cell (TFS #10093492) or Invitrogen™ Mini Gel Tank (*In vitro* Sumoylation Assay samples; TFS #15320424) and run in 1x MES Running Buffer (50mM MES, 50mM Tris Base, 3.47mM SDS, and 1mM EDTA) for 50 minutes at 160V by connection with a PowerPac™ Basic Power Supply (Bio-Rad #1645050). Sufficiently separated gels were transferred to a 100mL weighing boat and were stained for at least 3 hours with Quick Coomassie Stain (NeoBiotech #NB-45-00078-1L), prior to vigorous destaining with ddH_2O. 1D SDS-PAGE gels were visualised and imaged using a Vilber Fusion FX Spectra Chemiluminescence/Fluorescence Gel Doc System.

2.2.5 Phos-Tag SDS-PAGE Analysis

Phos-Tag SDS-PAGE analysis was conducted to quantitatively analyse the stoichiometry of phosothreonine incorporation at T690 within purified recombinant MBP-3C-Rad50$_{585\text{-}766\ pT690}$-His$_6$ phosphoprotein. An 8% (w/v) acrylamide/bis-acrylamide resolving gel was copolymerised with 50μM Phos-Tag™ acrylamide and 100μM $MnCl_2$ and hand-poured (8% (w/v) acrylamide/bis-acrylamide 37.5:1; Bio-Rad #1610148, 390mM Tris-HCl pH 8.8; SA #T1503, 50μM Phos-Tag™ acrylamide; Nard Institute Ltd # 304-93526 AAL-107S1, 100μM $MnCl_2$; SA #M3634, 0.1% (w/v) sodium dodecyl sulphate; SDS; SA #75746, 0.1% (w/v) ammonium persulphate; APS; SA #248614, and 0.1% (v/v) N,N,N',N'-tetramethylethane-1,2-diamine; TEMED; SA #17919) in a 1.0mm glass plate. A thin layer of pure isopropyl alcohol (TFS #10215390) was added dropwise to the top of the gel to maintain a horizontal gel surface, and the gel was allowed to polymerise for 1 hour at room temperature. After 1 hour, the isopropyl alcohol was poured off and the resolving gel was rinsed with ddH_2O, before gently blotting the excess liquid with a paper towel. A 4% (w/v) acrylamide/bis-acrylamide stacking gel (4% (w/v) acrylamide/bis-acrylamide, 125mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.1% (w/v) APS, and 0.1% (v/v) TEMED) was layered onto the resolving gel until overflowing, whereupon a 10-well comb was immediately inserted into the gel. The stacking gel was allowed to polymerise for 20 minutes at room temperature. The discontinuous Phos-Tag gel was then inserted into a Mini-PROTEAN Tetra Vertical polyacrylamide gel electrophoresis system tank (Bio-Rad #1658001EDU), and the tank was filled with 1x SDS-PAGE running buffer (25mM Tris pH 8.3, 192mM glycine, and 0.1% (w/v) SDS) before removing the comb. The purified recombinant MBP-3C-Rad50$_{585\text{-}766\ pT690}$-His$_6$ phosphoprotein was prepared as described previously and loaded into the first gel well. No molecular weight marker was added to any of the wells due to incompatibility with the Phos-Tag SDS-PAGE system, however the remaining wells

were loaded with 1x Laemmli Buffer to minimise band distortion during separation. The Mini-PROTEAN Tetra Vertical polyacrylamide gel electrophoresis system was connected to a PowerPac™ Basic Power Supply and run at 120V, room temperature for approximately 150 minutes. After successful separation of the dye front to the bottom of the gel, the gel was removed and stained with Quick Commassie Stain, as described previously, and imaged using a Vilber Fusion FX Spectra Chemiluminescence/ Fluorescence Gel Doc System. The stoichiometry or efficiency of phosphothreonine incorporation at Rad50-T690 was semi-quantitatively calculated by densitometry using Fiji image analysis software v2.16.0 (*Schindelin et al., 2019*), whereby the relative intensity of upper band was compared to the sum of the upper and lower band intensities, and expressed as a percentage:

$$Phosphothreonine\ Incorporation\ Efficiency(\%) = \frac{Upper\ band\ intensity}{Upper + lower\ band\ intensity} X100$$

2.2.6 Western Blotting for Detection of Recombinant Phospho-proteins

Western blotting was employed as an initial quality control to monitor successful site-specific incorporation of phosphoserine/threonine into recombinantly expressed phosphoproteins. Purified samples were prepared and separated on mPAGE™ 4-20% Bis-Tris Precast gels, as described previously. Sufficiently separated gels were removed and transferred to a 100mL weighing boat, before the gel was cut according to the molecular weight of the probed protein. The cut gel was inserted into a Western blot transfer sandwich that was assembled and inserted into a gel holder cassette for the wet transfer technique. From cathode to anode, the transfer sandwich was assembled by a sponge, Whatman® 3mm chromatography paper, SDS-PAGE gel, 0.45μm nitrocellulose membrane (TFS #88018), Whatman® 3mm chromatography paper, and a sponge. Each Western blot transfer sandwich component was equilibrated in 1x Transfer Buffer (25mM Tris pH 8.3, 192mM glycine, and 20% (v/v)

methanol; TFS #10245850). The gel holder cassette was inserted into a blotting electrode assembly module within the Mini-PROTEAN Tetra Vertical polyacrylamide gel electrophoresis system tank. The gel was transferred in 1x Transfer Buffer by applying a 28mA current for 1 hour at room temperature by connecting the gel tank to a PowerPac™ Basic Power Supply. After 1 hour, the nitrocellulose membrane was transferred to a 100mL weighing boat and successful electrophoretic transfer was validated by molecular weight marker transfer and staining for 5 minutes at room temperature with 1x Ponceau S Solution (0.1% (w/v) Ponceau S; TFS #161470100, and 5% (v/v) acetic acid; SA #A6283). Following sufficient staining, the 1x Ponceau S Solution was removed from the weighing boat, and the membrane was destained vigorously with 1x TBST (20mM Tris pH 7.6, 150mM NaCl, and 0.1% (w/v) polysorbate 20; Tween 20; SA #P1379). The membrane was then transferred to a 50mL Falcon tube and blocked for 1 hour at room temperature with Blocking Solution (5% (w/v) bovine serum albumin; BSA; SA #05470), and then incubated for a further 16-18 hours at 4°C on a roller in Primary Antibody Solution (Blocking Solution supplemented with primary antibody at the indicated final concentration; Table 2.12). The following day, the Primary Antibody Solution was transferred to a 15mL Falcon tube and stored at -20°C for future use, while the membrane was washed for 10 minutes at room temperature with 1x TBST. The membrane was incubated in Secondary Antibody Solution (Blocking Solution supplemented with secondary antibody HRP conjugate at the indicated final concentration; Table 2.12) for 3 hours at room temperature on a roller. The Secondary Antibody Solution was then discarded, and the membrane was washed thrice in 1x TBST for 5 minutes at room temperature. Finally, the membrane was transferred to a Vilber Fusion FX Spectra Chemiluminescence developer tray and exposed to 1mL SuperSignal™ West Pico PLUS Chemiluminescent Substrate (TFS #34580X4), before immediately imaging chemiluminescent bands using a Vilber Fusion FX Spectra Chemiluminescence/Fluorescence Gel Doc System.

Antibody	Species	Dilution Factor	Supplier/source
1° Anti-Phospho-CK2 Substrate Multi Monoclonal Mix	Rabbit	1:1000	Cell Signalling Technology #8738
1° Polyclonal Anti-Chk2-pT383	Rabbit	1:1000	Abcam #ab59408
1° Polyclonal Anti-Sae1-pS185	Rabbit	1:4000	Rockland #600-401-B24 (Gifted from Dr Alexander Lanz)
2° Polyclonal Anti-Rabbit Immunoglobulins horseradish peroxidase (HRP) conjugate	Goat	1:2000	Agilent Technologies Dako #P044801-2

Table 2.12 - Antibodies used to detect phosphorylated Rad50-pT690 and Sae1-pS185 incorporation this study. Each primary (1°) and secondary (2°) antibody is indicated with respective host species, as well as the dilution factor used for Western blotting. The manufacturer/source is also provided.

2.2.7 Mass Spectrometry Analysis for Detection of Site-Directed Phosphoserine/ Threonine Incorporation into Recombinant Phospho-proteins

Bottom-up proteomics was conducted to validate the site-specific incorporation of phosphoserine/threonine residues into recombinant phosphoproteins. Briefly, recombinant phosphoproteins were separated by 1D SDS-PAGE and stained with Quick Coomassie Stain, as described previously. The SDS-PAGE gel was destained with filter-sterilised ddH_2O and then moved to a laminar flow hood to minimise proteinaceous contamination of the gel. The protein band of interest was excised using a scalpel and was transferred to a fresh 1.5mL Eppendorf tube under aseptic conditions. The gel slice was then trypsinised and analysed by time-of-flight mass spectrometry, which was conducted by the University of Birmingham Advanced Mass Spectrometry Facility.

2.3 Interaction Studies

2.3.1 Analytical Size Exclusion Chromatography for Detection of Complex Formation

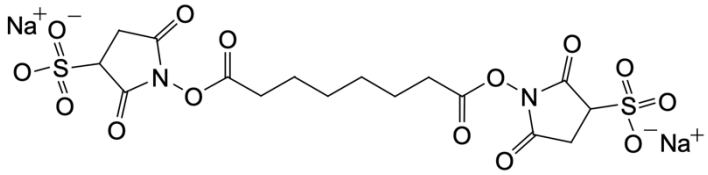
Gel filtration analysis was employed to assess the formation of heteromeric complexes between recombinant $PLK1_{367-603}$ (PBD) and CK2-phosphorylated $Rad50_{585-766}$ (pZCC). Purified PBD and pZCC proteins were mixed in a 1:1 molar ratio in PBD Lysis Buffer to ~248μM with a final volume of 400μL in a 1.5mL Eppendorf tube. The protein solution was incubated on a rotator at 4°C for at least 10 minutes, before injecting the solution into a 500μL sample application loop- yielding a ~200μM 1:1 protein solution. For BS^3 amine-to-amine crosslinking gel filtration analyses, a 500μL solution containing 25μM CK2-phosphorylated MBP-3C-$Rad50_{585-766}$-His_6 and 41μM PBD was also incubated at 4°C for at least 10 minutes, before injecting the solution into a 500μL sample application loop. Protein solutions were separated by molecular weight/hydrodynamic size using a Superdex 200 10/300 GL gel filtration column that was connected to an ÄKTA go™ chromatography system. The column was pre-equilibrated with 50mL of degassed PBD Lysis Buffer, set at a flow rate of 0.3mL/min, and incubated at 4°C. 1mL fractions were collected for a 40mL elution programme and fractions that corresponded to A_{280nm} peaks were collected for SDS-PAGE analysis.

2.3.2 Amine-to-Amine Cross-Linking Analysis for Detection of Complex Formation

Bis(sulfosuccinimydl)suberate (BS^3; TFS #A39266) and BS(PEG)9 (TFS #21582) amine-to-amine crosslinking analyses were performed to detect the formation of any complex between recombinant PBD protein and either MBP-3C-$Rad50_{585-766\ pS690}$ and CK2-phosphorylated MBP-3C-$Rad50_{585-766\ T/A690}$-His_6 proteins. BS^3 and BS(PEG)9 are homobifunctional crosslinkers (Figure 2.1) that contain amine-

reactive N-hydroxysuccinimide (NHS) esters with varying spacial reactivity, since BS^3 contains a spacer arm length of 11.4Å, while the BS(PEG)9 spacer arm length is 35.8Å.

BS^3



BS(PEG)9

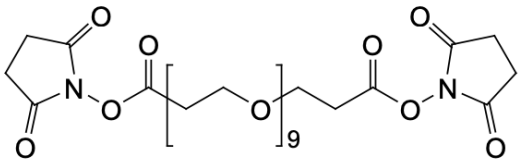


Figure 2.1 - Skeletal Structures of BS^3 and BS(PEG)9 homobifunctional amine-reactive crosslinking agents. Of note, the central PEG moiety in BS(PEG)9 is repeated 9 times.

BS^3 cross-linking analysis involved initial solubilisation of 2mg of the cross-linker in 277µL PBD Lysis Buffer, yielding a 12.5mM stock solution. 25µM recombinant CK2-phosphorylated MBP-3C-$Rad50_{585-766}$-His_6 proteins were each mixed with 41µM recombinant PBD and 2mM BS^3 in a 1.5mL Eppendorf tube to a total reaction volume of 450µL and incubated on a rotator at room temperature for 30 minutes. Following this, the reaction was quenched upon addition of 20mM Tris to a final volume of 500µL, and incubation on a rotator at room temperature for 15 minutes. BS^3 cross-linking of each individual protein was also conducted as a control for self-association. The cross-linked samples were then loaded onto a Superdex 200 10/300 GL gel filtration column, as previously described. Alternatively, BS(PEG)9 was solubilised in 465µL DMSO to obtain a 250mM stock solution, which was then added to a protein solution comprised of 5µM MBP-3C-$Rad50_{585-766\ pS690}$ or CK2-phosphorylated MBP-3C-$Rad50_{585-766\ T/A690}$-His_6 proteins and 25µM PBD to a final reaction concentration of 250µM. The cross-linking reaction was

again incubated on a rotator at room temperature for 30 minutes, and then quenched upon addition of 20mM Tris and incubation on a rotator at room temperature for 15 minutes. BS(PEG)9 cross-linking of each individual protein was also conducted as a control for self-association. The 20µL BS(PEG)9 samples were immediately analysed by SDS-PAGE analysis, as previously described.

2.3.3 Microscale Thermophoresis Binding Analysis

Microscale thermophoresis (MST) is an immobilisation-free biophysical technique used to quantify biomolecular interactions through the measurement of the directed movement of fluorescent biomolecules, such as a fluorescently labelled protein, in response to a temperature gradient generated by an infrared laser (*Jerabek-Willemsen et al., 2014*). This directed movement, or thermophoresis, typically results in the depletion of the fluorescent biomolecule from the heated region. When the IR laser heat source is deactivated, the fluorescent biomolecules return through mass diffusion (*Jerabek-Willemsen et al., 2014*). Consequently, the maximal change in average fluorescence intensity occurs between the 'hot' and 'cold' regions of the experiment (*Jerabek-Willemsen et al., 2014*). Moreover, the thermophoretic behaviour of a fluorescent protein depends on several intrinsic protein properties, including size, charge, and solvation energy. Upon binding to a non-fluorescent ligand, these properties are altered in a concentration-dependent manner (*Jerabek-Willemsen et al., 2014*). As a result, the thermophoretic signal of a fluorescent protein is also altered in a ligand concentration-dependent manner. Consequently, MST can precisely quantify the affinity (K_D) of protein-ligand interactions.

MST was employed for binding analysis between recombinant PBD and either MBP-3C-Rad50$_{585\text{-}766\ pT690}$-His_6 or CK2 phosphorylated MBP-3C-Rad50$_{585\text{-}766\ T/A690}$ proteins or PBDtide or Rad50-pT690 derivative

peptides, because the yield of phosphorylated MBP-3C-Rad50$_{585\text{-}766}$-His$_6$ proteins was insufficient for analysis by isothermal calorimetry. Recombinant PBD/PBD$_{H538AK540M}$ (Pincer) proteins were chosen as the desired fluorescent biomolecule for protein-peptide MST binding analysis experiments, whereas recombinant MBP-3C-Rad50$_{585\text{-}766}$-His$_6$ proteins were chosen as the desired fluorescent biomolecule for protein-protein MST binding experiments. Fluorescent amine-targeted labelling was achieved using the RED-NHS 2nd Generation Protein Labelling Kit (NanoTemper #M0-L011). Briefly, 10μg of RED-NHS 2nd Generation dye was reconstituted to 600μM with 25μL DMSO, of which 5μL was mixed in a 1:1 volumetric ratio with PBD Lysis Buffer. 10μL of 300μM was added to 90μL of 10μM recombinant PBD/Pincer or MBP-3C-Rad50$_{585\text{-}766}$ that was then incubated in the dark at room temperature for 30 minutes. Following this, the dye:protein solution was buffer-exchanged in PBD Lysis Buffer to remove excess dye using a 5kDa molecular weight cutoff PD SpinTrap™ G-25 desalting column. The degree of labelling (DOL) for each fluorescent protein conjugate was quantified using a NanoDrop 2000 Spectrophotometer by A_{280nm} and A_{650nm} measurement of 2μL labelled protein sample relative to a PBD Lysis Buffer control and calculation using the following equations:

$$[1]\ c(M) = \frac{A_{280nm} - (A_{650nm} \mathrm{x} 0.04)}{\varepsilon_{Protein}\ \mathrm{x\ d}}$$

$$[2]\ DOL = \frac{\mathrm{A650nm}}{195{,}000M - 1cm - 1}$$

where *d* is the given light pathlength of 1cm. A DOL value between 0.5 and 1.0 was considered optimal, however sufficient fluorescent labelling was further verified by a monodisperse pre-MST assay fluorescence intensity check value of between 200-1500 counts. The non-fluorescent titrant peptide was solubilised and diluted to ~200μM in PBD Lysis Buffer, while MBP-3C-Rad50$_{585\text{-}766}$-His$_6$ proteins were buffer-exchanged into 1L PBD Lysis Buffer using a a 3mL 10kDa molecular weight cutoff Slide-A-

Lyzer™ Dialysis Cassette at 4°C for 16-18 hours, and then concentrated with 10kDa molecular weight cutoff Vivaspin® 20 PES Centrifugal Concentrator to ~0.96-1.24mM.

MST-binding experiments were prepared by an initial 1 minute 13,000 x g centrifugation spin of the fluorescent protein and non-fluorescent titrant peptide/protein to remove any sample aggregates, and dilution of the fluorescent protein to 40nM in PBD MST Assay Buffer (PBD Lysis Buffer supplemented with 0.05% (w/v) Tween 20), which was added for sample homogeneity. 20μL of the non-fluorescent titrant peptide/protein was added to a 1.5mL Eppendorf tube. 16-fold Serial dilutions were conducted by pipetting 10μL of PBD MST Assay Buffer in 15x 1.5mL Eppendorf, and carefully pipetting and mixing 10μL of Eppendorf tube #1 with the 10μL PBD MST Assay Buffer of Eppendorf tube #2, and repeating this process until completion, however 10μL of the mixed Eppendorf tube #16 was aspirated and discarded. 10μL of the fluorescently labelled protein was then added to each tube and carefully mixed. Afterwards, a single Monolith Premium Capillary (MO; NanoTemper #MO-K025) was added to each Eppendorf tube to load the reaction mixtures by capillary action, and the capillaries were placed in a highest-to-lowest concentration order onto the Monolith NT.115 sample tray. MST binding experiments were performed using the MO. Control software v1.6 (NanoTemper #MO-S001C) and binding affinity routine using the following experimental parameters: MST-Power medium to high, excitation power of 20%, and thermostat setpoint of 22°C. All samples were quality controlled by automated pre- and post-MST assay fluorescence intensity capillary scans, MST trace shape analysis for sample aggregation and photobleaching-induced effects, and signal-to-noise ratio binding checks. The fluorescence cold and hot regions were automated by determination of signal-to-noise ratio, wherein a value of >5 was considered desirable, while >12 reflected an excellent assay. Raw fluorescence (relative fluorescence units; RFU), Fnorm (‰), and fraction bound curves (relative units; RU) were extracted from the MO.

Affinity Analysis software v2.3 (NanoTemper #MO-S001A) and analysed using a one site-specific binding model in GraphPad Prism v10.4.0. Of note, Fnorm describes the normalised relative fluorescence intensity in the heated state divided by the fluorescence intensity in the cold state between two specific timepoints.

2.3.4 Isothermal Titration Calorimetry Binding Analysis

Isothermal titration calorimetry (ITC) is an immobilisation-free and label-free biophysical technique used to quantify the thermodynamic parameters of biomolecular interactions through direct measurement of heat transfer associated with an interaction. These parameters include the affinity (K_D), enthalpy change (ΔH), and stoichiometry (n) at a constant temperature (T), which can be used to calculate the entropy change (ΔS) and Gibbs free energy change (ΔG) based on:

$$\Delta S = \frac{\Delta G - \Delta H}{T}$$

Upon binding to a ligand, a protein will either absorb or release heat, because of the formation of interactions and the disruption of pre-existing interactions that are inherent to proteins/ligand and with the surrounding solvent environment. ITC employs a microcalorimeter that detects the difference in heat between a reference cell containing water and a sample cell that are set at a desired experimental temperature. ITC also involves a syringe that titrates small volumes of a ligand into the sample cell. During binding events, the microcalorimeter will detect for heat differences both cells and apply a differential power to compensate for these heat changes, so that the thermal equilibrium is maintained. The differential power applied during a binding reaction is quantified and the extent of temperature compensation is directly proportional to the binding affinity of an interaction between a

protein and ligand. Consequently, the amount of heat transfer is maximal when the sample is non-saturated and diminishes as the protein sample becomes saturated with ligand.

ITC was conducted to provide a full thermodynamic characterisation of protein:peptide and protein:protein interactions, namely between recombinant Pellino $1/2_{FHA}$ proteins with $MDC1_{pTQxF}$ or $H2AX_{pS139}$ or $IRAK1_{pT141\ +3Y}$ phosphopeptides, recombinant $MDC1_{tBRCT}$ protein with the $H2AX_{pS139}$ phosphopeptide, recombinant PBD/Pincer protein with the derivative $Rad50_{pT690}$ phosphopeptides, and recombinant $Sae1_{pS185}$-His_6 phosphoprotein with recombinant SUMO2 protein.

For each interaction pair, the samples were extensively buffer exchanged in 1L of the appropriate Lysis Buffer at 4°C for 16-18 hours using either a 3mL 5kDa or 10kDa molecular weight cutoff Slide-A-Lyzer™ Dialysis Cassette for the recombinant protein and a 250μL 1kDa molecular weight cutoff Mini Dialysis Kit (Cytiva #GE80-6483-75) for the synthetic peptide. Dialysed samples were requantified by UV spectroscopy using a NanoDrop 2000 Spectrophotometer by A_{280nm} measurement of 2μL protein sample relative to a Lysis Buffer control. ITC binding experiments were conducted using a MicroCal PEAQ-ITC calorimeter (Malvern Panalytical #MAN0573-01). Typically, 40μL of 16μM recombinant protein was loaded into the sample titrant syringe and 280μL was carefully pipetted into the sample cell, however 50μM of recombinant $Sae1_{pS185}$-His_6 protein and 500μM of recombinant SUMO2 was used for the SUMO-interaction motif binding experiment. Binding measurements were conducted using the MicroCal PEAQ-ITC Control software v1.41 (Malvern Panalytical) at a thermostat setpoint of 25°C with a stir speed of 750rpm, and an average reference power of 41.9μW was used. In total there were 13 sample syringe injections for each binding experiment, wherein the initial injection was 0.4μL for 0.8s followed by 12 injections of 3.0μL for 6s. A 150s spacing was used between injections throughout the

course of the binding experiment. The sample titrant syringe and sample cell were extensively cleaned between assay runs. The raw heat curve and differential binding curve, as well as thermodynamic parameters, were calculated and plotted using the MicroCal PEA-ITC Analysis software v1.41 (Malvern Panalytical).

2.4 Biochemical Assays

2.4.1 *In vitro* Casein Kinase II Assays and Purification of Casein Kinase II Phosphorylated Proteins

Recombinant GST-CK2α was mixed in a 1:20 volumetric ratio with 1mg of either purified recombinant MBP-3C-Rad50$_{585-766}$-His$_6$ or untagged Rad50$_{585-766}$ proteins in CK2 Buffer (50mM Hepes pH 7.5, 150mM NaCl, 10mM $MgCl_2$) in 2mL Eppendorf tubes. 10mM Adenosine triphosphate (ATP; Roche #10127523001) was added to the CK2α/Rad50$_{585-766}$ protein solution to a final concentration of 500μM, which were then incubated in a water bath at 30°C for at least 2 hours. Following phosphorylation, 250μL of 50% (w/v) slurry Glutathione Sepharose 4B beads that were pre-equilibrated with Rad50$_{585-766}$ Lysis Buffer were added to the protein solution to bind to the recombinant GST-CK2α. The Eppendorf tubes were placed inside of a 50mL Falcon tube, which was incubated at 4°C on a rotator for at least 1 hour before transferring the phosphorylated solution to a 10mL Poly-Prep® Chromatography Column. The flow through was collected to obtain the pure phosphorylated recombinant Rad50$_{585-766}$ protein, which was validated by Western blotting using an anti-Phospho-CK2 substrate antibody.

2.4.2 Fluorescent Single Turnover *In vitro* SUMOylation Assays

Recombinant wild-type Sae1 and $Sae1_{pS185}$-containing SUMO E1 heterodimers, $SUMO1_{1\text{-}97\ S9C\ C52A}$ AlexaFluor 488 thiol-labelled conjugate, and $SUMO2_{1\text{-}93\ A2C\ C48A}$ AlexaFluor 647 thiol-labelled conjugate proteins were thawed on ice, while recombinant Ubc9 and $RanGAP1_{398\text{-}587}$ proteins were kindly provided by Dr Alexander Lanz and placed on ice. All samples were buffer exchanged, where necessary, into SUMO Assay Buffer using a 5kDa molecular weight cutoff PD SpinTrap™ G-25 desalting column, and re-quantified by UV spectroscopy, as described previously. 133µL *In vitro* SUMOylation assay master mixes containing either 100nM wild-type Sae1 or $Sae1_{pS185}$-containing SUMO E1 heterodimers, 500nM Ubc9, 10µM $RanGAP1_{398\text{-}587}$, and 20µM of either $SUMO1_{1\text{-}97\ S9C\ C52A}$ AlexaFluor 488 thiol-labelled conjugate or $SUMO2_{1\text{-}93\ A2C\ C48A}$ AlexaFluor 647 thiol-labelled conjugate proteins were constituted. 19µL of *In vitro* SUMOylation assay master mix was distributed across 6x 1.5mL Eppendorf tubes for each SUMO E1 heterodimer, while a 1mL 5mM ATP stock was kept on ice in a separate 1.5mL Eppendorf tube. *In vitro* SUMOylation assays were initiated upon addition of 1µL of 5mM ATP (250µM) to each 19µL sample and incubation at 30°C on a shaking heat block set to 500rpm. In contrast, the -ATP negative control sample was supplemented with 1µL of SUMO Assay Buffer. Samples were staggered across the time course for ease, which from first-to-last were: -ATP (10 minutes), 10 minutes, 5 minutes, 2.5 minutes, 1.5 minutes, and 1 minute. Sample SUMOylation reactions were quenched by addition of 20µL 2x Non-Reducing Laemmli Buffer and boiling at 95°C for 5 minutes. The samples were loaded onto 15-well NuPAGE™ 4-12% Bis-Tris Mini Protein Gel in the following order: molecular weight marker, -ATP, 1 minute, 1.5 minutes, 2.5 minutes, 5 minutes, and 10 minutes, and a single well gap was left between samples for each SUMO E1 heterodimer. The samples were separated by SDS-PAGE, as previously described, and then visualised using an AlexaFluor 488 or AlexaFluor 647 fluorescence-

specific wavelength filters provided by a ChemiDoc MP Imaging System (Bio-Rad #12003154). The gels were then stained with Quick Coomassie Stained, as previously described, to validate equivalent loading across all samples.

The raw fluorescence images were evaluated using Fiji image analysis software. Fluorescent images were uniformally thresholded across conditions and repeats by applying identical fluorescence intensity threshold values to minimise background intensity and maximise signal-to-noise ratio. The cumulative fluorescence intensity and individual fluorescence conjugates of each well was quantified. The fold change difference in cumulative fluorescence intensity values across wells were calculated, and each individual fluorescence intensity value was then scaled to the maximum cumulative fluorescence value to minimise pipetting error. RanGAP1 SUMOylation and total SUMOylation activity for each time point was then calculated using the following equations:

$$RanGAP1\ SUMOylation\ Conjugates = \frac{SUMO1/2 - RanGAP1_{398-597}\ Scaled\ Fluorescence}{Total\ SUMO1/2\ Scaled\ Cumulative\ Fluorescence}$$

$$Total\ SUMOylationactivity = 1 - Non - Conjugated\ SUMO1/2\ Scaled\ Cumulative\ Fluorescecnce$$

Values for each parameter were normalised and plotted in GraphPad Prism, and then fit to a Michaelis-Menten kinetics model, despite the limitation of using a single concentration of each of the *in vitro* SUMOylation components in this study. Initial relative RanGAP1 and Total SUMOylation rate values were estimated by linear regression analysis of the non-normalised intensity values from the first 2.5 minutes of the *in vitro* SUMOylation time courses. The relative proportion of 20μM SUMO conjugated within 1 minute was calculated from the estimated K_m value.

2.5 Structural Biology

2.5.1 X-ray Crystallographic Analysis of Pellino 1_{FHA}/$MDC1_{pT765}$ and PBD/$Rad50_{pT690}$ Complexes

Protein/phosphopeptide crystallisation proceeded by an initial pre-crystallisation test (PCT™; Hampton Research #HR2-143) of the complex across a range of 10, 15, 20, 25, and 30mg/mL recombinant Pellino 1_{FHA} or PBD with a 1.5 and 2.0 molar excess of $MDC1_{pT765}$ and Biotin-Rad50-pT690 phosphopeptides, respectively. The pre-crystallisation test was conducted according to the manufacturer's instructions, wherein 0.5mL of reagent A1 and 0.5mL A2 was pipetted into well A1 and well A2, respectively, of a VDX Plate with sealant per screened concentration. 500nL of each screened protein/phosphopeptide concentration was pipetted onto a single glass cover slide with 500nL of either reagent A1 or reagent A2. The cover slide was inverted and incubated at room temperature for 30 minutes, before the protein/phosphopeptide drop was analysed using a light microscope. Sample concentration drops that were clear or contained a light granular precipitate across either or both reagents were considered optimal for initial crystallisation screens.

Crystallisation was routinely conducted using the sitting drop vapour diffusion method using a NT8® v4 crystallisation robot (Formulatrix) and NT8 software (Formulatrix). ProPlex™ HT-96 (Molecular Dimensions #MD1-43), The LMB Crystallisation Screen™ HT-96 (Molecular Dimensions #MD1-99), Morpheus® HT-96 (Molecular Dimensions #MD1-47), and The Basic Chemical Space Screen HT-96 (BCS; Molecular Dimensions #MD1-105) screens were used for high-throughput 96-well initial screening of crystallisation buffers. 80µL of each initial screen condition were pipetted in a polystyrene MRC 96-well 2-drop crystallisation plate™ (Molecular Dimensions #MD11-00-100) using 200µL high

volume (HV) tips (Formulatrix #209584-B) under 95% humidity. 3µL of protein/phosphopeptide complex was added to rows A-H per column of a low volume disposable protein tray (Formulatrix #363080). 8x 200nL of protein/phosphopeptide sample was aspirated from each protein tray column, and a further 8x 200nL of each reservoir buffer condition column was aspirated and mixed with the protein sample using a low volume tip (Formulatrix #363040_B). 400nL of the protein/phosphopeptide/buffer solution was then aspirated as 8 drops per column. The low volume tips were extensively washed after each cycle, and this process was repeated until all 96 sitting drops contained 400nL protein/phosphopeptide in mother liquor. The above process was repeated using 2 columns of the low volume disposable protein tray and low volume tips for the PBD only and Biotin-Rad50-pT690 only double drop screens. Following this, the polystyrene MRC 96-well 2-drop crystallisation plate™ was sealed and incubated at 18°C. Protein/phosphopeptide crystallisation drops were analysed daily using a trinocular stereomicroscope with a 15x magnification objective.

2.5.1.1 Structure of the Human Pellino 1_{FHA}/MDC1_{pT765} Phosphopeptide Complex

The MDC1-pT765 phosphopeptide was mixed with the recombinant Pellino 1_{FHA} protein fragment in a 1170 µM:754 µM stoichiometric excess in Pellino Crystallisation Buffer (Pellino Lysis Buffer that absented 0.5mM) and crystallisation using the sitting drop vapour diffusion method was conducted as described above. Initial trials were performed using the ProPlex™ HT-96 crystal screen and microcrystals nucleated after 1 week in 1.5M $(NH_4)_2SO_4$, 0.1M Tris pH 8. An optimisation trial was conducted by varying the Pellino 1_{FHA}/MDC1_{pT765} phosphopeptide molar ratio (1:1, 1:1.5, 1:3, and 1:5) and supplementation of Pellino Crystallisation Buffer with 0.5mM TCEP, however the addition of TCEP inhibited crystal nucleation. A further trial was performed using the ANGSTROM Additive Screen™ (Molecular Dimension #MD1-100) in a 1:9 ratio with 1.67M $(NH_4)_2SO_4$, 0.11M Tris pH 8, whereby single

crystals nucleated and grew to ~400μm from drops consisting of 1.5M $(NH_4)_2SO_4$, 0.1M Tris pH 8, and 20% (v/v) PEG400. Crystals were cryoprotected by soaking in mother liquor supplemented with 25% (v/v) PEG400. Diffraction data were collected at the Diamond Light Source beamline I04 at a wavelength of 0.9795Å. The data were indexed and integrated by autoPROC v1.0.5 and XDS, scaled by AIMLESS v0.7.9, and analysed by POINTLESS v1.12.15 and STARANISO v2.3.94 (*Evans, 2006; Kabsch, 2010; Vonrhein et al., 2011,* and *Evans and Murshudov, 2013*). The data quality was analysed using Xtriage, indicating moderate anisotropy (*Zwart, Grosse-Kunstleve, and Adams, 2005*). The structure was solved by molecular replacement using Phaser-MR from the PHENIX suite v1.21.2-5419 with a search model comprised of coordinates from the recombinant human apo Pellino 2_{FHA} protein fragment structure (PDB accession code 3EGA; *McCoy et al., 2007; Lin et al., 2008; Adams et al., 2010; Tickle et al., 2018-2022*). The initial output model showed clear residual density for the $MDC1_{pT765}$ phosphopeptide at a 1s contour level, so the model was manually rebuilt and adjusted in Coot v1.1.10 from the CCP4 Program Suite v9.0.004 (*Emsley et al., 2010; Agirre et al., 2023*). The model was then refined using PHENIX.refine and validated in PHENIX using the comprehensive validation program, which includes MolProbity tools (*Echols et al., 2014; Williams et al., 2018*).

2.5.1.2 Structure of the Human PBD and Attempts at Crystallisation of a Potential Human PBD/Biotin-Rad50-pT690 Phosphopeptide Complex

The Biotin-Rad50-pT690 phosphopeptide was mixed with the recombinant PBD protein fragment in a 716μM:358μM stoichiometric excess in PBD Lysis Buffer and crystallisation was performed as described above. Initial trials were conducted using ProPlex™ HT-96, The LMB Crystallisation Screen™ HT-96, Morpheus® HT-96, and The BCS Screen HT-96 screens, and crystals of at least 3 different morphologies and varying sizes nucleated and grew in ~20% (84/384) of conditions screened. A

double-drop optimisation screen containing either 716μM Biotin-Rad50-pT690 phosphopeptide or 358μM recombinant PBD protein fragment was performed with the same high-throughput screens, as described above, and crystals of at least 2 different morphologies and varying sizes nucleated and grew in 90% (75/84) of initial hits for PBD only containing drops. Crystals from the remaining 9 conditions from the initial screen were cryoprotected by soaking the mother liquor with the indicated cryoprotectant (Appendix V). Diffraction data were collected at the Diamond Light Source beamline I24 (The LMB Crystallisation Screen™ HT-96 datasets from wells D1, D7, and G10) and the European Synchrotron Radiation Facility (ESRF) beamline MASSIF-3 (The LMB Crystallisation Screen™ HT-96 datasets from well F9). The data were indexed and integrated using either fast_dp or Xia2, including XDS, and scaled and analysed using AIMLESS and POINTLESS (*Evans, 2006; Kabsch, 2010; Evans and Murshudov, 2013,* and *Winter et al., 2018*). The structure was solved by molecular replacement using Phaser-MR from the PHENIX suite with a search model comprised of coordinates from the recombinant human PBDtide-bound PBD protein fragment structure (PDB accession code: 1UMW; *Elia et al., 2003;* McCoy et al., 2007; *Adams et al., 2010; Tickle et al., 2016*). All models absented residual density for the Biotin-Rad50-pT690 phosphopeptide. Nonetheless, the model was refined using PHENIX.refine and validated in PHENIX using the comprehensive validation program, which includes MolProbity tools (*Echols et al., 2014; Williams et al., 2018*).

2.5.2 Structural Model Predictions using AlphaFold

Protein:phosphopeptide complexes, namely the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide and PBD:$Rad50_{pT/E690}$ phosphopeptide complexes, were predicted using artificial intelligence-based models that are provided by the ColabFold v1.5.5: AlphaFold2 using MMSeqs2 (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scroll

To=kOblAo-xetgx) and AlphaFold 3 (https://alphafoldserver.com/) webservers (*Jumper et al., 2021; Mirdita et al., 2022; Abramson et al., 2024*). The indicated primary sequences (Appendix VI) were input into each webserver. The *num_relax* option was set to 5 in ColabFold to generate 5 models that underwent energy minimisation using the Amber ff14SB force field and were ranked according to Predicted Local Distance Difference Test (pLDDT), Inter-Residue Predicted TM-Score (iPTM), and Predicted Alignment Error (PAE) intrinsic metrics (*Maier et al., 2015; Jumper et al., 2021; Abramson et al., 2024*). The output models from the Pellino 1_{FHA}:$MDC1_{T765}$ ColabFold prediction was manually rebuilt in Coot to phosphorylate the $MDC1_{T765}$ peptide residue.

2.5.3 Visualisation and Analysis of Experimental and Predictive Structural Models

All experimental and predictive structural model coordinates were visualised in The PyMOL Molecular Graphics System v3.1.1 (*Schrödinger, LLC*) using .pdb files, and movies were generated using UCSF ChimeraX v 1.9rc202411140624 (*Meng et al., 2023*). The atomic coordinates from the Pellino 1_{FHA}:$MDC1_{pT765}$ AFM and AF3 predictive structural models were compared against the experimental crystal structure using the Amino Acid Sequence into Tertiary Structure (AS2TS) system Local-Global Alignment (LGA) webserver (http://proteinmodel.org/AS2TS/LGA/lga.html) (*Zemla, 2003; Zemla et al., 2005*). Global Distance Calculation (GDC) scores for all Pellino 1_{FHA} and $MDC1_{pT765}$ atoms were plotted in GraphPad Prism, while root-mean square deviation (RSMD), GDC of all and main chain interface-specific and pT765 residue-specific atoms, and the LGA scoring function (LGA_S) values were extracted to provide a comprehensive analysis of the structural similarity between the experimental crystal structure and AlphaFold predictive models (*Keedy et al., 2009*).

Chapter 3

Structural and Biochemical Characterisation of MDC1-pTQxF Mediated Recruitment of Pellino 1/2 to DNA Damage Sites

3.1 Introduction

3.1.1 Pellino E3 ligases

Pellino (Peli) proteins were first discovered through yeast two hybrid screens as an interaction partner with the *Drosophila melanogaster* protein Pelle, an orthologue of human serine/threonine-specific interleukin-1 receptor associated kinases 1 and 4 (IRAK1/4) (*Grosshans, Schnorrer, and Nüsslein-Volhard, 1999*). The Pelle:Peli interaction is an essential requirement for *Drosophila* innate immunity signalling in response to gram-positive bacterial infections, and is conserved throughout vertebrates and invertebrates (*Haghayeghi et al., 2010*). For example, all members of the human Peli family interact with IRAK1 and IRAK4 (*Jiang et al., 2003; Strelow, Kollewe, and Wesche, 2003; Huoh and Ferguson, 2014*). Pelis have since been implicated as central mediators of toll-like receptor (TLR) and interleukin-1 receptor (IL-1R)-mediated innate immunity pathways, as well as potentiating T-cell receptor (TCR) activation during adaptive immunity (*Jiang et al., 2003; Choi et al., 2006; Chang, Lin, and Sun, 2009*). More recently, Peli proteins have been identified as positive regulators of homologous repair (HR) in an ATM signalling-dependent manner during double strand break (DSB) signalling (*Ha et al., 2019*).

In humans, there are four members of the Pellino (Peli) family of RING-containing E3 ligases: Peli 1 and 2, as well as the alternative Peli 3a and 3b splice variants that are all encoded by at least 6 exons within chromosomes 2, 14, and 11, respectively (*Resch, Jockusch, and Schmitt-John, 2001; Jensen and Whitehead, 2003*). Peli 3a possesses 7 exons due to a 24-amino acid encoded exon insertion, located between exons 1 and 2 (*Jensen and Whitehead, 2003*). Pelis are characterised by the presence of an N-terminal atypical forkhead-associated (FHA) domain, responsible for specifically

binding to phosphothreonine (pT) protein epitopes, and a C-terminal RING domain, which acts as a molecular adaptor to facilitate the transfer of Ub from either chain-initiating or chain-elongating E2s to target substrates during the formation of either non-proteolytic K63-linked or proteolytic K11/K48-linked Ub chains (*Schauvliege, Janssens, and Beyeart, 2006; Lin et al., 2008; Ordureau et al., 2008,* Figure 3.1).

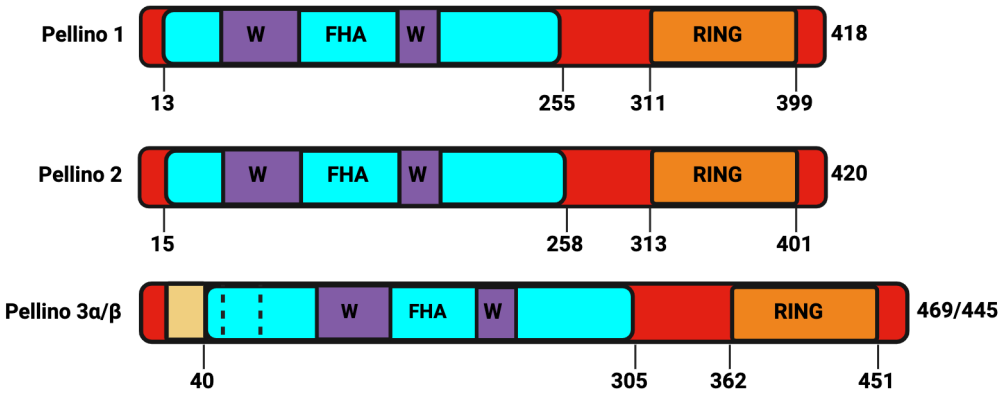


Figure 3.1 - The human Pellino family domain architectures. The N-terminal forkhead-associated domain, responsible for binding to pThr epitopes, is indicated in cyan, and the 'wing' inserts the split the domain is indicated in purple. The novel 27-amino acid N-terminal extension present in Pellino 3 isoforms is indicated in yellow. The 24-amino acid extension that is present in Pellino 3α, yet absent in Pellino 3β is indicated by the dashed lines. The C-terminal RING domain that provides Pellino family members with E3 ligase activity is indicated in orange. Residue positions that define the borders of each domain are located at the bottom. This figure was created with BioRender.com.

Peli family exhibit a high level of sequence conservation (Figure 3.2), with Peli 1 and 2 sharing 81.4% sequence identity, whilst both share 66.4-70.8% sequence identity with Peli 3 isoforms due to a unique 27 amino-acid N-terminal extension in both Peli 3 variants. The Peli 3a and 3b splice variants share 94.9% sequence identity, because the aforementioned exon insertion is absented in Peli 3b. Despite the high level of sequence conservation among Peli family members, the Peli family members are not functionally redundant. Specifically, this has been demonstrated by divergent FHA domain substrate preferences and specialised roles of different members within immune signalling pathways (*Huoh and Ferguson, 2014; Moynagh,2014*).

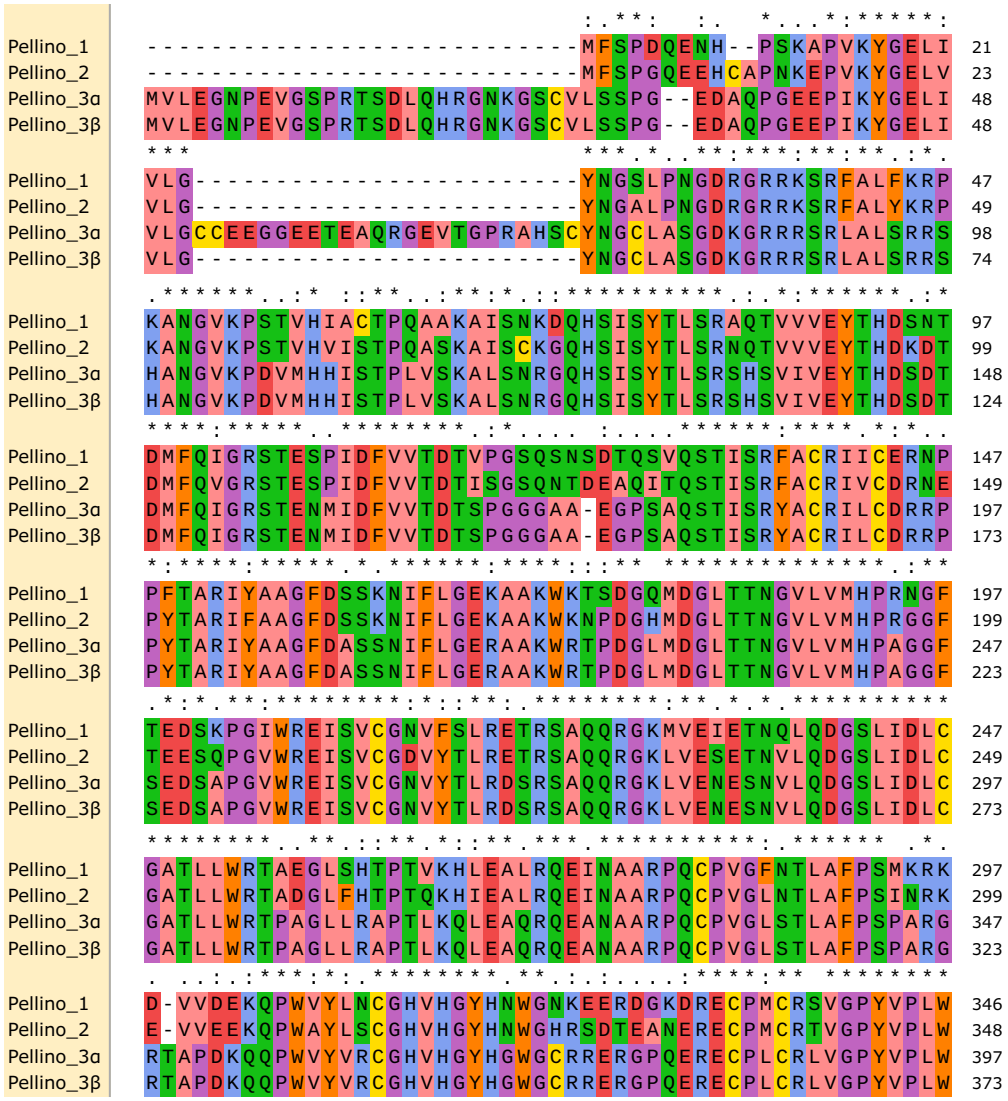


Figure 3.2 - Primary sequence alignment of the Pellino family. The primary sequence of human Pellino family members was aligned using the Clustal Omega algorithm (*Sievers et al., 2011*). Amino acids are coloured according to physicochemical properties (red- acidic, blue- basic, green- polar, salmon- hydrophobic, orange- aromatic, purple- conformationally special, and yellow- cysteine). Absolutely conserved amino acids are denoted with an asterisk (*****). Amino acids with similar properties are indicated with a colon (**:**), and weakly similar amino acids are indicated with a full stop (**.**). Figure created with SnapGene software v8.0 (www.snapgene.com).

3.1.2 The Pellino FHA Domain and Substrate Specificity

Following the initial discovery of Peli proteins, it was clear that they provide a structural scaffold framework during innate immune signalling cascades, however it was unknown how Peli scaffold activity was imparted at the molecular level. Whilst the previously mentioned C-terminal RING

domain in all Peli members was identified as a functional catalytic E3 module responsible for formation of K63 polyubiquitin chains on IRAK1, it was dispensable for mediating Pellino-IRAK1 interactions (*Lin et al., 2008*). Indeed, the Peli N-terminal fragment was responsible for binding to IRAK and other substrates, however the primary sequence of this fragment did not intuitively reveal any sequence homological clues (*Lin et al., 2008*). In 2008, Lin and colleagues determined the crystal structure of the Peli 2 N-terminal fragment at 1.8Å (Figure 3.3).

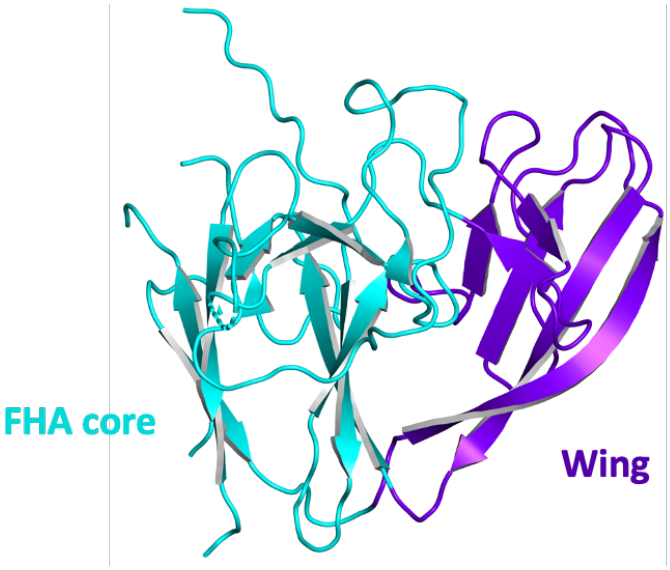


Figure 3.3 - The 1.8Å crystal structure of the Pellino 2FHA domain. PDB accession code: 3EGA from *Lin et al., 2008*. The secondary structure backbone is represented using the cartoon format. The FHA core and are unique 'wing' appendage are indicated in cyan and purple, respectively. Figure made using the PyMOL Molecular Graphics System v3.1.1. (*Schrödinger, LLC*).

Astonishingly, the structure revealed a single globular domain composed of 17 β-strands forming 4 β-sheets, of which 2 β-sheets ($\beta_{1\text{-}2}$-$\beta_{6\text{-}9}$-$\beta_{13\text{-}17}$) formed a β-sandwich that exhibited high structural similarity with canonical FHA domains (*Lin et al., 2008*). Interestingly, the remaining 2 β-sheets ($\beta_{3\text{-}5}$-$\beta_{10\text{-}12}$) split the core FHA domain as a structural appendage in an unprecedented way, thereby providing a molecular basis for the lack of intuitive sequence homology with other FHA *domains (Lin et al., 2008*). Structural and sequence homology analysis highlighted that the Peli 2 FHA (Peli 2_{FHA})

contains a core region that forms an extensive basic pocket containing highly conserved residues that are involved in pThr-substrate binding in canonical FHA domain counterparts (*Lin et al., 2008*).

Indeed, mutation of a highly conserved phosphate O_{γ}-contacting arginine, R106 to an alanine, or IRAK1 dephosphorylation by calf intestinal phosphatase (CIP) treatment, abrogated interactions between the Peli 2 FHA core and IRAK1, thereby demonstrating that the Peli 2 N-terminal fragment functions alike canonical FHA domains (*Lin et al., 2008*). Additionally, this FHA functionality was supported through R106 mutation and phosphorylation-dependent interaction studies involving Peli 2 and a *Saccharomyces cerevisiae* Rad9-derived phosphothreonine peptide (*Lin et al., 2008*). Further biochemical analysis using sequential IRAK1 truncation and T/A phosphonull mutants identified that Pelis interact specifically with phospho-IRAK1 at T141, and mutation of the pThr surrounding sequence greatly reduces binding affinity (*Huoh and Ferguson, 2014*). This result implied that, like canonical FHA domains, optimal Peli phosphorylation-dependent interactions are specifically determined by the local sequence of a central pThr residue (*Huoh and Ferguson, 2014*). Specifically, pThr local sequence screening experiments -using synthetic oriented peptide libraries- have demonstrated that the amino acid located 3 positions C-terminal to the pThr (pThr +3) moiety is a major determinant of FHA domain substrate-binding specificity (*Durocher et al., 2000*). In turn, FHA domains can be stratified into 4 major groups according to pThr +3 residue specificity: pTxxD, pTxxI/L, pTxxS/A, and pTxxY/M. Peli family member$_{FHA}$ placement according to these major groups, and apparent lack of functional redundancy within TLR/IL-1R signalling, was explored through binding experiments involving modulation of the +3-residue determinant within a synthetic IRAK1-pT141 phosphopeptide according to the major group preferences, namely +3D, +3I, +3S, and +3Y,

and pulldown experiments with known interactors involved in TLR signalling, including IRAK1, RIP1, and TRAF6 *(Huoh and Ferguson, 2014;* Table 3.1*)*.

Family Member	pT +3 Binding Preference	IRAK1	RIP1	TRAF6
Pellino 1	Y≈S≈I<<D	Y	Y	N
Pellino 2	Y<S<I<D	Y	N	Y
Pellino 3α	Y<S≈I<<D	N	N	Y
Pellino 3β	Y<S≈I<<D	Y	Y	Y

Table 3.1 - Pellino family members pT +3 and TLR substrate binding specifities. The pT +3 binding specificities were determined from fluorescence polarisation peptide binding assays using synthetic peptides derived from $IRAK1_{pT141}$. Similar pT +3 binding preferences are denoted with an approximately equal sign (≈), while a lower or an extremely lower binding preference is denoted with a single less than sign (<) and double less than sign (<<), respectively. TLR substrate binding to each Pellino family member were conducted by GST-Pellino family pulldown assays with Xpress-tagged substrates from HEK293T cell lysates. All results were taken from *Huoh and Ferguson, 2014*.

In all cases, +3D IRAK1-pT141 phosphopeptides displayed the weakest binding affinity, whilst +3Y were the tightest, however the affinity between Peli 1 and +3Y, +3I, and +3S were equivalent (*Huoh and Ferguson, 2014*). Peli 2, however, exhibited preference of Y<S<I<D, and Peli 3 splice variants bound +3S and +3I-derived peptides to a similar degree. Peli 1, 2, and 3b interacted with pTxxA/L/S/Y-containing IRAK1, whereas Peli 1 and 3b only interacted with pTxxL/M/S/Y-containing RIP1, and Peli 2 and 3a only specifically interacted with pTxxD/I/L/M-containing TRAF6 (*Huoh and Ferguson, 2014*). Thus, these findings provided an insight into the partial role of FHA domain binding preference within the context of tissue-specific E3 ligase activity of individual Peli isoforms.

3.1.3 Pellino Proteins in Innate Immunity Pathways

The innate immune system is the first and most efficient line of host defence during pathogen infection and host cell damage (*Li and Wu, 2021*). While the presence of physical or chemical

barriers largely limits host cell infection by pathogenic microorganisms, a second layer of defence is provided by non-specific immunosurveillance and immune responses provided by innate immune cells, including- but not limited to- monocytes, macrophages, neutrophils, natural killer (NK) cells, and eosinophils (*Li and Wu, 2021*). Although often regarded as a largely non-specific system, the innate immune response has evolved a specific class of pattern recognition receptors (PRRs) that specifically recognise conserved features of pathogenic microorganisms and factors released by damaged host cells, termed pathogen associated molecular patterns (PAMPs) and damage associated molecular patterns (DAMPs), respectively (*Janeway,1989; Wang et al., 1999*). Consequently, PRRs fundamentally allow innate immune cells to distinguish pathogens as non-self and recognise damaged host cells, through PAMPs and DAMPs, respectively, to elicit intricate proinflammatory signalling cascades that are carefully orchestrated through phosphorylation and ubiquitination-mediated regulation (*Liu, Qian, and Cao, 2016*). In turn, PRR-mediated signalling cascades result in activation of proinflammatory transcriptional programmes that stimulate the synthesis and release of proinflammatory cytokines and chemokines (*Li and Wu, 2021*). These molecules not only amplify the innate immune response through innate cell recruitment to sites of infection or injury, but also promote crosstalk with the adaptive immune response for a more specialised and coordinated response upon future infections (*Iwasaki and Medzhitov, 2015*). Additionally, chemokines and cytokines facilitate the resolution, removal and destruction of pathogens and damaged cells by phagocytic engulfment, release of antimicrobials, and activation of DNA damage repair and inhibition of pro-apoptotic pathways (*Vabulas et al., 2001; Kaiser and Offermann, 2005; Sumikawa et al., 2006; Brzostek-Racine et al., 2011*).

In general, the PRR family consists of five key players: membrane-bound toll-like receptors (TLR1-10) and C-type lectin receptors (CLRs; Dectin1-2), and cytoplasmic nucleotide-binding oligomerisation domain (Nod)-like receptors (NLRs; Nod1-2), retinoic acid-inducible-protein 1 (RIG-1)-like receptors (RLRs; RIG-1, MDA5, and LGP2), and absent in melanoma-2 (AIM2)-like receptors (ALRs; AIM2 and IF116) (*Li and Wu, 2021*). The TLR family are one of the first PRRs discovered and are the most extensively characterised. TLRs are bitopic integral membrane glycoproteins that consist of an N-terminal extracellular ligand-binding ectodomain (ECD), a single transmembrane helix that traverses either plasma (TLRs 1-2, 4-6 and 10) or intracellular organelle (TLRs 3 and 7-9) membranes, and a highly conserved cytoplasmic Toll-1L-1 receptor (TIR) domain *(Botos, Segal, and Davies, 2011)*. More specifically, the ECD adopts a solenoid structure formed by 19-25 tandem copies of 24 residue leucine-rich repeat (LRR) motifs that are involved in TLR ligand-induced 'm'-shaped homodimerisation and heterodimerisation at the plasma membrane, and subsequent TLR signal transduction through TIR dimerisation (*Botos, Segal, and Davies, 2011*). The ECD recognises specific PAMPs and DAMPs, such as lipopolysaccharides (LPS), heat shock proteins, histones, self/bacterial/viral DNA, and viral RNA (vRNA) through divergent homo- and heterodimeric ligand-binding pockets (*Chow et al., 1999; Asea et al., 2002; Hemmi et al., 2004; Xu et al., 2011*). The cytoplasmic Toll/IL-1 receptor (TIR) domain is highly conserved between TLRs and the interleukin-1-like receptor (IL-1R). TIRs form homotypic interactions through self-association and association with TIR-domain containing adaptor proteins, including myeloid differentiation primary response protein 88 (MyD88) and TIR domain-containing adaptor-inducing interferon-b (TRIF), to form signal transduction scaffolds (*Medzhitov et al., 1999; Xu et al., 2000*).

3.1.4 The Role of Pellino 1 in TLR3/4-specific NF-κB Proinflammatory Signalling

Peli 1 plays a central and specialised role in endosomal TLR3 and TLR4 TRIF-dependent signalling networks (*Chang, Jin, and Sun, 2009; Yan, Cui, and Feng, 2023;* Figure 3.4). Upon stimulation by either viral dsRNA (TLR3) or bacterial LPS (TLR4), both TLRs dimerise and recruit TRIF directly (TLR3) or indirectly through the TRIF-related adaptor molecule (TRAM; TLR4), and the RING domain-containing tumour necrosis factor receptor (TNFR)-associated factor 3 (TRAF3) (*Yamamoto et al., 2003; Häcker et al., 2006; Yan, Cui, and Feng, 2023;* Figure 3.4). K63-linked autoubiquitination or HECT3-mediated ubiquitination of TRAF3 at K138 stimulates recruitment and activation of the TANK-binding kinase 1 (TBK1)/I-kappa-B epsilon (IKKε) serine/threonine kinase complex, which in turn, activates Peli 1 E3 ligase activity through phosphorylation at S76, T288, and S293 (*Smith et al., 2011; Tseng et al., 2011; Li et al., 2018;* Figure 3.4). Activated Peli 1 catalyses K63-linked ubiquitination of receptor-interacting serine/threonine-protein kinase 1 (RIPK1) at K115, resulting in recruitment of the IKKα:IKKβ:NF-κB essential modulator/NEMO/IKKγ (IKK) complex, and TAK1:TAB1-3 complexes via ubiquitin-binding NEMO Optineurin Abin (NOA) and zinc finger domains, respectively (*Wang et al., 2001; Kanyama et al., 2004; Laplantine et al., 2009; Wang et al., 2017;* Figure 3.4). Proximity-driven TAK1-mediated phosphorylation of the IKKβ catalytic subunit T-loop at S177 primes IKKb autophosphorylation at S181, thereby triggering IKK-mediated phosphorylation of IκBα (*Zhang et al., 2014;* Figure 3.4). Under non-stimulated conditions, IκBα sequesters the proinflammatory gene transcription factor proteins, p50, p52, p65, RelB, c-Rel, Dosal, Dif, and Relish (NF-κB), within the cytoplasm through steric hindrance of the NF-κB nuclear localisation sequences (NLSs) (*Beg et al.,1992;* Figure 3.4). IKK-mediated IκBα phosphorylation at S32 and S36 within an N-terminal regulatory domain, however, triggers SCF-mediated K48-linked ubiquitination of IκBα, thus targeting IκBα for 26S-mediated proteasomal degradation (*Brown et al., 1995;* Figure 3.4). Consequently, the NF-κB NLSs are unmasked, allowing p50 and p65 to translocate to the nucleus and activate proinflammatory gene expression (Figure 3.4). Thus, Peli 1 is central to TLR3/4 activation of NF-κB-catalysed proinflammatory gene expression, as supported by Peli 1 knockout

mice studies. For example, Peli 1-deficient mice exhibit reduced NF-κB activation and proinflammatory gene expression, but also septic shock resistance (*Chang, Jin, and Sun, 2009*). Peli 3, however, acts as a negative regulator of TLR3 signalling and IFN-β transcription. Peli 3 knockout mice studies have highlighted that Peli 3-mediated K63-linked polyubiquitylation of the TRAF6 E3 ligase prevents interactions with and K63-linked polyubiquitylation of IRF7, resulting in reduced nuclear translocation of IRF7 (*Siednienko et al., 2012*).

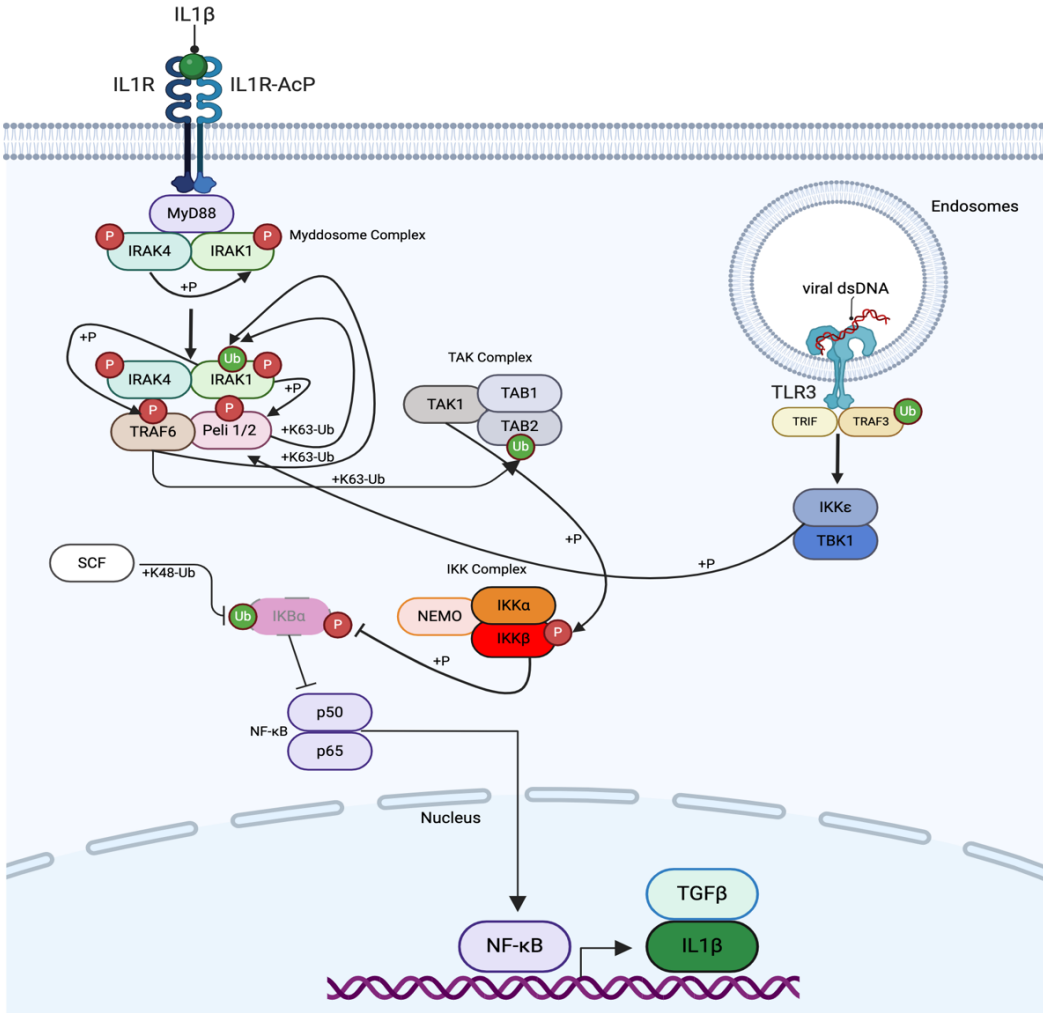


Figure 3.4 - Pellino 1/2 family members in TLR and IL-1R signalling cascades. Key phosphorylation events (+P) are denoted with a red P symbol. Key ubiquitin chain formation events that lead to complex formation (+K63-Ub) or target a substrate for degradation (+K48-Ub) are denoted with a green Ub symbol. Arrows denote complex transitions, post-translational modifications, and substrate translocation events. The arrow inside the nucleus indicates gene transactivation. Blunt-ended arrows indicate an inhibitory post translational modification event. A dashed and transparent colouring for a protein indicates that the protein is targeted for 26S proteasomal degradation. Adapted from *Yan, Cui, and Feng, 2023*.

3.1.5 Pellino Proteins in IL-1R Signalling

Peli 1 has also been implicated in the MyD88-dependent IL-1 proinflammatory cytokine pathway in macrophages, which results in activation of NF-κB and mitogen-activated protein kinase (MAPK) signalling (*Strickson et al., 2017;* Figure 3.4). IL-1b is a proinflammatory cytokine that is involved in both the innate and adaptive immune systems through recruitment activation of neutrophils and macrophages, and T and B lymphocytes, respectively. IL-1β is expressed as an inactive precursor that is activated by caspase 1/8-mediated processing, in response to TLR3/4 *stimulation (Li et al., 1995; Shenderov et al., 2014*). The IL-1β agonist binds to its cognate IL-1RI receptor with nanomolar affinity, stimulating recruitment of the IL-1 receptor accessory protein (IL-1RAcP) co-receptor, and subsequent TIR domain dimerisation-mediated activation (*Fields, Günther, and Sundberg, 2019;* Figure 3.4). MyD88 is recruited to the activated ternary complex TIR domain, which triggers recruitment of IRAK1 and IRAK4 serine/threonine kinase via N-terminal death domains (DDs) and thus formation of the Myddosome complex *(Wang et al., 2017;* Figure 3.1.2*)*. MyD88 mediates IRAK4 homodimerisation and transautophosphorylation, resulting in IRAK4 activation (*Wang et al., 2017*). Active phosphorylated IRAK4 concomitantly dissociates from the IRAK4 homodimer, and instead forms a heterodimer with IRAK1, which is necessary for IRAK1 primed activation via IRAK4-mediated phosphorylation of IRAK1-T209 (*Wang et al., 2017;* Figure 3.1.2). Full activation of IRAK1 is elicited upon T387 autophosphorylation at its activation loop, which may relieve N-terminal DD and C-terminal tumour necrosis factor (TNF) receptor-associated factor 6 (TRAF6)-binding motif (TB)-mediated autoinhibition (*Nguyen et al., 2009; Wang et al., 2017*). Successive hyperphosphorylation of the IRAK1 proline/serine/threonine-rich domain (ProST) causes dissociation of IRAK1 from the Myddosome complex (*Kollewe et al., 2004*). Active IRAK1 then recruits TRAF6 and mediates its

homodimerisation, whereupon TRAF6 E3 ligase activity is activated (*Jiang et al., 2002; Yin et al., 2009;* Figure 3.4). Importantly, active IRAK1 also associates with Peli 1 and 2, and perhaps Peli 3 splice variants, through phospho-dependent interactions between IRAK1-pT141 and the N-terminal Peli FHA domains (*Huoh and Ferguson, 2014;* Figure 3.4). Bound IRAK1 also triggers Peli E3 ligase activity through phosphorylation of Peli at S76, S78, T80, S82, T86, within a cryptic 'wing' insert that splits the FHA domain, and T288 and S293, located adjacent to the RING domain (*Smith et al., 2009*). Active TRAF6 and Peli both undergo autoubiquitylation and mediate K63-linked polyubiquitylation of IRAK1, which stimulates cytoplasmic translocation of the MyD88-IRAK4-IRAK1-Peli1-TRAF6 complex, in response to Peli and cellular inhibitor of apoptosis protein 2 (cIAP2)-mediated K48-linked proteasomal degradation of TRAF3 (*Ordureau et al., 2008; Walsh et al., 2008; Mao et al., 2010;* Figure 3.4). TRAF6 recruits TAK1:TAB1-3 bind to the cytoplasmic complex through the previously mentioned K63-linked ubiquitin-binding ZnF domain, and mediates K63-linked polyubiquitylation of TAB2, following linear ubiquitin chain assembly complex (LUBAC)-mediated M1-linked linear Ub chain formation (*Yan, Cui, and Feng, 2023;* Figure 3.4). The M1/K63-hybrid chain stimulates recruitment of the IKK *complex* (*Kanyama et al., 2004;* Figure 3.4).

As mentioned previously, TAK1 activates IKK-mediated phosphorylation and degradation of IκBα, culminating in activation of the NF-κB proinflammatory transcriptional programme (*Brown et al., 1995; Winston et al., 1999*). Additionally, active IKK/MAP3K7 is involved in MAPK and c-Jun NH2-terminal kinase (JNK) signalling, while the IKKβ subunit phosphorylates the interferon regulatory factor 5 (IRF5) transcription at S462, which stimulates IRF5 dimerisation and nuclear translocation; an important step during transcriptional activation of IL-12 and IFN-β cytokine genes (*Moriguchi et al., 1996; Zhang et al., 2009; Lopez-Pelaez et al., 2014; Kinsella et al., 2018*).

Whilst the above demonstrates that Peli proteins play key roles within IL-1R signalling, there is conflicting data surrounding the dispensability of Pelis. In support of a Peli dependency, anti-inflammatory TGF-β and bone morphogenic protein (BMP) stimulation induces expression of Smad6 and Smad7, which cooperate through mad homology 2 (MH2) domains to specifically sequester Peli 1 away from the Myddosome:TRAF6 complex, resulting in reduced NF-κB activation (*Choi et al., 2006; Lee et al., 2010*) . In contrast, IκBα phosphorylation and NF-κB transcriptional activation of proinflammatory *Tnf* and *IL-6* genes was unperturbed in Peli 1 knockout mouse embryonic fibroblasts (*Chang, Jin, and Sun, 2010*). However, some functional redundancy between Peli family members, especially Peli 2, and TRAF6 may account for these findings.

3.1.6 DNA Damage Response signalling at double-stranded DNA breaks

DSBs are the rarest (~ 50 spontaneous DSBs per cell per cell cycle) and most cytotoxic DNA lesions (*Vilenchik and Knudson, 2003*). DSBs occur when two complementary and proximal single strand phosphate backbone breaks occur, in response to endogenous and exogenous factors, such as ionising radiation (IR), acute and chronic exposure to DNA-damaging agents (clastogens) including chemotherapeutics, and upon replication fork collapse in response to replication stress (*Corry and Cole, 1968; Suzuki et al., 1970; Michel, Ehrlich, and Uzest, 1997*). Unresolved DSBs increase the proclivity for genomic instability through: DNA insertions/deletions, obstruction of DNA replication, chromosomal rearrangements, perturbed amphitelic sister chromatid segregation during mitosis, and chromothripsis (*Janssen et al., 2011; Stephens et al., 2011; Crasta et al., 2012*). Genomic instability culminates as changes in cell fate, including senescence or apoptosis commitment, or instead promotes tumourigenesis (*Elledge, 1996; Haper and Elledge, 2007*). Given that maintenance of genomic stability is essential to life, the cell employs the DNA damage response

(DDR) pathway to coordinate the detection of DSBs, arrest cell cycle progression, and recruit DNA repair factors (*Harper and Elledge, 2007*). Both phosphorylation and ubiquitination cascades are vital for DSB signal transduction and recruitment of DNA repair mediators, which regulate effector proteins, and in turn, determines cell cycle checkpoint activation, DNA repair pathway choice, and cell fate.

Typically, the DDR is initiated by DNA-damage induced recruitment and anchoring of the 2:2:1 heteropentameric Mre11/Rad50/NBS1 (MRN) complex to free blunted dsDNA ends and ssDNA overhangs within chromatin through a bridge formed by the globular Rad50 ATPase domain, comprised of the N- and C-terminal walker A/B ATP-binding motifs, and MRE11 bound to Rad50 (*de Jager et al., 2001*). The Rad50 and MRE11 heterotetramer DNA tether serves to increase the local concentration of DNA to stimulate Nbs1-mediated ATM recruitment (*Uziel et al., 2003; Dupré et al., 2006*). Nbs1, bound independently to MRE11, possesses a FxF/Y motif within a C-terminal ataxia telangiectasia mutated (ATM)-binding domain, which specifically interacts with the N-terminal ATM spiral domain (*Warren and Pavletich, 2022*). Upon recruitment, the dimeric and inactive ATM undergoes trans-autophosphorylation at S1981, coinciding with displacement of the autoinhibitory-state FRAP ATM TRRAP (FAT) domain from the C-terminal kinase domain, and dissociation of ATM into active monomers (*Warren and Pavletich, 2022*). Once activated in response to IR, active ATM serine/threonine kinases catalyses phosphorylation of over 700 proteins at SQ and TQ motifs, including the DSB signal-amplification serine/threonine checkpoint kinase 2 (Chk2) effector, the C-terminus of the minor histone variant H2AX at S139 (γH2AX) (*Matsuoka et al., 2000; Burma et al., 2001;Kolas et al., 2007;Matsuoka et al., 2007*). In particular, γH2AX is considered a *de facto* marker of DNA damage, as it can be visualised by immunofluorescence microscopy in IR-treated cells as

distinct punctate nuclear structures, termed ionising-radiation-induced foci (IRIF), because γH2AX spreads across DSB-contacting chromatin for 1-2Mb (*Collins et al., 2020*). γH2AX acts a focal point for the recruitment of downstream DDR proteins, including the IRIF-mediating scaffold protein, MDC1 through a H2AX-pS139-dependent interaction with a phospho-binding module comprised of 2 tandem BRCA1 C-terminal repeats (BRCTs) (*Stucki et al., 2005*).

3.1.7 MDC1 Acts as a Molecular Scafold in the DNA Damage Response

MDC1 is large modular protein that plays multiple roles as a molecular scaffold in the DDR (Figure 3.5) (*Stewart et al., 2003*).

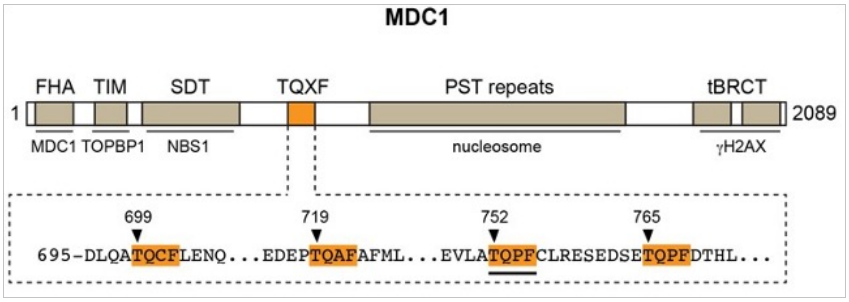


Figure 3.5 - The domain architecture of MDC1 and phosphosite clusters. The sequence of individual TQxF repeats is highlighted.

It is characterised by an N-terminal forkhead-associated domain that interacts in *trans* with the ATM-dependent T4 phosphorylation site to facilitate homodimer stabilisation in response to double-stranded DNA damage (*Liu et al., 2012*) and the C-terminal BRCT-repeat domain that interacts with H2AX-pS139 to facilitate DNA damage-dependent recruitment of MDC1 (*Stucki et al., 2005*). The intervening region is largely unstructured but contains a number of functionally important regions including a poorly understood set of 13 PST repeats that are involved in H2AX- and DNA damage-independent interactions with H2A and H2B nucleosomal acidic patches (*Salguero et al., 2019*). The remaining sequence contains several clusters of conserved, phosphorylated motifs. S168 and S196

(pS168 and pS196) function as TOPBP1-interaction motifs (TIM) that interacts with the TOPBP1 BRCT repeats 1 and 2 during mitosis (*Leimbacher et al., 2019*). This is followed by a cluster of 6 SDTD-[V/A/D]-[D/E] motifs (residues 218-456) constitutively phosphorylated by casein kinase 2 (CK2) isofiorms (*Chapman and Jackson, 2008*). MRN retention at DSB sites occurs via interactions between MDC1-pSDpTD sites and an NBS1 super-modular N-terminal FHA and tandem BRCT repeat region (*Lloyd et al., 2010*). Finally, the phosphorylated MDC1-TQxF (MDC1-pTQxF) cluster further recruits the DDR E3 ligase, RNF8, which contains an N-terminal FHA and C-terminal RING domains (*Kolas et al., 2008*). MDC1-TQxF specifically interacts pThr-dependently with the RNF8 FHA domain (*Huen et al., 2007*). RNF8 and UbE2N/Ubc13 mediate histone 1 (H1) K63-polyubiquitin chain formation which, in turn, is specifically recognised by RNF168 through the first of two ubiquitin-dependent DSB recruitment modules (UDM1) (*Thorslund et al., 2015*). The UDM1 can be further subdivided into 3 submodules: an LR motif 1 (LRM1), which is proposed to impart K63-linked Ub specificity, a Ub-interacting motif [UIM]-and MIU-related UBD (UMI), and a motif interacting with Ub 1 (MIU1) (*Thorslund et al., 2015*). RNF168 is also an E3 ligase with an N-terminal RING domain that mediates K27 polyubiquitin chains on K13 and K15 of H2AX to recruit further factors, such as 53BP1, BRCA1-BARD1, Rap80, and RNF169 (*Gatti et al., 2015; Becker et al., 2021*). Importantly, temporally regulated cell cycle-dependent kinase activity adds a further layer of regulation by promoting or inhibiting 53BP1 and BRCA1 recruitment, thereby governing whether damaged DNA is end-resected or not, and consequently, non-homologous end-joining (NHEJ) or homologous repair (HR) DNA repair pathway commitment (*Ceccaldi, Rondinelli, and D'Andrea, 2015*).

3.1.8 A Role for Peli Paralogues in the DDR

Recent evidence has emerged that implicates Peli 1 as a central phosphorylation-dependent DSB response E3 ligase involved in potentiating ATM signalling and downstream HR-mediated repair (*Ha et al., 2019*). This work suggested that Peli 1 localises to microirradiation-induced DNA damage foci and laser tracks in an ATM-dependent and apparent γH2AX-dependent manner through recruitment of Peli 1 via its FHA domain (*Ha et al., 2019*). Exogenous overexpression of GFP-tagged variants of the remaining 3 Peli family members also localised to laser tracks and DSB reporter sites. Crucially, Peli 1 DSB localisation was activated in response to ATM-mediated phosphorylation of S121 and T127 SQ/TQ motifs, located adjacent to predicted pThr-binding sites within the split FHA domain, and concomitant recruitment by a proposed Peli1FHA: γH2AX interaction mechanism (*Ha et al., 2019*; Figure 3.6).

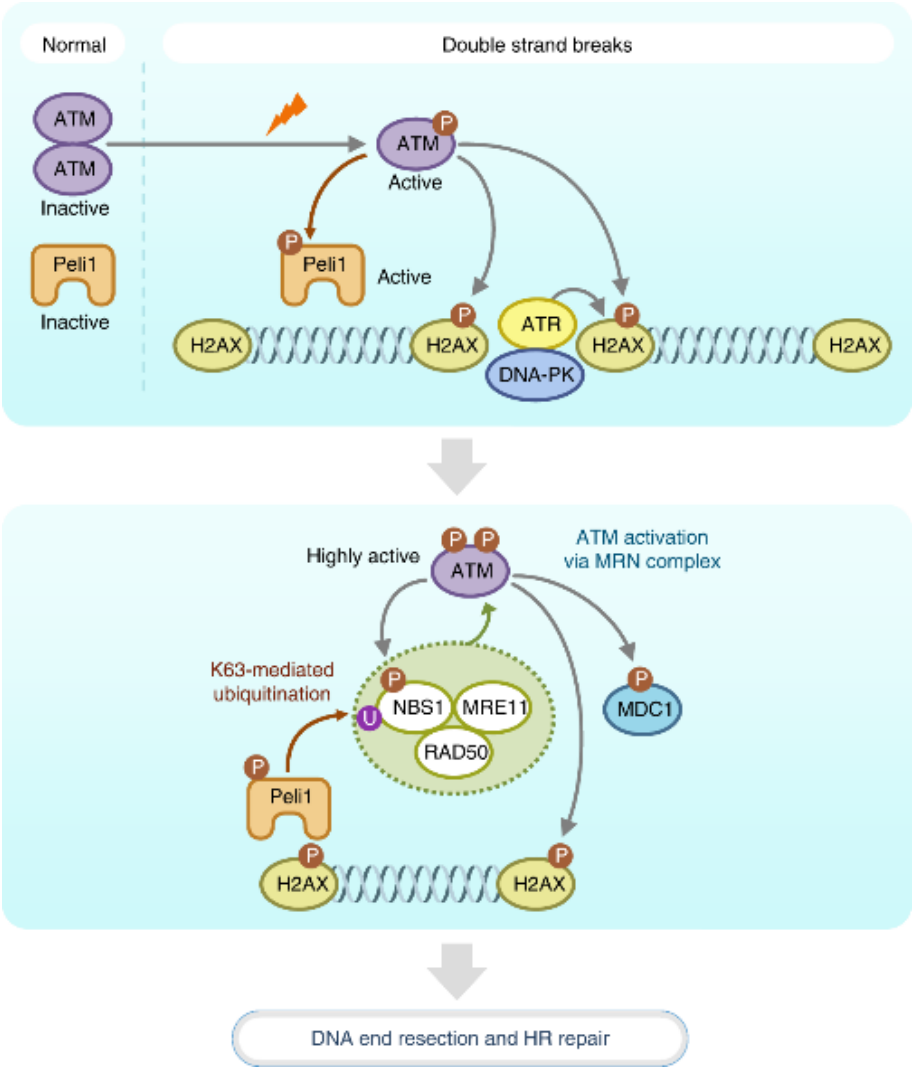


Figure 3.6 - Model of γH2AX-dependent Pellino 1 recruitment to DNA damage sites and Pellino 1-mediated DNA damage signalling. Under normal physiological conditions ATM is maintained as inactive dimers, while Pellino 1 is inactive and not recruited. In response to double-stranded DNA breaks (DSBs), indicated by a lightning bolt, ATM dissociates into active monomers that phosphorylates (red P symbol) H2AX and Pellino 1. Increased ATM activity is dependent on the MRE11-Rad50-NBS1 (MRN complex), which leads to ATM-dependent phosphorylation of MDC1 and the NBS1 subunit of the MRN complex. Phosphorylated Pellino 1 is recruited to DSBs in a phosphorylated H2AX-dependent manner, whereupon it stimulates K63-linked polyubiquitination (purple Ub symbol) of NBS1 to potentiate ATM signalling, and drive end-resection and homologous recombination (HR)-mediated DNA repair. Reproduced from *Ha et al., 2019*.

Peli 1 knockout mutant MEFs exhibited an increased number of mitotic cells with chromosomal breaks, whilst Peli 1 FHA truncation mutant MEFs displayed an increased sensitivity to IR, and a reduction in the total levels of DNA-end resection factors, including CtIP, Exo1, phospho-BRCA1,

Rad51, and RPA32 (*Ha et al., 2019*). Additionally, overall mono and poly-ubiquitination of substrates at DSB sites appeared reduced and recruitment of MRE11, Rad50, phosphorylated/activated-ATM, MDC1 and 53BP1 were also reduced in FHA truncation mutant MEFs. These findings were fundamental for implicating Peli recruitment of the MRN complex subunit, Nbs1, potentially through its central 2 tandem BRCT repeats and subsequent Peli RING-mediated K63-linked polyubiquitylation of Nbs1 at K686 and K690 (*Ha et al., 2019;* Figure 3.6). Nbs1 K63-polyubiquitination at these sites stabilises Nbs1, and functions as a positive ATM signalling potentiator to ultimately promote end-resection-mediated HR (Figure 3.6). In support of this, Peli 1 depletion led to a lower level of site-specific HR-mediated repair in response to DSBs, quantified using the GFP-*I-SceI* HR reporter assay (*Ha et al., 2019*).

The involvement of the Peli family as response E3 ligases within the DDR was unprecedented. In particular, it is striking that two members of E3 ligases, namely Peli 1 and RNF8, share similar domain organisation and functionality (Figure 3.7), however, potentially diverge with respect to potential DSB recruitment mechanisms, including γH2AX and MDC1-TQxF-phosphorylation-dependent interactions, respectively (*Huen et al., 2007; Ha et al., 2019*).

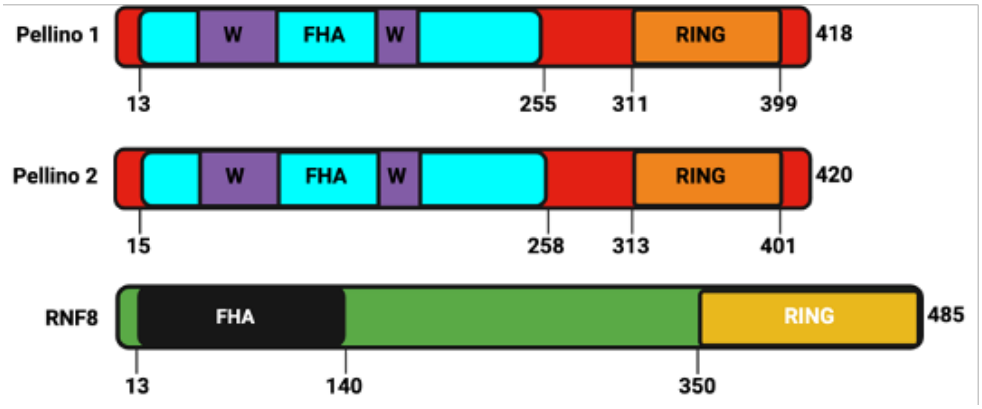


Figure 3.7 - Pellino and RNF8 E3 ligases share a common domain architecture. The N-terminal forkhead-associated domain (FHA), responsible for binding to pThr epitopes, is indicated in cyan for Pellino and black for RNF8, and the 'wing' inserts (W) that split the Pellino FHA domain is indicated in purple. The C-terminal RING domain that provides Pellino and RNF8 with E3 ligase activity is indicated in orange and yellow, respectively. Residue positions that define the borders of each domain are located at the bottom. This figure was created with BioRender.com.

Moreover, it appears that the presence of both DDR E3 ligases is not attributed to functional redundancy, because there are clear differences in substrates that are targeted by K63-liked polyubiquitin changes, specifically Nbs1 and histone 1, respectively, which are catalysed by UbE2N/Ubc13 and mediated by each respective C-terminal RING domain (*Thorslund et al., 2015; Ha et al., 2019*). In turn, it appears that the activity of Peli 1 and RNF8 are associated with differential downstream DSB repair pathway choice, with Peli 1 driving HR, and RNF8 involved in both HR and NHEJ (*Kobayashi et al., 2014; Ha et al., 2019; Chen et al., 2021*). This of course raised the exciting prospect that the recruitment and activation of either DDR E3 ligase, or subsequent commitment to a specific repair pathway choice, may be diametrically opposed through divergent DDR protein recruitment pathways. For example, Peli and RNF8 recruitment and activity may be regulated in response to specific DSB insults, through spatiotemporal regulation of kinases that provide distinct phosphorylation-dependent recruitment interfaces, and possibly that each respective E3 ligase may regulate the recruitment and/or activity of one another.

3.1.9 Limitations of the γH2AX-Dependent Pellino 1 Double-Stranded DNA Damage Recruitment Model

The apparent dependency of an interaction between γH2AX and Peli 1 within the proposed Peli DSB recruitment model was likely incorrect, because FHA domains do not bind to their respective ligands in a phosphoserine dependent manner. In support of this criticism, every single FHA:phosphorylation-dependent interaction characterised thus far- through extensive direct biochemical and structural interaction studies, involving a diverse range of eukaryotic and eubacterial recombinant FHA proteins and synthetic phosphopeptides, as well as synthetic combinatorial oriented phosphopeptide libraries- have continuously demonstrated that FHA binding modules display a strict specificity for binding partners containing at least one pThr epitope and not pSer, nor pY epitopes (*Durocher et al., 2000; Yaffe and Cantley, 2000; Pennell et al., 2010*). Moreover, molecular dynamic simulations have revealed that strict pThr epitope specificity is driven by a combination of factors, including reduced atomic flexibility of pThr peptides compared to pSer counterparts and a reciprocal reduced flexibility of conserved FHA phosphate-contacting residues upon complex formation, yielding a slight entropic advantage for pThr epitope binding over pSer (*Pennell et al., 2010*). Additionally, pThr epitopes confer an enthalpic advantage over pSer, due to a conserved FHA surface pocket that specifically accommodates the pThr -methyl moiety through weak van der Waals interactions upon phosphate docking. This, in turn, sterically hinders the pThr rotamer, and ultimately, stabilises the hydrogen bond and salt bridge network between conserved FHA residues and the pThr phosphate (*Pennell et al., 2010*). Despite this evidence, the Peli FHA domain, split by the bi-partite wing that is absent from other FHA family members, contains a more extensive basic surface mediated by the wing appendage, which could potentially have uniquely accommodated a pSer epitope (*Li et al., 2008*). Furthermore, direct Peli-binding experiments using

a phosphoserine peptide, nor pulldown assays with phosphoserine-containing physiological substrates have not yet been formally explored experimentally. Importantly, despite extensive attempts to biophysically characterise Peli FHA:phosphopeptide complexes, in order to resolve the residues involved in phospho-ligand recognition, no structures of such complexes had been solved.

Taking these criticisms and experimental unknowns into account, it appeared likely that γH2AX-dependent Peli recruitment may instead occur via an indirect mechanism involving a protein mediator that interacts with both Peli and γH2AX. For example, the DDR scaffold protein, MDC1, was an excellent candidate, because the C-terminal tandem BRCT repeats were known to be involved in γH2AX-mediated recruitment of MDC1, and the FHA domain of the highly similar DDR E3 ligase, RNF8, is recruited to MDC1 through phosphorylation-dependent interactions with a central cluster of 4 MDC1-TQxF motifs (*Stucki et al., 2005; Huen et al., 2007*). Moreover, experiments that investigated the relative localisation of GFP-tagged Peli 1 to DSB sites upon microlaser tracks, and in response to transient knockdown of MDC1, were hampered by partial siRNA knockdown of the endogenous MDC1 pool as observable by Western blotting (*Ha et al., 2019*). In turn, this result was interpreted as a lack of MDC1 involvement, and rather γH2AX, during recruitment of Peli 1, but a failure to completely ablate endogenous MDC1 levels would not rule out this possibility.

3.2 Preliminary Work Suggests an Alternative Pellino 1 DNA Double-Stranded Break Recruitment Model Involving the MDC1-TQxF Cluster

The work described in this chapter was initiated following experiments performed by our collaborator, Dr Mònica Torres-Esteban, from the Stucki lab at ETH Zurich investigating novel interactors of phosphorylated MDC1-TQxF motifs within the context of the DDR. The core

observation arose from unbiased MS peptide-pulldown screens involving N-terminally biotinylated synthetic MDC1-T752 peptides, corresponding to residues 743-757, with either the T752 position unmodified or phosphorylated (Figure 3.8). As expected RNF8 was identified as a MDC1-T752 phosphorylation-dependent interactor, thereby acting as a positive control for identification of novel MDC1-TQxF interacting proteins. Remarkably, both Peli 1 and Peli 2 were also identified as novel interactors (Figure 3.2.1C).

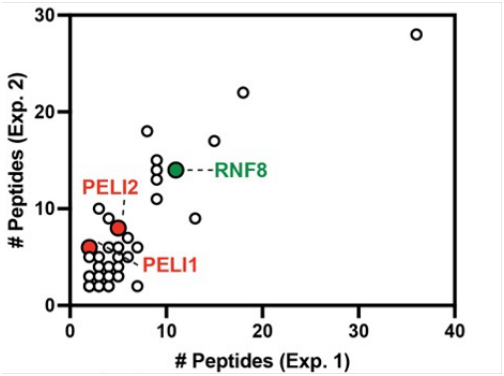


Figure 3.8 - MDC1-pTQxF peptide-pulldowns coupled with tandem mass spectrometry (LC/MS) proteomics identifies Pellino 1/2 (red) as novel interactors of MDC1-pTQxF motifs. The number of peptides uniquely isolated by N-terminally-biotinylated MDC1-T752 phosphopeptide pulldown is shown. RNF8 (green), a known MDC1-pTQxF interactor, serves as a positive control. These experiments were repeated two times. **Experiments were conducted by Dr. Mònica Torres Esteban (Stucki lab, ETH Zurich).**

The MDC1-pTQxF peptide pulldown experiments were then further expanded to include individual synthetic peptides corresponding to each of the 4 MDC1-TQxF motifs and the level of Peli 1/2 pulled down was compared to RNF8 levels (Figure 3.9A). This was to be able to determine whether there were any differences in MDC1-TQxF binding preferences between RNF8 and Peli proteins. Indeed, pT699 pulled down the highest levels of both RNF8 and Peli1/2, while pT752 appeared to preferably pulldown RNF8 and pT765 pulled down more Peli 1/2 than RNF8. This interaction was further supported by peptide pulldown experiments using both non-phosphorylated and phosphorylated synthetic T752 and T719 peptides, and N-terminally GST-tagged Peli 2 recombinant protein derived

from residues 1-275, corresponding to the previously solved FHA domain with a 14-amino acid N-terminal extension (Figure 3.9B). Again, these experiments demonstrated that the interaction between differential MDC1-TQxF motifs is both phosphorylation-dependent and involves the Peli FHA domain. Crucially, peptide pull down experiments with a γH2AX peptide were able to pulldown MDC1, but not Peli 1/2 (Figure 3.9C). Importantly, these experiments suggest that the previous Peli1 DSB recruitment model was incorrect, whereby Peli recruitment to DNA damage sites was instead dependent on ATM-mediated MDC1-TQxF sites, like RNF8.

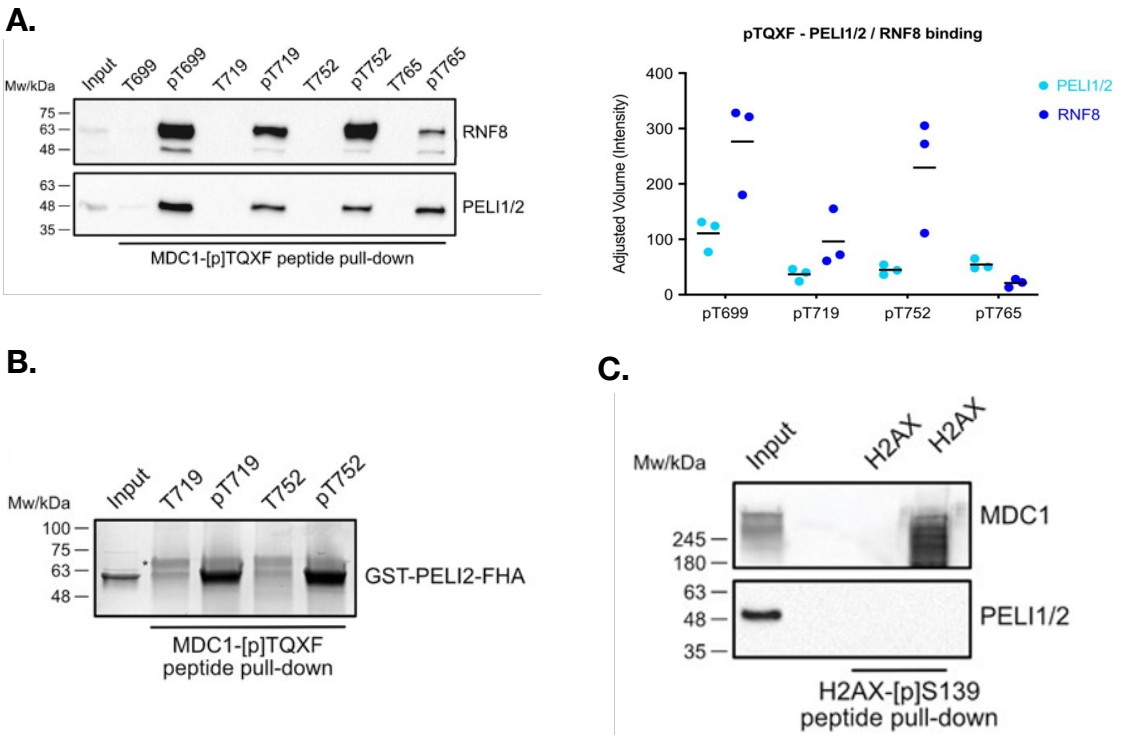


Figure 3.9 – Pellino 1 and 2 interact with MDC1-pTQxF phosphopeptides and not a γH2AX phosphopeptide. (A) All MDC1-pTQxF motifs (MDC1-T699, MDC1-T719, MDC1-T752, and MDC1-T765) interact with Pellino 1/2 proteins. Western blots of individual MDC1-pTQxF peptide pulldowns (left panel) and semi-quantification of MDC1-TQxF peptide pulldowns suggest preferential RNF8 binding (right panel) and Pellino 1/2 proteins exhibit minor MDC1-pT699 and MDC1-pT765 preferential binding. The adjusted volume mean intensity represents the mean band intensity of either RNF8 or Pellino 1/2. **(B) Western blots of MDC1-T719 and MDC1-752 pulldown assays with recombinant GST-Pellino 2$_{FHA}$ further support a phosphorylation-dependent interaction between Pellino 1/2 and MDC1-pTQxF motifs.** The upper band marked with an asterisk (*) denotes streptavidin elution. (C) A γH2AX phosphopeptide does not pulldown Pellino 1/2 proteins. **All experiments were conducted by Dr. Mònica Torres Esteban (Stucki lab, ETH Zurich).**

Following these initial observations, *in situ* proximity ligation assay (PLA) experiments were conducted. These experiments involved quantification of the relative levels of Peli 1/2:MDC1 PLA foci per cell under physiological conditions, and upon pre-treatment with 3 Gy of ionising radiation (IR) to examine the DNA-damage dependence of MDC1-TQxF and Peli1/2 interactions (Figure 3.10). In untreated cells, a basal level of Peli1/2:MDC1-TQxF interactions occur, which may be a result of basal ATM activation in response to transfection strategies (Figures 3.2.2B and 3.2.2C). However, the PLA signal intensity significantly increased upon IR treatment, thereby indicating that the interaction between Peli1/2 and MDC1-TQxF motifs occurs DNA damage-dependently in cells.

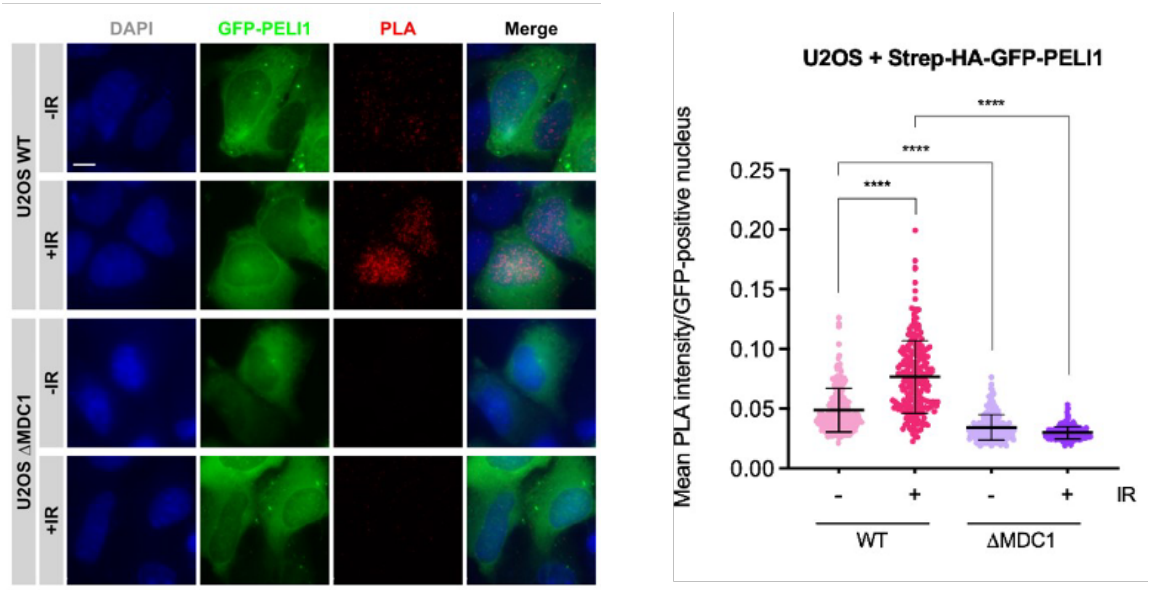


Figure 3.10 - Pellino 1/2 and MDC1 interact in cells in a DNA damage-dependent manner. PLA experiments (left panel) suggest that exogenous Pellino 1 protein interacts with endogenous MDC1 in a DNA damage dependent manner. A 10μm scale bar is indicated at the bottom. Quantification of PLA signal per DAPI-positive nuclei from in situ PLA experiments (right panel) further support a DNA damage-dependent interaction between exogenous Pellino 1/2 proteins and MDC1 in cells. **All experiments were conducted by Dr. Mònica Torres Esteban (Stucki lab, ETH Zurich).**

Cellular Peli DSB recruitment experiments were then conducted by inducing DNA damage through laser microirradiation, which would be observable as laser tracks (Figures 3.10A and 3.10B). The colocalisation of stably transfected N-terminally GFP/RFP tagged Peli 1 with endogenous MDC1, or

N-terminally GFP tagged MDC1, was monitored following CRISPR/Cas9-mediated knockout of the endogenous MDC1 pool, and subsequent stable transfection of exogenously expressed N-terminally tagged full-length wild-type MDC1, or a quadruple T-A phosphonull mutant (MDC1-AQxF). In addition, the localisation of γH2AX was also visualised as a marker of DNA damage, and to robustly analyse the Peli DSB site accumulation via γH2AX- or MDC1-dependent models. These experiments conclusively revealed that Peli accumulation at laser track DSB sites was dependent on the presence of intact phosphorylatable MDC1-TQxF motifs, while γH2AX DSB localisation was unperturbed in all conditions.

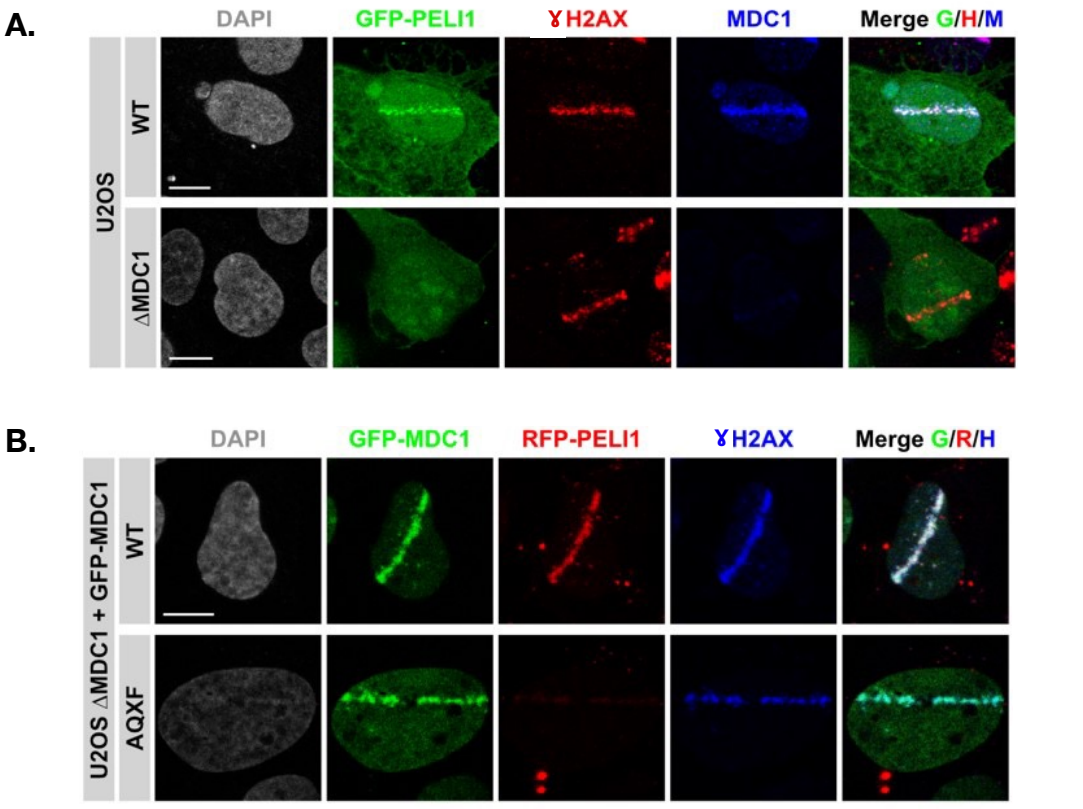


Figure - 3.11. Pellino 1 Recruitment to DSB sites is dependent on the phosphorylation status of MDC1-TQxF sites. (A) Pellino 1 recruitment to microlaser irradiation-induced DSB sites is dependent on MDC1. DAPI (greyscale), EGFP (green), Alexa Fluor 568 (red), and Alexa Fluor 700 (blue) emission wavelength filters were used. **(B) Pellino 1 recruitment to microlaser irradiation-induced DSB sites is dependent on intact MDC1-TQxF sites.** All scale bars are at the bottom of each DAPI image and represent 10μm. **All experiments were conducted by Dr. Mònica Torres Esteban (Stucki lab, ETH Zurich).**

3.3 Aims and Objectives

The findings of the preliminary experiments clearly indicate that Peli recruitment to DSB sites is dependent on an interaction between the Peli 1/2 FHA domain and intact phosphorylatable MDC1-TQxF motifs, and not through a direct interaction with the C-terminus of γH2AX. The goals of this project were to structurally and biochemically characterise the phosphorylation-dependent interactions between the $Peli_{FHA}$ domain and MDC1-TQxF motif. From a biochemical perspective, we sought to to provide quantitative measurements of the thermodynamic parameters that govern any direct binding between Peli 1/2 FHA domains and each MDC1-TQxF motif. Since the local sequence surrounding the TQxF core motif diverges significantly across individual MDC1-TQxF cluster repeats, ITC affinity measurements would provide key insights into Peli sequence binding preferences that locally surround pThr and pThr +3 binding determinants. From a structural perspective, we sought to solve the structure of a Peli-FHA:MDC1-pTQxF phosphopeptide complex using x-ray crystallography. This objective was fundamental to understand the molecular basis of substrate recognition by the Peli FHA domain, which is unique across FHA family members and has not been previously described. Consequently, a Peli:phosphopeptide crystal structure would elucidate differences in molecular recognition between archetypal FHA members, and the role, if any, of the cryptic wing inserts in substrate binding. Moreover, these findings could be applied to understand how Peli substrates involved in innate immune signalling pathways are recognised, and provide a solid foundation for the understanding of the molecular crosstalk between innate immune and DNA damage response signalling pathways.

3.4 Results

3.4.1 Recombinant Expression and Purification of Pellino 1/2-FHA and MDC1-tBRCT Domains

In the first instance, the *Peli 1* and *Peli 2* cDNA corresponding to amino acids 13-273 and 15-275, respectively, were cloned in-frame with an N-terminal GST-tag and succeeding 3C protease cleavage site (LEVLFQ↓GP) into a pGEX6P-1 bacterial expression vector (generated by Dr. Éilís Bragginton). The respective fragments, termed Peli 1_{FHA} and Peli 2_{FHA}, were chosen for expression and recombinant protein purification based upon 83% sequence identity between the region encompassing the Peli 1 fragment and a Peli 2 fragment that was previously used to solve the X-ray crystal structure of the Peli 2 FHA domain (*Li et al., 2008*). Additionally, the *MDC1* cDNA corresponding to amino acids 1883-2089 was also cloned in-frame into pGEX6P-1 for expression and purification of the MDC1 C-terminal tandem BRCT repeats domain ($MDC1_{tBRCT}$). All 3 proteins were initially purified by glutathione sepharose affinity chromatography and overnight 3C digestion, and then further purified by gel filtration chromatography to homogeneity at >90% purity (Figure 3.12). The typical yields were approximately 2.5-10mg/L culture. $MDC1_{tBRCT}$ peak elution volumes were between 2.7-3.0mL later than the Peli $1/2_{FHA}$ proteins, corresponding to the 2.2-2.4kDa larger predicted molecular weights, respectively (Figure 3.12B).

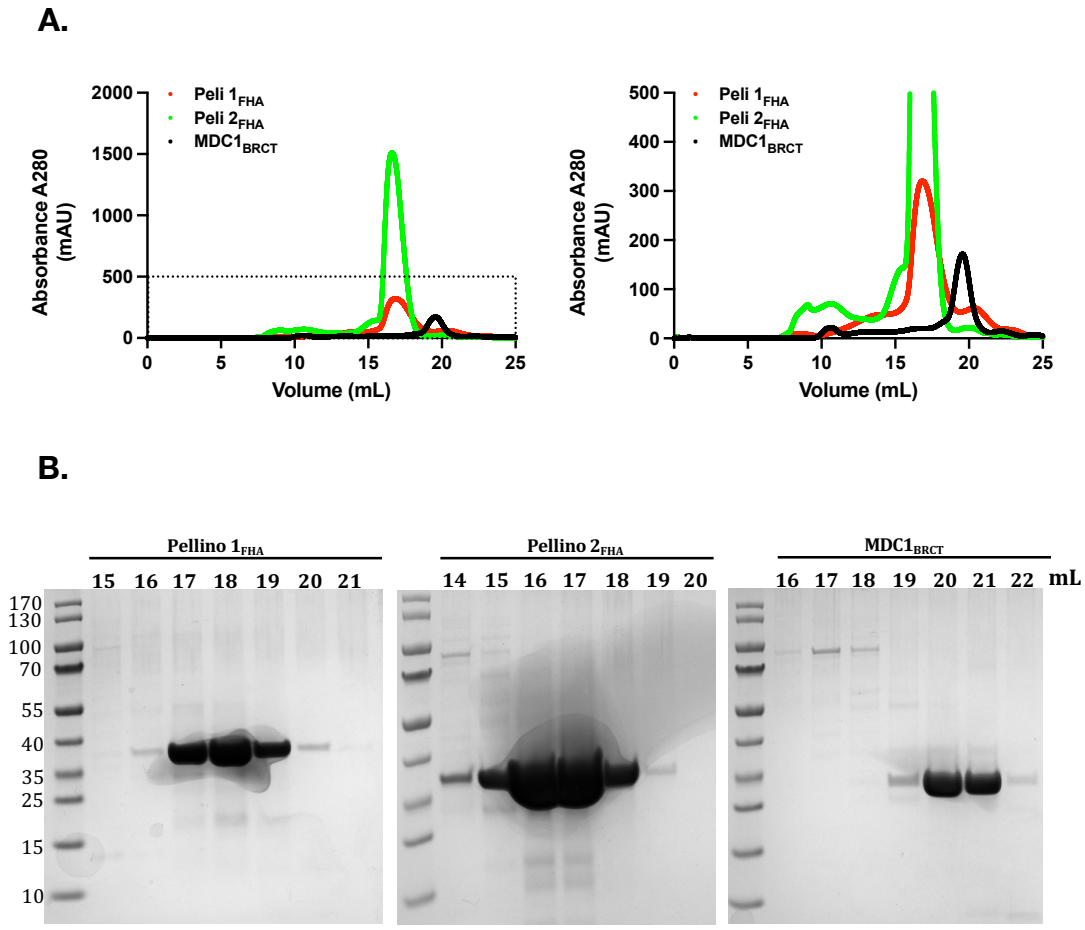


Figure 3.12 - Expression and purification of recombinant Pellino 1/2$_{FHA}$ and MDC1$_{tBRCT}$ domains. (A) Size-exclusion chromatography (SEC) elution profiles of purified Pellino 1/2$_{FHA}$ and MDC1$_{tBRCT}$ domains. The overall elution volume (mL) and ultraviolet (UV) absorbance profile at 280nm (mAU) is indicated for Peli 1$_{FHA}$ (red), Peli 2$_{FHA}$ (green), and MDC1$_{tBRCT}$ (black) domains from a Superdex 200 10/300 GL column (left panel). The dashed line denotes the area of the SEC profile that was used as a cutoff to observe Peli 1$_{FHA}$ and MDC1$_{tBRCT}$ profiles in (right panel). **(B) SDS-PAGE analysis of SEC elution peaks for Pellino 1/2$_{FHA}$ and MDC1$_{tBRCT}$ domains.** The fraction elution volume is indicated on the far right.

The purpose of the recombinant MDC1$_{tBRCT}$ fragment was to serve as a positive control for ITC binding measurements involving a N-terminally acetylated (Ac) synthetic γH2AX phosphopeptide (Ac-GGKATQApSQEY), because γH2AX is a major prerequisite for MDC1 recruitment to chromatin at DSB sites. Of note, the γH2AX phosphopeptide was not C-terminally amidated, because a free C-terminus is required for hydrogen bonding with the basic guanidinium side chain moiety of MDC1-

R1933 *(Stucki et al., 2005*). Individual N-terminally acetylated and C-terminally amidated MDC1-pTQxF and IRAK1$_{pT141+3Y}$ phosphopeptides were also synthesised (Table 2.2.3). The IRAK1$_{pT141+3Y}$ phosphopeptide was included as a positive control for Peli interaction experiments, since a synthetic IRAK1$_{pT141}$ phosphopeptide containing a Y residue in the +3 pThr position is the tightest known interactor of all Peli FHA domains, and a pThr +3Y residue is structurally similar to MDC1-TQxF +3F (*Huoh and Ferguson, 2014*). Both MDC1$_{pT699}$ and MDC1$_{pT719}$ synthetic phosphopeptides contained pThr +4A residues and not the respective physiological +4L and +4M residues, to aid solubility and monodispersity during synthesis. Importantly, all MDC1-pTQxF and IRAK1 $_{pT141+3Y}$ also included a synthetic N-terminal AY linker motif. The A was to prevent steric hindrance imposed on the Y, which itself aided peptide concentration quantitation using the Beer-Lambert Law based upon absorption of UV light at 280nm (A_{280}).

3.4.2 Pellino 1/2-FHA Domains Directly Bind to MDC1-pTQxF Phosphopeptides and Not a γH2AX Phosphopeptide

ITC binding experiments conducted using Peli1/2$_{FHA}$ and IRAK1$_{pT141+3}$ phosphopeptide clearly demonstrated that our recombinantly purified fragments are functional phospho-binding modules with K_Ds within a 3-4 -fold range of previous measurements (*Huoh and Ferguson, 2014*; Table 3.4.2). This slight difference may be associated with differences in the presence/absence of fluorescent biomolecules between quantitative binding experimentation workflows (*Huoh and Ferguson, 2014*). Like other FHA family member interactions with their corresponding phospholigands, the interaction between Peli 1/2$_{FHA}$ proteins and MDC1-pTQxF phosphopeptides is direct and enthalpically driven (Figure 3.13).

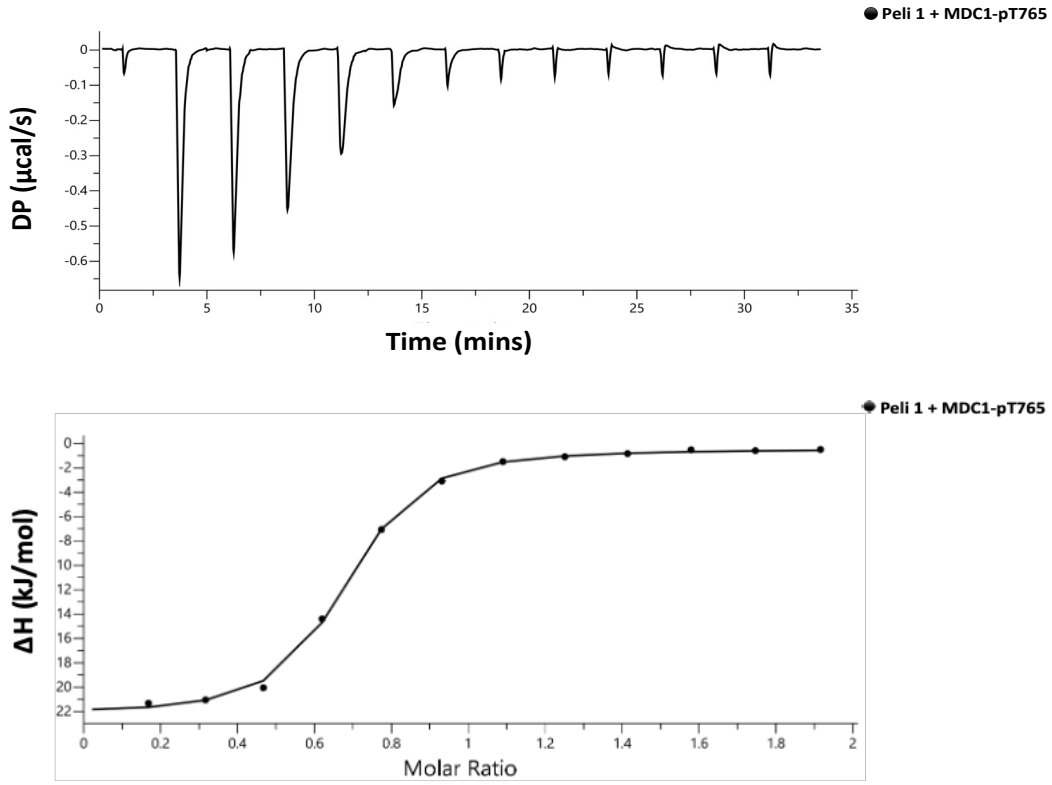


Figure 3.13 - ITC binding experiments demonstrate that Pellino 1/2$_{FHA}$ domains directly bind to MDC1-pTQxF phosphopeptides. A representative raw ITC thermogram (top panel) demonstrates the direct and enthalpically driven binding between Pellino 1/2$_{FHA}$ domains and MDC1-pTQxF phosphopeptides. The differential power (DP) applied (µcal/s) to maintain thermal equilibrium between the sample cell containing 16µM Pellino 1$_{FHA}$ (black symbol) and reference cell following each MDC1$_{pT765}$ phosphopeptide injection is indicated on the y-axis. The x-axis indicates the time (in mins) of the experiment. A representative integrated heat curve/binding isotherm (bottom panel) demonstrates that Pellino 1/2$_{FHA}$ domains bind directly to MDC1-pTQxF phosphopetides and no detectable binding with a γH2AX phosphopeptide. The change in enthalpy (ΔH; kJ/mol) derived from integration of the raw ITC thermograms from Pellino 1$_{FHA}$:MDC1$_{pT765}$ (black symbols) versus Pellino 1$_{FHA}$:γH2AX (red symbols) binding experiments is indicated on the y-axis. The molar ratio of each phosphopeptide:Pellino 1$_{FHA}$ protein at each injection is indicated on the x-axis.

Complex formation may be attributed with differences in enthalpic and entropic contributions between Peli family members, wherein Peli 2 pays the larger entropic cost, but makes more favourable interactions (Peli 1$_{FHA}$ average ΔH:-84.5±8.3kJmol^{-1} and average TΔS: -49.9±6.9kJmol^{-1}, Peli 2$_{FHA}$ average ΔH: -120.9±27.4 kJmol^{-1} and average TΔS: -86.0±26.9kJmol^{-1}; Table 3.8.2). The

latter observation may reflect a larger degree of apo-Peli 2_{FHA} conformational flexibility compared to apo-Peli 1_{FHA}, despite the high level of sequence identity, but this requires further biophysical experiments. In any case, the K_D across all MDC1-pTQxF phosphopeptides was similar between both Peli family members (Table 3.2). The $MDC1_{pT765}$ phosphopeptide bound ~3-5-fold tighter to Peli $1/2_{FHA}$ domains than the other tested MDC1-TQxF phosphopeptides. Due to the unnatural +4A in $MDC1_{pT699}$ and $MDC1_{pT719}$, this quantification-based comparison requires further experimentation to detail any contribution of the +4. Importantly, we did not observe any detectable binding between the synthetic γH2AX phosphopeptide and Peli $1/2_{FHA}$ proteins (Figure 3.14 and Table 3.2).

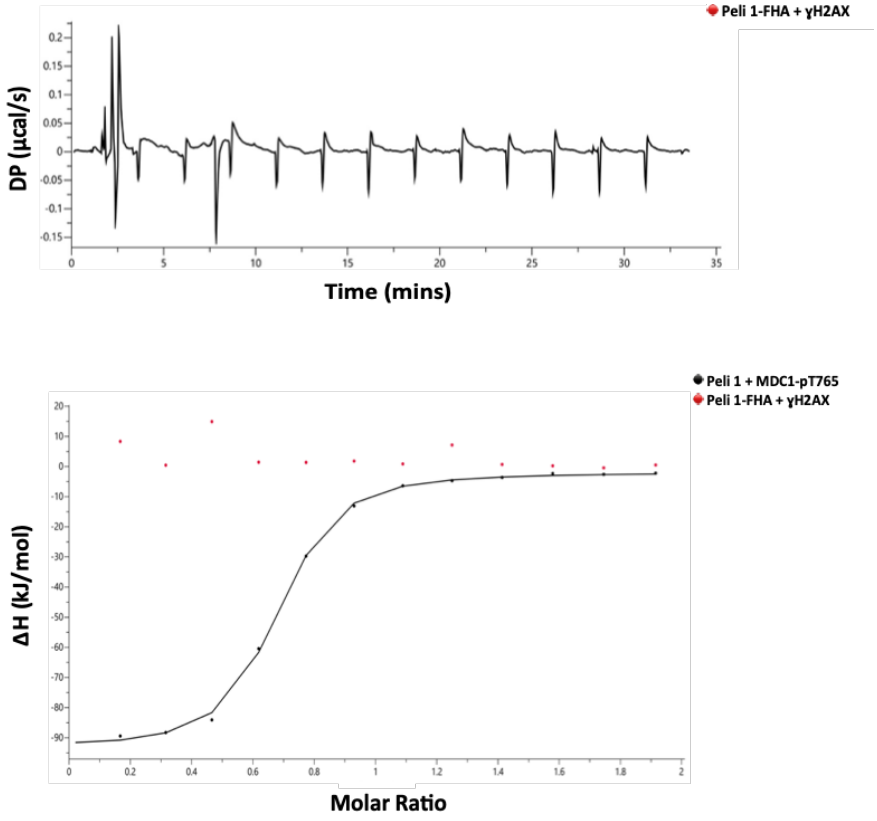


Figure 3.14 - ITC binding experiments demonstrate that Pellino $1/2_{FHA}$ domains do not directly bind to a γH2AX phosphopeptide. A representative raw ITC thermogram (top panel) exhibits no detectable binding between Pellino $1/2_{FHA}$ domains and a γH2AX phosphopeptide (red). A representative integrated heat curve/binding isotherm (bottom panel) demonstrates that Pellino $1/2_{FHA}$ domains bind directly to MDC1-pTQxF phosphopetides (black) and not with a γH2AX phosphopeptide (red).

In contrast, the $MDC1_{tBRCT}$ domain bound directly to γH2AX with a K_D of 853±128nM, which is approximately ~2.5-fold tighter than previously observed (*Stucki et al., 2005;* Figures 3.15, and Table 3.2). Our data, hereby, supports an MDC1-pTQxF-dependent mechanism of Peli 1/2 protein recruitment to DSBs that indirectly involves γH2AX through recruitment of MDC1 by its C-terminal tandem BRCT repeats.

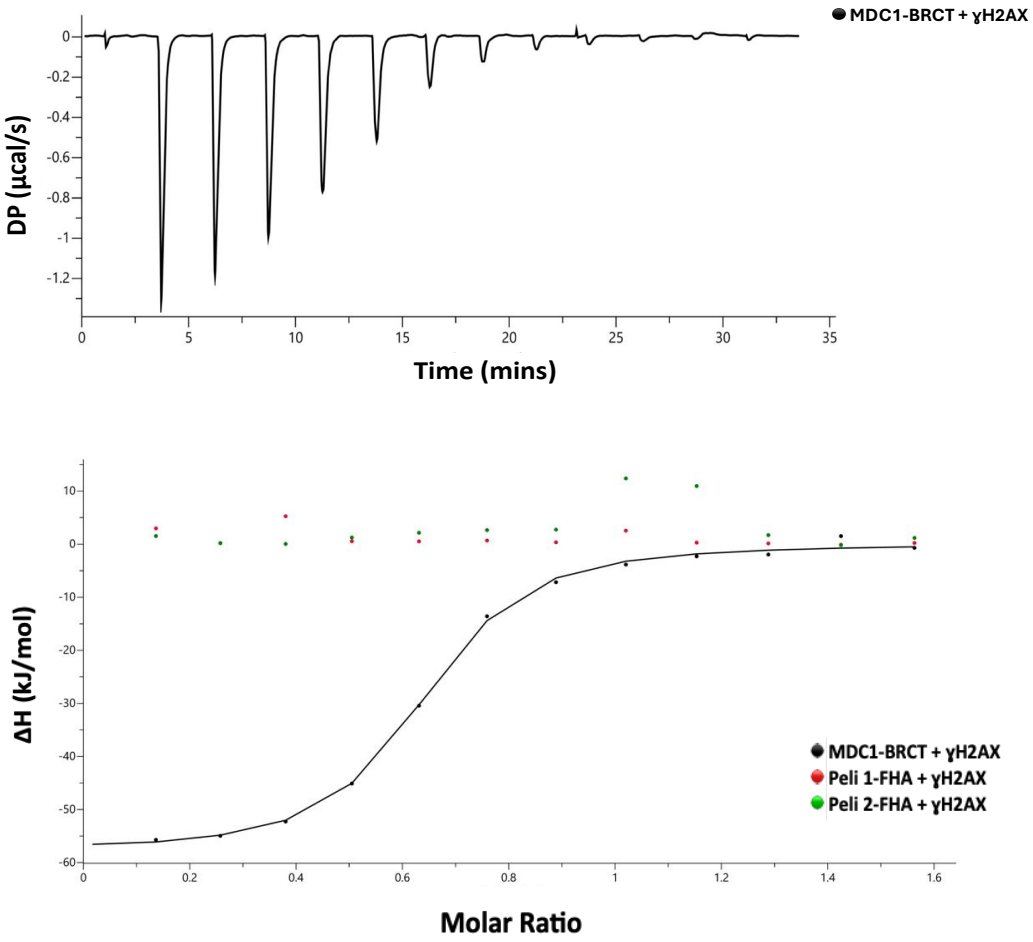


Figure 3.15 - ITC binding experiments demonstrate that $MDC1_{tBRCT}$ domains directly bind to the γH2AX phosphopeptide, but Pellino 1/2 domains do not. A representative raw ITC thermogram (top panel) exhibits direct and enthalpically-driven detectable binding between the $MDC1_{tBRCT}$ domain and a γH2AX phosphopeptide (black). A representative integrated heat curve/binding isotherm (bottom panel) demonstrates that $MDC1_{tBRCT}$ directly bind to the γH2AX phosphopeptide (black), while Pellino 1_{FHA} (red) and Pellino 2_{FHA} (green) domains do not.

Pellino 1$_{FHA}$					
Peptide	**K_D (μM)**	**N**	**ΔH (KJmol^{-1})**	**ΔG (KJmol^{-1})**	**TΔS (KJmol^{-1})**
MDC1-pT699 (Ac-AYDLQA**pT**QCFA-Am	2.01	0.36	-77.5	-32.5	-44.9
MDC1-pT719 (Ac-AYEDEP**pT**QAFA-Am	1.14	0.78	-91.2	-34.0	-57.3
MDC1-pT752 (Ac-AYEVLA**pT**QPFC-Am	1.14	0.74	-77.2	-33.9	-43.3
MDC1-pT765 (Ac-AYEDSE**pT**QPFD-Am	0.23	0.62	-92.1	-38.0	-54.1
IRAK1-pT141 (Ac-AYSSAS**pT**FLYP-Am	0.42	0.70	-72.6	-36.5	-36.1
H2AX-pS139 (Ac-GGKATQA**pS**QEY)	N.B.D.	/	/	/	/
Pellino 2$_{FHA}$					
Peptide	**K_D (μM)**	**N**	**ΔH (KJmol^{-1})**	**ΔG (KJmol^{-1})**	**TΔS (KJmol^{-1})**
MDC1-pT699 (Ac-AYDLQA**pT**QCFA-Am	1.12	0.55	-83.4	-34.0	-49.4
MDC1-pT719 (Ac-AYEDEP**pT**QAFA-Am	1.22	0.52	-145	-33.8	-111
MDC1-pT752 (Ac-AYEVLA**pT**QPFC-Am	1.06	0.69	-118	-34.1	-83.7
MDC1-pT765 (Ac-AYEDSE**pT**QPFD-Am	0.37	0.69	-137	-36.7	-100
IRAK1-pT141 (Ac-AYSSAS**pT**FLYP-Am	0.70	0.59	-108	-35.2	-73.0
H2AX-pS139 (Ac-GGKATQA**pS**QEY)	N.B.D.	/	/	/	/
MDC1t$_{BRCT}$					
Peptide	**K_D (μM)**	**N**	**ΔH (KJmol^{-1})**	**ΔG (KJmol^{-1})**	**TΔS (KJmol^{-1})**
H2AX-pS139 (Ac-GGKATQA**pS**QEY)	0.85	0.60	-58.2	-34.6	-23.5

Table 3.2 - Summary of ITC binding experiments conducted. The primary amino acid sequence of each individual phosphopeptide is provided, whereby Ac and Am denotes N-terminal acetylation and C-terminal amidation, respectively. Detectable binding between each purified phospho-binding module and phosphopeptide is indicated with a green phosphosite (pT or pS), while no detectable binding (N.D.B) is indicated in red. The affinity/dissociation constant (K_D; in μM), molar stoichiometry of binding (N), change in enthalpy (ΔH; in kJmol^{-1}), change in Gibbs free energy upon binding (ΔG; in kJmol^{-1}), and change in entropy at a constant temperature (TΔS; in kJmol^{-1}) are provided. All experiments were conducted at least twice. All thermodynamic parameters were calculated using the MicroCal PEA-ITC Analysis software v1.41 (Malvern Panalytical). **Data were produced by Dr. Éilís Bragginton and Matthew Stewart.**

3.4.3 Crystal Structure of the Pellino 1-FHA:MDC1-pT765 Phosphopeptide Complex

We pursued solving a high-resolution crystal structure of a Peli 1/2:MDC1-pTQxF complex, because there are no known crystal structures of any Peli family member FHA domains bound to a phospholigand. At the outset, the Peli 1_{FHA} and the $MDC1_{pT765}$ candidate complex was pursued, because the affinity of binding was the tightest from all ITC binding experiments conducted in this study (Table 3.2), and there was the possibility that the Peli 2_{FHA} was truly more conformationally flexible than Peli 2_{FHA}, despite previous apo-Peli 2_{FHA} crystallisation success (*Li et al., 2008*). Reducing agent was omitted from initial co-crystallisation sitting-drop vapour diffusion screens. Nonetheless, microcrystals nucleated in ProPlex™ HT-96 well G1 containing 0.1M Tris pH 8 with 1.5M $(NH_4)_2SO_4$ 6 days after initial screens were performed, and grew to maximum size after 12 days (Figure 3.16, left panel). The addition of 500μM TCEP to co-crystallisation buffer was sufficient to inhibit crystal nucleation, whereas addition of either 5% (w/v) *myo*-inositol (Figure 3.16, middle panel) or 20% (w/v) PEG400 (Figure 3.16, right panel) to the reservoir solution and mother liquor were sufficient to obtain single orthorhombic crystals of sufficient size (~0.4mm) for x-ray diffraction data collection at Diamond Light Source, Oxford beamline I04 at a wavelength of 0.9795Å.

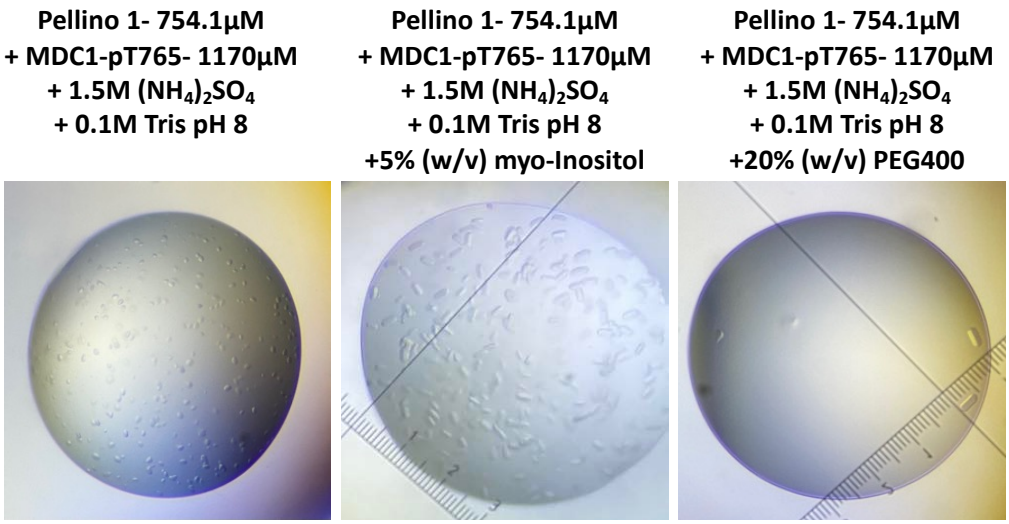


Figure 3.16 - Crystallisation trials of the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex. Initial ProPlex™ HT-96 well G1 hit from initial crystallisation screens after incubation at 18°C for 12 days (left panel). Supplementation of the initial screen hit with 5% (w/v) *myo*-inositol aids growth of Pellino 1_{FHA}:$MDC1_{pT765}$ crystals (middle panel). Final crystal trials involve supplementation of the initial screen hit with 20% (w/v) PEG400 to aid growth of individual Pellino 1_{FHA}:$MDC1_{pT765}$ crystals (right panel). Crystals were visualised with a trinocular stereo microscope at 15x magnification.

The X-ray crystal structure of the N-terminal Peli 1 fragment encompassing residues 13-255 bound to the synthetic $MDC1_{pT765}$:AYEDSEpTQPFD phosphopeptide was solved using molecular replacement with the atomic coordinates from the previously reported apo-Peli 2_{FHA} crystal structure (PDB accession code: 3EGA, Figure 3.3) as a search model (*Li et al., 2008*). Initial maps phased with the initial molecular replacement model showed clear residual density for the $MDC1_{pT765}$ phosphopeptide at a 1σ contour level, which enabled a near complete model of the complex (residues 13-30, 35-119, and 130-255) to be manually built and refined at 2.55Å (Figure 3.17, Table 3.3) (*Emsley et al., 2010*). The Peli 1_{FHA} is composed of 16 β-strands, which in turn, form 4 distinct β-sheets, and an α-helix that is not present in the Peli 2_{FHA} structure (*Li et al., 2008*). This latter finding may reflect the differences in average TΔS values between Peli family members from the above ITC experiments (Table 3.2). Like the apo-Peli 2_{FHA} crystal structure, the Peli 1_{FHA} adopts an atypical FHA-fold comprised of a central FHA core (residues 13-48, 97-156, and 183-255) that

can be subdivided into two β-sheets (β_1-β_2, β_6-β_8, and β_{12}-β_{16}) that form a β-sandwich containing a large loop insertion between β_{14}-β_{15} (Figure 3.17C). The FHA core is split by a bi-partite 'wing' appendage (wing split I: residues 49-96, wing split II: 157-182) that consists of the remaining two β-sheets (β_3-β_5, β_9, and β_{10}-β_{11}), and an α–helix that separates β_3-β_4 (Figures 3.8.4.2B and 3.8.4.2C). Interestingly, residual electron density supporting a disulphide bond between C61 and a symmetry-related C61 residue was discovered, thereby demonstrating that the lack of reducing agent within the reservoir and mother liquor solution was crucial for crystal lattice formation (Figure 3.17D).

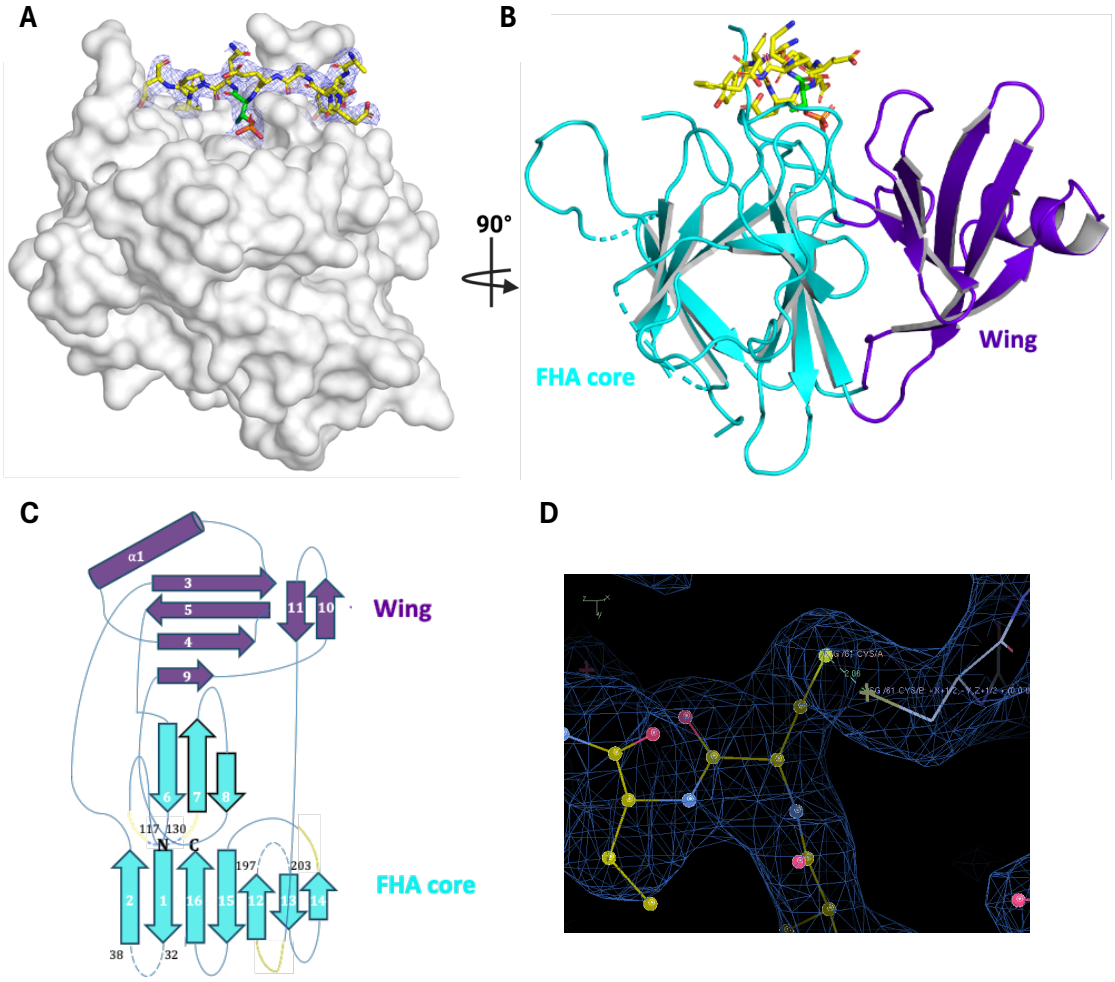


Figure 3.17 - The 2.55Å crystal structure of the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex. (A) Surface representation of the Pellino 1_{FHA} domain bound to the $MDC1_{pT765}$ phosphopeptide. The solvent-accessible surface of the atypical Pellino 1_{FHA} domain is represented in white. The bound $MDC1_{pT765}$ phosphopeptide backbone carbon atoms are coloured in yellow, except for the pThr, which is coloured in green. Individual elemental atoms coloured using the byelement colour scheme. The residual electron density at a 1σ contour level supporting the phosphopeptide is coloured in blue**. (B) Secondary strucuture representation of the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex.** The structure in A is rotated about the y-axis clockwise by 90°. **(C) 2D Secondary structure topology diagram of the Pellino 1FHA domain.** Individual β-strands are represented as numbered arrows whose directionality roughly corresponds to 3D atomic space. The individual α-helix is represented as a bar and connecting loops are represented as blue lines, while yellow colouring is used to indicate regions involved in binding to the MDC1pT765 phosphopeptide. Dashed lines are used to represent areas that are unsupported by density. The N- and C-termini are also represented above β_1 and β_{16} strands, respectively. **(D) Residual electron density supports a disulphide bond between Pellino 1-C61 and symmetry-related Pellino 1-C61, responsible for crystal lattice formation.** Pellino 1_{FHA} atoms are represented in the stick format with the carbon backbone coloured in yellow, while the symmetry related Pellino 1-C61 residue carbon backbone is coloured in light blue. This figure was made using Coot v1.1.10 from the CCP4 Program Suite v9.0.004 (*Emsley et al., 2010*; *Agirre et al., 2023*).

Pellino 1 (13-255):MDC1-pT765 (STARANISO)			
Data Collection Diamond I04			
Space group	$P2_12_12_1$		
Cell dimensions	a, b, c 52.9 Å 75.1 Å 165.1 Å α, β, γ 90° 90° 90°		
Wavelength	0.9795 Å		
Resolution limit	2.55 Å		
	Overall (68.35-2.55 Å)	**Inner Shell (68.35-8.01 Å)**	**Outer Shell (2.77-2.55 Å)**
R_{merge}	0.215	0.060	1.682
R_{meas}	0.224	0.063	1.803
R_{pim}	0.062	0.019	0.626
Total n° of observations	204,610	9,148	6,591
Total n° of unique observations	15,881	794	794
<I/σ>	8.9	21.1	1.3
Completeness (ellipsoidal)	91.9 %	99.9	63.1
Redundancy	12.9	11.5	8.3
Refinement			
Resolution limits	68.35-2.55 Å		
N° of reflections	15,866 (155)		
R-factor (R_{work})	0.21 (0.33)		
R-factor (R_{free})	0.21 (0.37)		
Ramachandran favoured	96.26 %		
Ramachandran outliers	0.66 %		
Clashscore	7.42		
Average B-factor	33.43 $Å^2$		
Model			
Protein	2 X Pellino 1 (13-255):MDC1-pT765		
	13-30, 35-119, 130-255 aa		
	5 sulphate ions		
	70 water molecules		
Total n° of atoms	3,817		
R.M.S deviations			
Bond lengths	0.009 Å		
Bond angles	1.04°		

Table 3.3 - Summary of crystallographic statistics for the Pellino 1_{FHA}:$MDC1_{pT765}$ complex structure. Refinement and model statistics were calculated in PHENIX using the comprehensive validation program, which includes MolProbity tools (*Echols et al., 2014; Williams et al., 2018*).

The MDC1$_{pT765}$ phosphopeptide binds in an extended conformation to a basic pocket within the FHA core, formed by loops separating β_6-β_7, β_{11}-β_{12}, and β_{14}-β_{15} (Figures 3.4.3.3A-C).

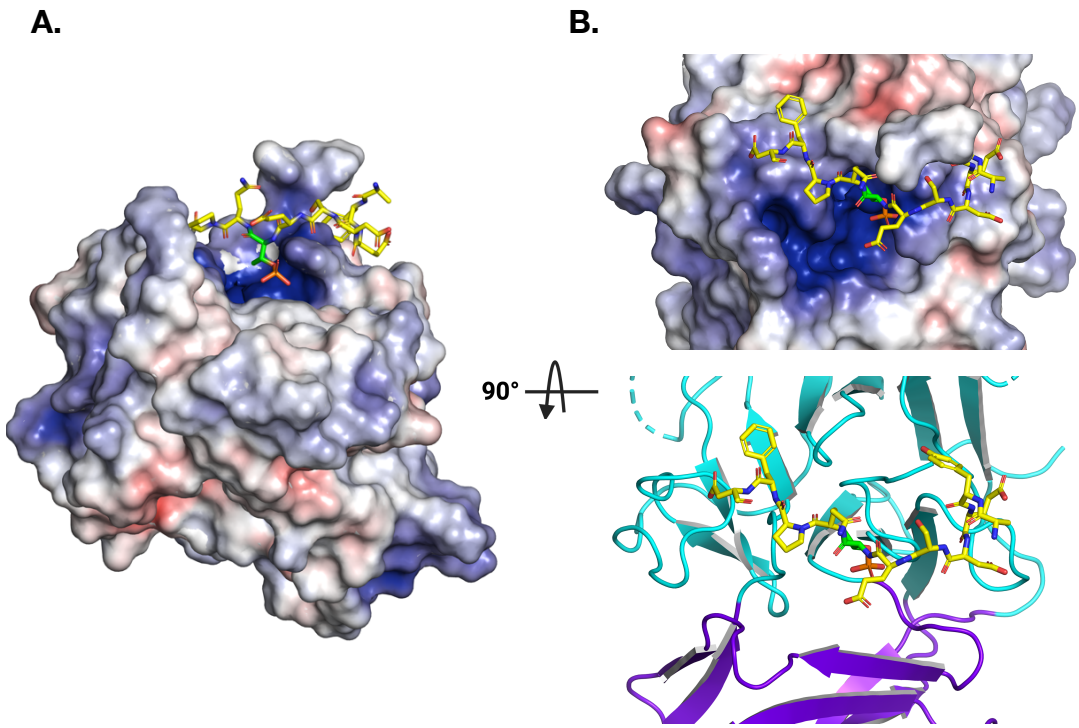


Figure 3.18 - The MDC1p$_{T765}$ phosphopeptide binds to the Pellino 1 FHA core basic pocket. In A, the surface of the Pellino 1$_{FHA}$:MDC1$_{pT765}$ phosphopeptide complex in Fig 3.17A is coloured by electrostatic potential (electropositive, electronegative, and electroneutral is indicated in blue, red, and white, respectively), as calculated by the Adaptive Poisson-Boltzmann Solver Plugin using the PyMOL Molecular Graphics System v3.1.1 (*Schrödinger, LLC; Jurrus et al., 2018*). In B, the structure is rotated 90° clockwise about the x-axis (B, top panel). The secondary structure of the Pellino 1$_{FHA}$ domain in this orientation is LAO represented in the cartoon format (B, bottom panel).

The pT765 γ-phosphate moiety forms four hydrogen-bonding interactions with side and main chain atoms from R104, S135, and R136 located in the loop separating β_6-β_7 of the FHA core β-sandwich (Figure 3.19). As previously mentioned, alanine mutation of the R104 equivalent residue in Peli 2 (R106) to a phosphonull A residue ablates interaction with IRAK1$_{pT141}$ phosphopeptides (*Li et al., 2008*).

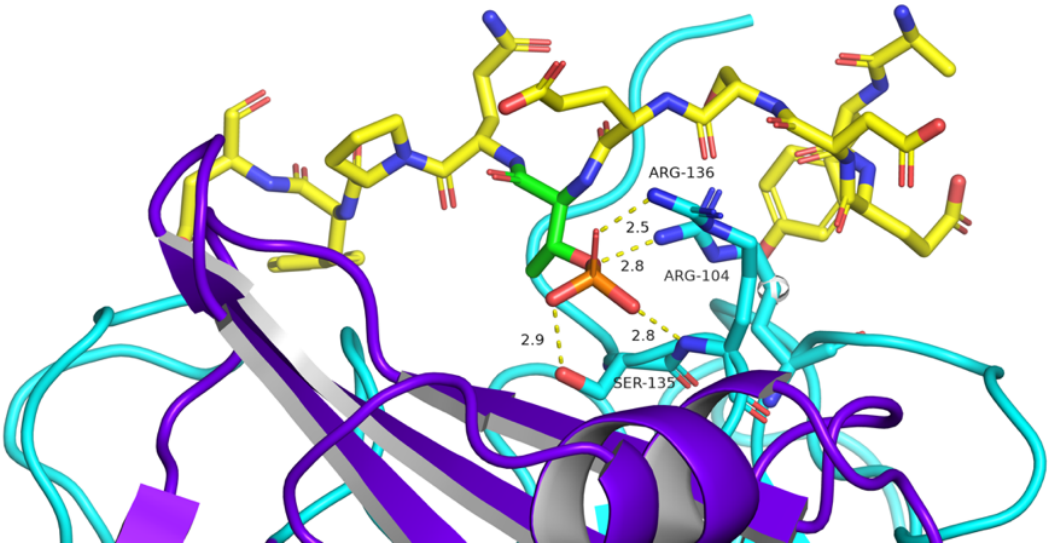


Figure 3.19- Pellino 1_{FHA} residues R104, S135, and R136 directly interact with the pThr residue in the $MDC1_{pT765}$ phosphopeptide. Pellino 1_{FHA} domain is represented in the cartoon format and coloured accordingly as in B (bottom panel). Atoms for pThr-contacting Pellino 1_{FHA} residues R104, S135, and R136 are represented as sticks. Each hydrogen-bonding contact between Pellino 1_{FHA} and the pThr residue are represented with dashed yellow lines and the distance between atoms (Å) are provided.

In particular, the phosphate-contacts involving R104, S135, and R136 residues are highly conserved throughout the DDR FHA family and evolutionary homologues, including MDC1 (PDB accession code: 3UOT; R58, S72, and K73), RNF8 (PDB accession code: 2PIE; R42, S60, and R61), Chk2 (PDB accession code: 1GXC; R117, S140, and K141), and the *Mycobacterium Tuberculosis* FHA-containing protein Rv0020c (PDB accession code: 3POA; R29, S43, and R44) (Figure 3.20) (*Li et al., 2002; Huen et al., 2007; Pennell et al., 2010; Roumestand et al., 2011; Jungmichel et al., 2012*).

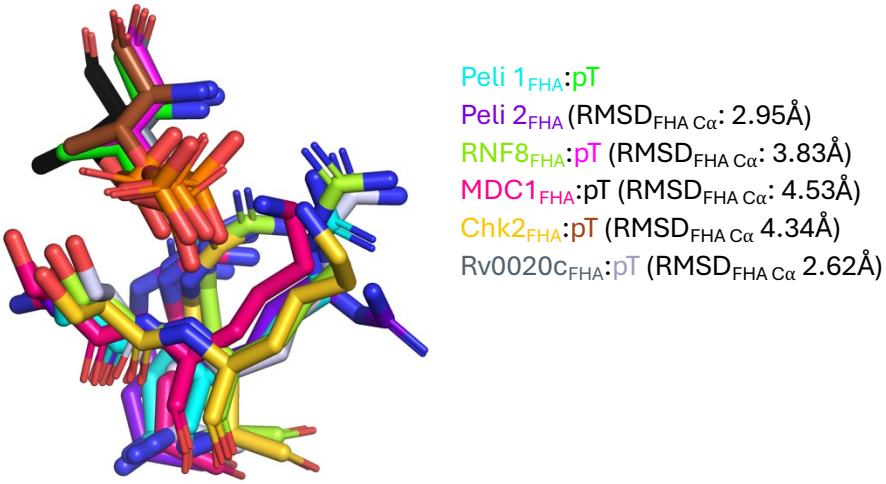


Figure 3.20 Pellino 1_{FHA} pThr-contacting residues are structurally conserved throughout the Pellino and FHA families. Atomic coordinates from crystal structures of the human apo-Pellino 2_{FHA} core (purple; accession code: 3EGA, *Lin et al., 2008*), human RNF8$_{FHA}$ (yellow green) bound to its optimal phosphopeptide (magenta; accession code: 2PIE), human MDC1$_{FHA}$ (hot pink) bound to a MDC1-pT4 phosphopeptide (black; accession code: 3UOT), human Chk2$_{FHA}$ (gold) bound to its optimal phosphopeptide (brown; accession code: 1GXC), and *M. tuberculosis* Rv0020c$_{FHA}$ (slate grey) bound to its optimal phosphopeptide (blue white; accession code: 3POA) were structurally aligned to the Pellino 1_{FHA} core independently of amino acid sequence using the cealign command in the PyMOL Molecular Graphics System v3.1.1. Only atoms for residues involved in contacting pThr residues are represented using the stick format. The root-mean square deviation (RMSD) for all alpha carbon atoms (Cα) is provided.

There are an additional 13 hydrogen bonds between the Peli 1_{FHA} and MDC1_{pT765} phosphopeptide that contribute to the ~750Å^2 interaction interface surface area, of which 4 are derived from the synthetically introduced -5Y AY motif residue with R104, D116, and S132, respectively. The remaining 9 hydrogen bonds are mediated by loops β_6-β_7, β_{11}-β_{12}, and β_{14}-β_{15}, involving R104, S132, R136, N186, and S223 with the following MDC1_{pT765} phosphopeptide positions -2S, +1Q, +3F, and +4D (Figure 3.21). To our knowledge, this is the first instance of a Peli family member interacting with a phosphopeptide containing a +3F residue.

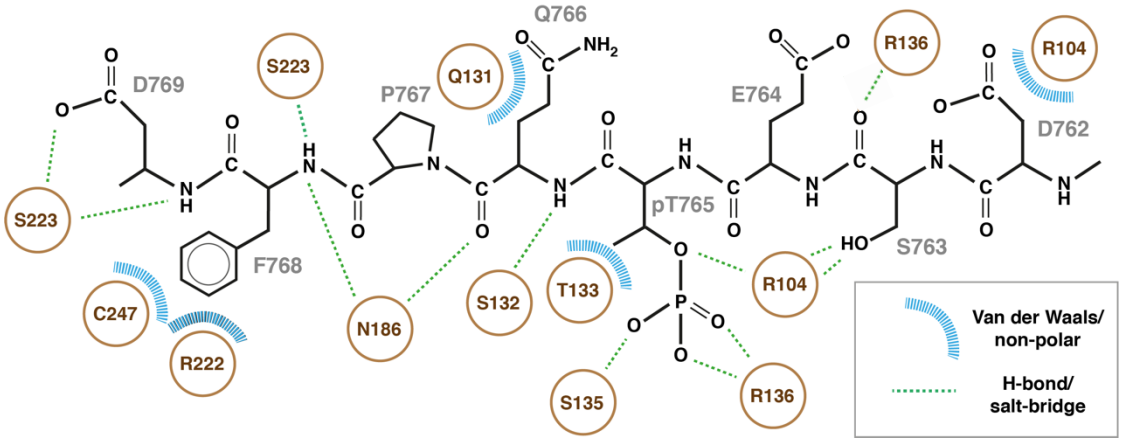


Figure 3.21 - A schematic representation of all interactions between Pellino 1_{FHA} and the $MDC1_{pT765}$ phosphopeptide. Hydrogen bonding and salt bridge contacts are represented with green dashed lines, while van der Waals and non-polar contacts are represented with blue dashed lines.

Previous structural analyses of the apo-Peli 2_{FHA} have proposed that the bi-partite 'wing' appendage may contribute to substrate binding through extension of the phospholigand interaction interface (*Li et al., 2008*). In turn, it was proposed that the 'wing' could allow for binding of multiply phosphorylated substrates, or through interactions with surface features of substrates to supplement FHA core-specific phospholigand interactions, or even to confer specificity of Peli family interactions, including regulatory interactions with the C-terminal RING domain (*Li et al., 2008*). Within the context of binding a singly phosphorylated MDC1-pTQxF peptide, the 'wing' appendage contributes indirectly. Instead, the 'wing split II' structurally supports the large loop insertion between β_{14}-β_{15}, subsequently forming a shallow basic platform formed predominantly by residues R222 and S223 (Figure 3.22, top panel). In particular, S223 forms a side chain hydrogen bond with both the +3F α-amino group and +4D acidic side chain and the side chain of β_{11}-β_{12} loop N186 forms an additional hydrogen bond with the +3F α-amino group, and the remaining residues pack against +3F +4D by cation-π and Van der Waals interactions (Figure 3.22, bottom panel).

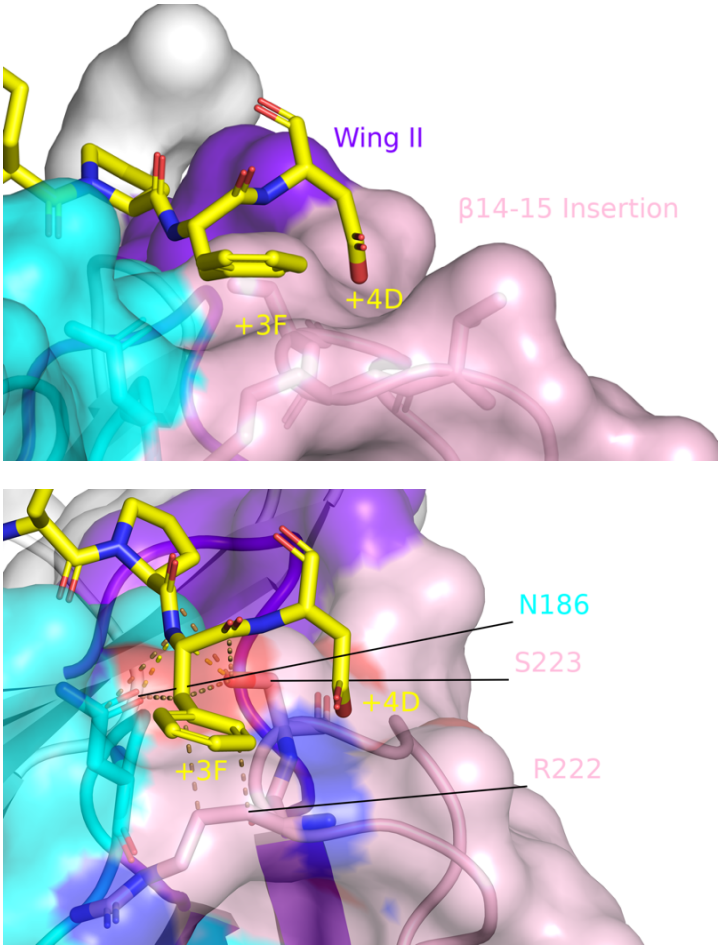


Figure 3.22 - The Pellino family 'wing' insert contributes indirectly to phosphopeptide pThr +3 selectivity. The 'wing II' insert structurally supports the $Peli_{FHA}$ core β_{14}-β_{15} loop insertion (top panel), which contains residues R222 and S223 that, in conjunction with the β_{11}-β_{12} loop residue N186, directly interact with $MDC1_{pT765}$ pThr +3F and +3D residues (bottom panel). The solvent-accessible surface of the Pellino 1_{FHA} domain is represented with 20% transparency to visualise the underlying secondary structure represented in the cartoon format. All the FHA core is coloured in cyan, except for the β_{14}-β_{15} insertion, which is represented in pink. The 'wing I' insertion is coloured in white, while the 'wing II' structural buttress is coloured in purple. Brown and yellow dashed lines indicate van der Waals/non-polar and hydrogen-bonding contacts, respectively, between Pellino 1 N186, R222, and S223 residue atoms with $MDC1_{pT765}$ pThr +3F and +4D residue atoms.

Interestingly, the structural mechanism for +3 selectivity differs between Peli 1_{FHA} and $RNF8_{FHA}$. For example, the +3F benzyl side chain group is rotated and anchored away from the Peli pThr-binding pocket by N186 and S223, whereas the optimised RNF8 FHA domain-binding phosphopeptide +3Y

residue binds to an equivalent N83 residue within an insertion loop adjacent to the RNF8 pThr-binding pocket (Figure 3.23) (*Huen et al., 2007*).

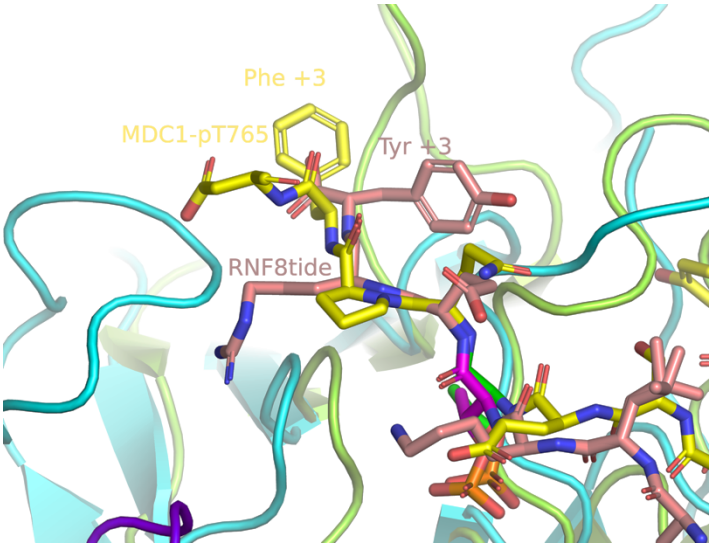


Figure 3.23 - Structural alignments of Pellino 1_{FHA}:$MDC1_{pT765}$ and $RNF8_{FHA}$:RNF8tide complexes reveal differences in pThr +3 binding selectivity.

In summary, the Peli 1_{FHA}:$MDC1_{pT765}$ complex crystal structure has demonstrated that the interface only directly involves residues that are highly conserved throughout the FHA family. In particular, the phosphate-contacting R104, S135, and R136 residues, the hydrogen bond between N186 and +1Q main chain C=O, and 2 hydrogen bonds between N186 and S223 with the main chain +3F are consistent with almost all FHA domains. Like other FHA domains, the γ-methyl group of the pThr residue is accommodated by weak favourable van der Waals interactions with residues T133, I134, S135, and N186 that line a Peli 1_{FHA} surface pocket, which stabilises the Peli 1_{FHA}:$MDC1_{pT765}$ hydrogen-bonding interaction network (Figure 3.24). Consequently, these properties enable FHA family members, including Peli 1, to specifically select for pThr over pSer. We have also shown that the 'wing split II' contributes indirectly to phosphopeptide binding through providing structural support to the large loop insertion between β_{14}-β_{15}, where +3F and +4D contacting-residues T221, R222, and S223 reside.

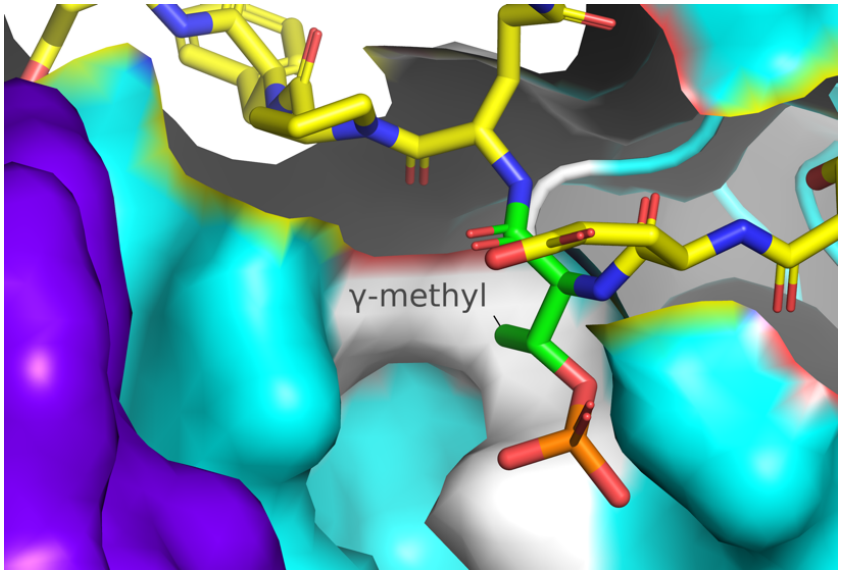


Figure 3.24 - The Pellino 1_{FHA} γ-methyl pocket involved in pThr selectivity. The solvent-accessible surface of the Pellino 1_{FHA} domain and $MDC1_{pT765}$ phosphopeptide is depicted γ-methyl pocket residues I133, T134, S135, and N186 coloured in white.

3.4.4 AlphaFold Models Accurately Predict the Structure of the Pellino 1-FHA:MDC1-pT765 Complex

Throughout this study, one of the prevailing questions was whether AlphaFold (AF) artificial intelligence (AI) systems can accurately predict phosphorylation-dependent protein:ligand structures (*Senior et al., 2020; Bagdonas et al., 2021; Jumper et al., 2021; Tunyasuvunakool et al., 2021*). In part, this was because early AF v2.0.0 and AF-Multimer v2.1.0 (AF-M) ColabFold prediction webservers available during the beginning of this study were not able to query post-translationally modified amino acid input sequences (*Jumper et al., 2021; Mirdita et al., 2022*). AF3 was released in May 2024, and possesses the capability to predict biomolecular complexes that involve protein post-translational modifications, both modified and unmodified DNA and RNA, as well as common small molecule ligands, such as ions, redox agents, and fatty acids (*Abramson et al., 2024*).

As of August 2024, our Peli1$_{FHA}$:MDC1$_{pT765}$ crystal structure had not yet been deposited to the Protein Data Bank (PDB) structural database, and so we were able to predict Pellino 1$_{FHA}$:MDC1$_{pT765}$ phosphopeptide complex using AlphaFold 2 and 3 models without the use of a structural template. In contrast, it was likely that the AlphaFold models were able to predict the Peli 1$_{FHA}$: MDC1$_{pT765}$ phosphopeptide complex structure by leveraging the availability of pre-existing structural templates, since there are structures for at least 18 members of the human FHA family, including the apo-Peli 2$_{FHA}$ structure, which all contain the highly conserved G-R-[x_n]-S-[R/K] pThr-binding motif, and all exhibit a high level of structural conservation with the Peli 1$_{FHA}$ domain (*Lin et al., 2008*). Moreover, there are at least 32 members of the human FHA family alone, which were identified using the atomic coordinates for Pellino 1$_{FHA}$, RNF8$_{FHA}$, MDC1$_{FHA}$, Chk2$_{FHA}$ crystal structures as search models in the Foldseek protein search computational tool (*van Kempen et al., 2024;* Table 3.4). These 32 members were cross-validated using the InterPro FHA domain database, sequence analysis based on the highly conserved GR-SR/K pThr-binding motif (*Paysan-Lafosse et al., 2023*). Consequently, the high level of structural and sequence conservation across all human FHA family members was likely to be able to provide AlphaFold with the potential to leverage multiple sequence alignments to identify those FHA proteins with pre-existing structures. In turn, AlphaFold was likely to be able to accurately model the co-evolutionary relationship of paired residues between FHA family members and their phospholigands and use this information to predict accurate structural models. Taking these considerations into account, we were presented with the unique opportunity to test the accuracy of AF-M and AF3 during prediction of the Peli 1$_{FHA}$:MDC1$_{pT765}$ complex.

Human FHA Domain	Structures	Human FHA Domain	Structures
AFDN	1WLN (*Hayashi et al., 2004*)	Peli1	This study
AGGF1		Peli2	3EGA (*Lin et al., 2008*)
APLF	5E50 (*Cherry et al., 2015*)	Peli3 (α and β)	
CEP170		PHF12	
CEP170B	4JON (*JCSG, 2013*)	PHLDB1	
CHFR	1LGQ (*Stavridi et al., 2002*)	PNKP	2W3O (*Ali et al., 2009*)
Chk2	1GXC (*Li et al., 2002*)	RADIL	
FHAD1		RASIP1	
FOXK1		RNF8	2PIE (*Huen et al., 2007*)
FOXK2		SLC4A1AP	4H87 (*JCSG, 2012*)
Ki-67	1R21 (*Li et al., 2004*)	SLMAP	6AR0 (*Bae et al., 2017*)
MCRS1		SNIP1	7ABH (*Townsend et al., 2020*)
MDC1	3UOT (*Jungmichel et al., 2012*)	STARD9	
Nbs1	3I0M (*Lloyd et al., 2009*)	TCF19	
NIPP-1	2JPE (*Kumeta et al., 2008*)	TIFA	4YM4 (*Weng et al., 2015*)
		TIFAB	8WWY (*Nakamura et al., 2024*)

Table 3.4 - The human FHA domain family. All family members were identified using the atomic coordinates for Pellino 1_{FHA}, $MDC1_{FHA}$, $RNF8_{FHA}$ and $Chk2_{FHA}$ structures to search for homologous proteins by the Foldseek protein search computational tool, verified by InterPro FHA domain database, and sequence analysis based on the highly conserved G-R-[x_n]-S-[R/K] pThr-binding motif (*Paysan-Lafosse et al., 2023*; *van Kempen et al., 2024*). All human members with structures are indicated with the relevant PDB accession code and reference.

Initially, we predicted the Peli 1_{FHA}:$MDC1_{pT765}$ complex using AF-M v2.3.2 by inputting the amino acid sequence of the regions corresponding to the solved Peli 1_{FHA} domain (residues 13-263) and a 9-mer $MDC1_{pT765}$ peptide, containing only the natural MDC1 amino acids both 4 positions upstream and downstream of the pT765 phosphosite (Appendix VI). Importantly, the T765 residue was inputted in its non-phosphorylated form, because this was not provided by earlier AlphaFold models, but a γ-phosphate moiety was manually added to output maps using the COOT software. This process was

repeated using the AF3 webserver, but instead the MDC1 9-mer input peptide was phosphorylated at T765 prior to prediction. Both AF-M and AF3 generated 5 output models, and individual candidates were selected based on AF ranking confidence metrics (Figure 3.25).

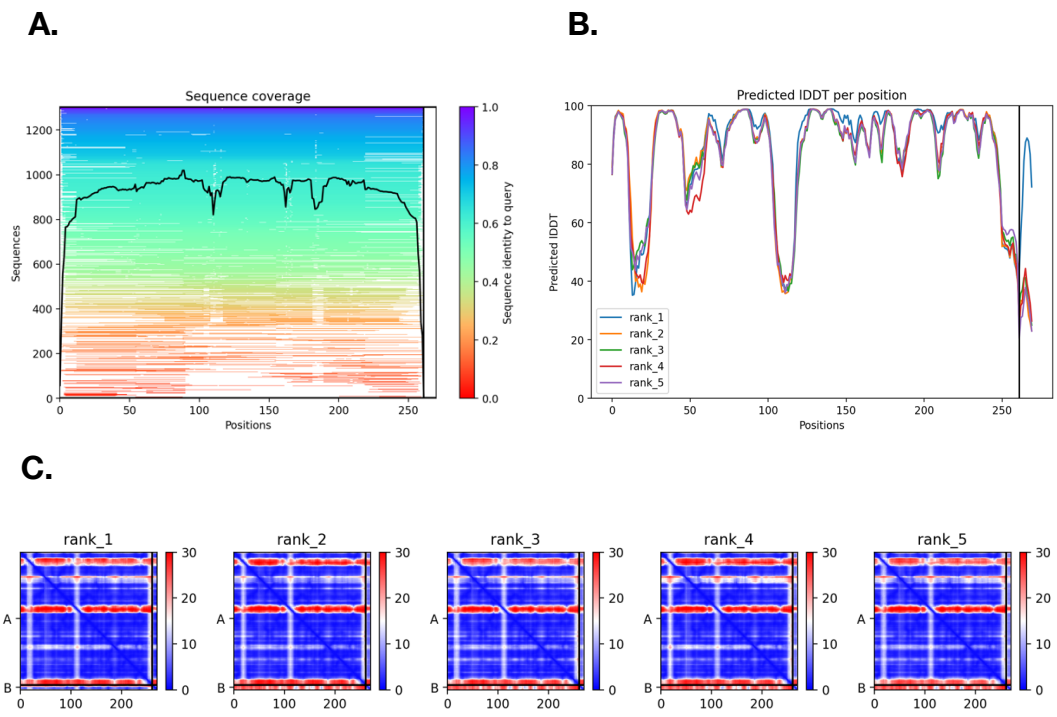


Figure 3.25 - AlphaFold 2 Multimer (AF-M) metric outputs for the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex structural models. (A) AF-M MSA identifies ~700-1000 homologous primary amino acid sequences that share ~60-70% sequence identity to the Pellino 1_{FHA} query and 0 homologous $MDC1_{T765}$ 9-mer peptide sequences. The left-hand side x-axis indicates each amino acid position within the Pellino 1_{FHA} domain query sequence, while the right-hand segment indicates the sequence within the 9-mer $MDC1_{T765}$ peptide. The y-axis indicates the number of homologous sequences identified. The graph is coloured according to the sequence identity (0-100%). The number of homologous sequences identified, and the level of sequence identity per amino acid residue is depicted by the solid black line. **(B) The AF-M IDDT metric identifies a single model with a high level of $MDC1_{T765}$ local residue confidence.** pLDDT is a per-residue confidence measurement of the expected distance deviation of a predicted model versus a true experimental structure (0-100) Each AF-M output model is indicated by a coloured line**. (C) The AF-M PAE metric identifies a single model with a high confidence level of oriented interacting residues.** The predicted aligned error (PAE) for a pair of residues is given on a scale of 0-30Å is depicted using a blue-to-red gradient colour scheme. For each individual ranked model, the PAE per residue for intra-Pellino 1_{FHA} paired residues is depicted by the main square, intra-$MDC1_{T765}$ paired peptide residues on the bottom right diagonal square, and inter-Pellino 1_{FHA}:$MDC1_{T765}$ and $MDC1_{pT765}$:Pellino 1_{FHA} paired residues are indicated on the bottom and right off diagonal rectangular plots, respectively.

Confidence metrics included multiple sequence alignments (MSA), predicted local distance difference test (pLDDT), predicted aligned error (PAE), predicted template modelling (pTM), interface predicted template modelling (ipTM). Of note, the multiple sequence alignment (Figure 3.25A), pLDDT (Figure 3.25B), and PAE (Figure 3.25C) output plots are provided by ColabFold AF-M but are not provided by the AF3 webserver. Top ranking structures for both models were then analysed by pLDDT scores (Figure 3.26A) and initial structural alignment with our crystal structure (Figure 3.26B). In general, both AF models exhibit high pLDDT scores in across the phosphopeptide complex, reflecting a high level of confidence in predicted atom positioning relative to the expected true structural positioning. Low scored pLDDT areas in the Peli 1_{FHA} domain include the loop separating β_1-β_2, the loops either side of the α-helix, and the unstructured region in the loop separating β_6-β_7. In the $MDC1_{pT765}$ phosphopeptide, AF-M predicts low pLDDT scores for the EDS N-terminal residues, while AF3 exhibits a low confidence about placement of the Glu N-terminal residue. In agreement with the pLDDT scores, both models structurally align to the crystal structure with extremely low root mean square deviation values (AF-M: 1.93Å and AF3: 2.66Å), reflecting a high level of structural similarity of the predicted models to the experimentally-derived structure.

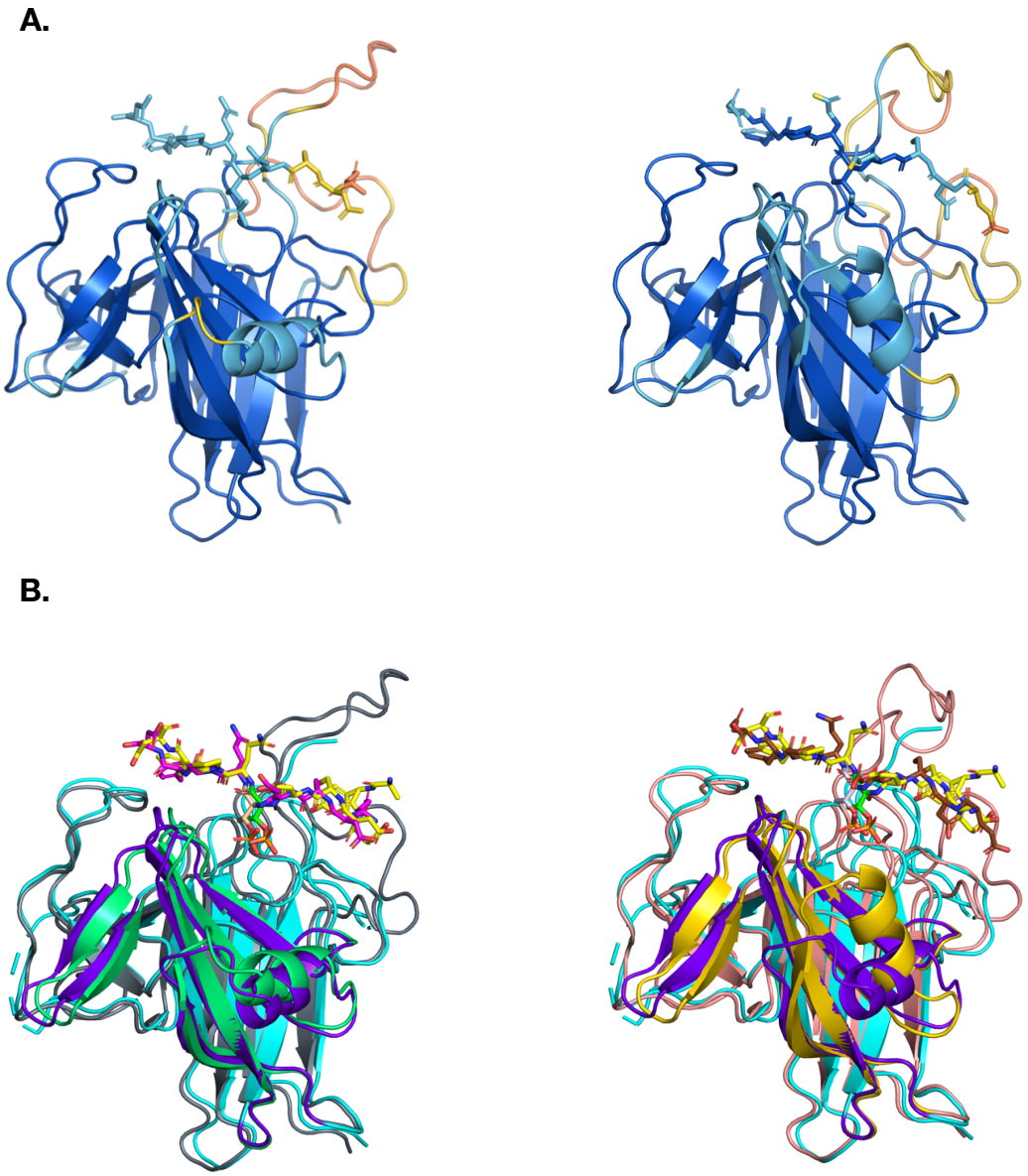


Figure 3.26 - AlphaFold 2 Multimer (AF-M) and 3 (AF3) models accurately predict the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex structure. (A) The AF-M (left panel) and AF3 (right panel) predicted Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex models are coloured according to per residue local confidence (pLDDT) score. Dark blue indicates very high confidence (pLDDT >90), light blue indicates high confidence (90 > pLDDT > 70), yellow indicates low confidence (70 > pLDDT > 50), and orange indicates very low confidence (pLDDT < 50). **(B) Structural alignment of the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex crystal structure against AF-M (left panel) and AF3 predictive (right panel) models demonstrate a high level of predictive accuracy.** The AF-M predicted model is depicted with the FHA core in slate grey, 'wing' in lime green, and $MDC1_{pT765}$ carbon backbone atoms are all coloured magenta, except the pThr, which is coloured in wheat. The AF3 predicted model FHA core is coloured in salmon, 'wing' in gold, $MDC1_{pT765}$ carbon backbone atoms are all coloured brown, except the pThr, which is coloured in blue white. The structures were aligned using the cealign command in the PyMOL Molecular Graphics System v3.1.1.

Both models were compared to the crystal structure more comprehensively using the Local-Global Alignment (LGA) program provided by the Amino Acid to Tertiary Structure (AS2TS) webserver (http://proteinmodel.org/AS2TS/LGA/lga.html). The LGA program was used to quantify the structural similarity between our experimental crystal structure and AlphaFold models, since the program includes structural comparison metrics such as a Global Distance Calculation Score (GDC), Local-Global Alignment Structural Similar Score (LGA_S), RMSD, and Global Distance Test High Accuracy Score (GDT_HA) after strict inter-atomic distance cutoffs (0.5Å, 1Å, 2Å, 4Å) (*Zemla et al., 2003; Keedy et al., 2010*) (Figure 3.27A and Table 3.5). Importantly, these evaluation metrics are often implemented in the biennial Critical Assessment of Structure Prediction (CASP) competitions to comprehensively assess the structural similarity between predictive models and experimental structures and are less quantitiatively influenced by outliers than simple RMSD values. Moreover, the interactions at the Pellino 1_{FHA}:$MDC1_{pT765}$ interface of each AF model were analysed using the PDBSum webserver (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/), and validated by manual investigated using PyMOL (*Laskowski et al., 2018;* Figure 3.27B).

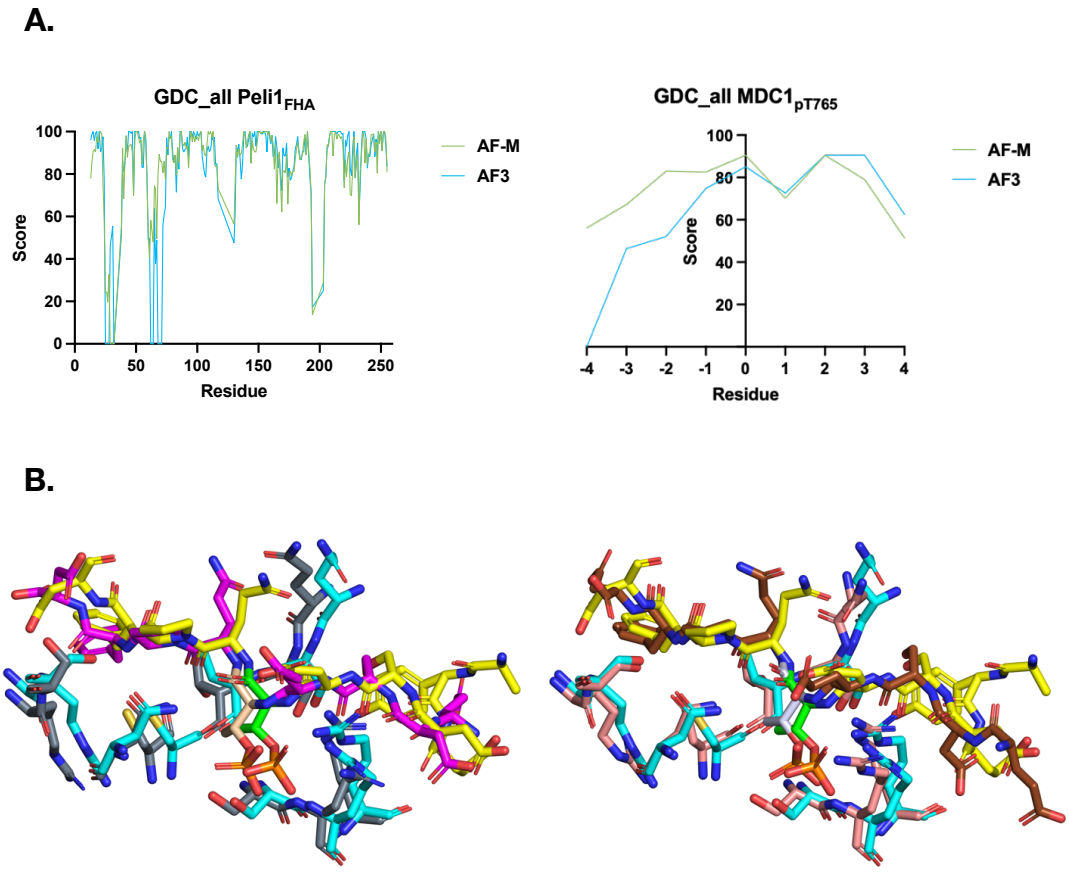


Figure 3.27 – LGA structural analysis of AF models compared to the crystal structure indicate that AF-M performs slightly better. (A) Local-Global Alignment (LGA) Global Distance Calculated (GDC) scores for all Pellino 1_{FHA} residue atoms (GDC_all) indicate that the AF-M model slightly outperforms the AF3 in predictive accuracy. GDC_all scores (y-axis) were calculated per residue (x-axis) using the LGA webserver. **(B) Structural alignment of Pellino 1_{FHA}:$MDC1_{pT765}$ interface residues support the AF-M model slightly outperforming AF3 in predictive accuracy.**

LGA analysis of both models after removal of the unstructured regions and the unnatural phosphopeptide AY linker illustrate the near perfection of both models (Table 3.5 and Figure 3.27), with AF-M performing marginally better than AF3 across almost all structural evaluation metrics. In particular, this was reflected in LGA_S (AF-M: 94.0% vs AF3: 90.7%), the more stringent GDT_HA (AF-M: 81.9% vs AF3: 78.9), GDC_all for pT765 residue atoms (AF-M: 90.5% vs AF3: 88.7%) and GDC_all atoms for interface-specific atoms (AF-M: 81.5±5.6% vs AF3: 84.6±10.1%).Clearly, these analyses demonstrate that both AF-M and AF3 are able to accurately predict the overall fold of the

complex, and is capable of placing the Peli-interacting pThr residue close to the true position and orientation. Minor inaccuracies are associated with interface residue sidechain atom placement, and residues that lie near to unstructured regions.

Area of Interest	Local-Global Alignment Parameter	AF-M	AF3
Overall Complex	Global Distance Calculation Main Chain Atoms (%)	87.9	85.5
	Global Distance Calculation All Atoms (%)	84.2	82.4
	Local-Global Alignment Structural Similarity Score	94.8	91.6
	Global Distance Test High Accuracy Score (at 0.5, 1, 2 and 4Å Distance Cutoffs)	81.9	78.9
	Root-Mean Square Deviation (Å, Structurally Aligned Residues)	1.14 (223/252)	1.11 (214/252)
pT765	Global Distance Calculation Main Chain Atoms (%)	90.5	88.2
	Global Distance Calculation All Atoms (%)	90.5	88.7
Complex Interface-Residues	Global Distance Calculation Main Chain Atoms (%)	85.5±5.2	90.0±8.2
	Global Distance Calculation All Atoms (%)	81.5±5.6	84.6±10.1

Table 3.5 - Summary statistics from the Local-Global Alignment (LGA) analysis of AF-M *versus* AF3 predictive models compared to the crystal structure of the Pellino 1_{FHA}:$MDC1_{pT765}$ complex. Of note, the complex interface values were calculated as the mean and standard deviation of global distance calculation main chain atoms and all atoms from only Pellino 1_{FHA} and $MDC1_{pT765}$ residues that were identified as significant contributors to the interface surface.

Both models succeeded in predicting the majority of highly conserved Peli 1:$MDC1_{pT765}$ interactions, including the hydrogen bond between Peli1-S135 and the pThr γ-phosphate group, the N186 and +1Q hydrogen bond, both of the +3F hydrogen bonds with Peli 1-N186/S223 and the supporting weak interactions from the basic platform, and the weak van der Waals interactions between the pThr γ-methyl group and Peli 1 surface pocket (Figure 3.27). Interestingly, AF3 was able to predict one of the conserved hydrogen bonds between Peli 1-R136 and the $MDC1_{pT765}$ γ-phosphate, however both models failed to predict any R104 pThr contacts. AF-M did not predict any pThr interactions with Peli 1-R136, because of R136 rotation away from the interface. Moreover, 2 hydrogen bonds were falsely predicted by both models, namely one between Peli 1-R37 and the -4E position, and Peli 1-R136 with the +4D.

The results of this experiment clearly demonstrate that both AF-M and AF3 are capable of predicting $Peli_{FHA}$:phospholigand complexes to a high level of accuracy. In this particular case, AF-M performed marginally better than AF3, but the prediction processing speed was drastically different, since AF-M outputted models after several hours, while AF3 outputted models within minutes. Importantly, the relative success of our AF predictions clearly depends on the high level of structural and sequence homology of FHA domains and their interactions with pThr-containing phosphopeptides, which were supported by template model building using ~70 pre-existing structure templates from the past ~25 years of FHA structural research.

3.5 Discussion

3.5.1 A revised model of Peli1/2 actvity in the DDR

Our data clearly supports the following updated Peli 1/2 DSB recruitment model (Figure 3.28). Upon IR-induced DSBs, the MRN complex is recruited to damage sites, leading to subsequent dissociation-induced activation of ATM kinase and interaction with NBS1 (*de Jager et al., 2001; Uziel et al., 2003; Dupré et al., 2006; Warren and Pavletich et al., 2022*). Active ATM kinase phosphorylates H2AX-S139 located at the C-terminal tail, yielding γH2AX, triggering recruitment of MDC1 via its C-terminal tandem BRCT repeats (*Burma et al., 2001; Stucki et al., 2005*). MDC1 functions as a scaffold to recruit a plethora of DDR proteins to augment ATM-signalling during IRIF formation. One such scaffolding property, is the DSB site retention of the MRN complex and ATM through NBS1-mediated interaction with a cluster of 6 Casein Kinase 2 (CK2) di-phosphorylated MDC1-SDTD-[V/A/D]-[D/E] motifs (*Chapman and Jackson, 2008; Lloyd et al., 2008*). ATM additionally phosphorylates MDC1 at TQxF motifs, which triggers recruitment of either RNF8 or Peli 1/2 through the highly specific canonical interactions between the respective typical and atypical-split FHA domains and pThr epitopes (*Huen et al., 2007*). At this juncture, the downstream repair pathway choice is determined by whether RNF8 or Peli 1/2 is recruited to the MDC1-pTQxF cluster. Following RNF8 recruitment, NHEJ or HR pathways are activated through processive K63-linked polyubiquitination-dependent recruitment of RNF168 and further DSB repair factors (*Gatti et al., 2015 Thorslund et al., 2015*). On the other hand, Peli protein recruitment is associated with concomitant ATM phosphorylation of Peli 1 S121 and T127 (Peli 2 S123 and T132) SQ/TQ motifs, which may act as an allosteric activator of Peli C-terminal RING activity (*Ha et al., 2019*). Peli activation is associated with K169, K202, and K266 autoubiquitylation, which are also targeted by

potential competitive Ubc9-directed SUMOylation, and formation of K63-linked polyubiquitin linkages on NBS1 at K686 and K690 (*Smith et al., 2008; Kim et al., 2011; Ha et al., 2019*). Polyubiquitinated NBS1 stabilises MRN and ATM complexes, which is fundamental for the subsequent recruitment of DNA-end resection factors, including CtIP, p-BRCA1, EXO1, RPA32, and Rad51, required for activation of HR (*Ha et al., 2019*).

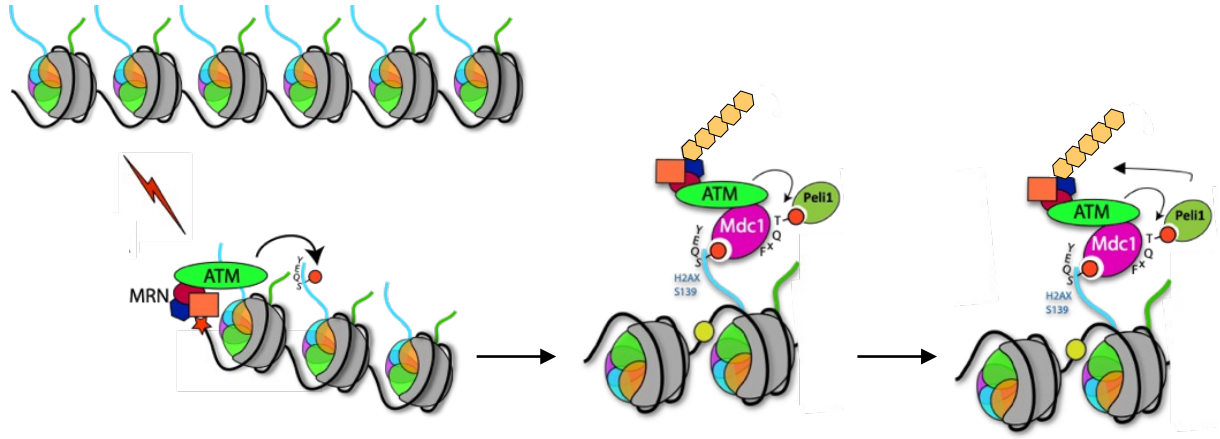


Figure 3.28 An updated model of ATM-dependent and MDC1-pTQxF-dependent Pellino 1 recruitment to DNA damage sites and Pellino 1-mediated DNA damage signalling. Upon Double-stranded DNA damage (lightning bolt), the MRN complex (red circle, blue hexagon and orange rectangle) is recruited, leading to dissociation of ATM kinase (green oval) into active monomers. Active ATM catalyses phosphorylation (red dot) of the C-terminal tail (cyan line) of H2AX at S139, forming γH2AX. MDC1 (pink egg) is recruited to γH2AX foci through its C-terminal tandem BRCT domain, whereupon ATM catalyses phosphorylation of MDC1 at the central cluster of 4 TQxF sites. Pellino 1 (dark green oval) is recruited to DNA damage sites by binding to the pTQxF sites via its N-terminal FHA domain. Pellino 1 then catalyses K63-linked polyubiquitination of NBS1 (gold hexagons), leading to stabilisation of the MRN complex and augmentation of ATM signalling.

Whilst our Peli DSB recruitment model conclusively disproves the previous γH2AX-dependent recruitment model, our understanding of the role of Peli in the DDR remains poorly understood. Our collaborators have recently shown that Peli is unable to functionally compensate for RNF8 loss, as demonstrated by a lack of 53BP1 IRIF formation upon shRNF8 treatment (data not shown). Moreover, analysis of 31 CRISPR screens against 27 different genotoxic agents demonstrate that RNF8 and Peli 1/2 knockout cells display significant differences in DNA-damage drug sensitivity

profiles (*Olivieri et al., 2020*). For example, RNF8 knockout cells are sensitive to bleomycin (enriched NHEJ repair response), camptothecin (enriched HR and FA/ICL repair responses), cisplatin (enriched HR and FA/ICL repair responses), potassium bromate (enriched HR and NHEJ repair responses), bleomycin (enriched NHEJ repair response), and IR (enriched NHEJ repair response). On the other hand, Peli 1 knockout cells were only moderately and minorly sensitive to bleomycin (enriched NHEJ repair response) or benzo(a)pyrene 7,8-diol-9,10-epoxide (enriched NER repair response), respectively, while Peli 2 knockout cells were insensitive to all clastogens. These data would imply that Peli 1 has a minor role in NHEJ and NER, while Peli 2 is functionally redundant within the DDR. The latter finding is at odds with the chromosomal instability phenotype observed by *Ha et al. (2019)* upon truncation of the Peli 1_{FHA} domain, and our data that demonstrate that Peli2 interacts with phosphorylated MDC1-TQxF motifs (*Ha et al., 2019*). Moreover, previous analyses have shown that Peli1_{FHA} reduces HR activity (*Ha et al., 2019*). There are various strategies that could resolve this issue, such as identification of Peli $1/2_{FHA}$-specific interactors or Peli $1/2_{RING}$-specific substrates from cells treated with different genotoxic agents. One such strategy involves Peli 1/2 loss-of-function experiments that depend on either Peli 1/2 siRNA-mediated knockdown or CRISPR-mediated double knockout cells, and complementation with exogenous mutant constructs. However, previous attempts at both strategies by our collaborators has proven technically challenging, and ultimately, unsuccessful (data not shown). An alternative approach may involve recombinant purification of site-directed photoactivatable cross-linker (PAC)-containing activity-based activity probes (ABPs), which have previously been used to characterise endogenous RING E3 activity profiles (*Hao et al., 2004; Mathur et al., 2020*). This platform depends on the use of amber stop codon suppression orthogonal translation systems, coupled with a structure-guided approach, to incorporate the PAC pyrrolysine derivative, AbK, into Peli or Ub site-specifically (*Hao et al., 2004;*

Mathur et al., 2020). Initial experiments could involve Abk incorporation at $Peli_{FHA}$:phospholigand or predicted substrate-RING-Ubc13~Ub interaction surfaces, within wild-type and FHA/RING mutant backgrounds, to stabilise both phosphorylation-dependent $Peli_{FHA}$ interactors and $Peli_{RING}$ substrates upon treatment with various DNA-damaging agents. The resulting observations from these experiments could, in turn, inform further biochemical and cell biology approaches to determine the specific role of Peli E3 activity within the DDR. For example, Peli forms K63-linked polyubiquitin chains on NBS1 at K686 and K690. K63-Ub chains are often implicated in recognition by ubiquitin-binding domains (UBDs). In support of this, SIRT1-mediated acetylation of the NBS1 K690 is associated with DNA repair inhibition, and phosphorylation of NBS1 at S278 and S343 (*Yuan et al., 2007*).

3.5.2 AlphaFold modelling of phospho-dependent Peli binding

Given that our AF models are fundamentally dependent on existing structures in the training set it is possible that this strategy may not succeed in predicting novel and/or more divergent interactions. Nonetheless, AF is unrivalled in its capacity to rapidly predict highly accurate models that can inform structure-function relationships that underlie complex biological pathways, such as the DDR. Consequently, both AF-M and AF3 could be used to predict models that aid a more in-depth understanding of the relationship between Peli family structures and apparent non-redundant functional roles within both DDR and TLR/IL-R signalling pathways. This may be explored by investigating the differences in binding preferences across Peli family members with MDC1-TQxF and IRAK1 phosphopeptides, as well as other physiological interactors, such as RIP1 and TRAF6, which have not yet been explored at the molecular level. Additionally, AF-M/AF3 may be used to resolve the potential autoinhibitory role of the Peli 3a N-terminal insertion, the binding contributions

of multiple phosphorylated epitopes with respect to the Peli 'wing' appendage, and identification of key sites of RING regulation sites (*Huoh and Ferguson, 2014*). For example, AF may identify the R linchpin residue, which is fundamental for promotion of the 'closed-state' substrate-RING E3-E2~Ub transfer complex (*Pruneda et al., 2011; Pruneda et al., 2013*). Additionally, the role of 'wing' phosphorylation during RING activation could also be explored. Indeed, previous research appears to suggest that the 'wing' acts as an allosteric phosphorylation-dependent rheostat of E3 RING activation (*Smith et al., 2009*). Ultimately, these predictive models will likely accelerate FHA and Peli-specific research, but it is vital that prediction models are validated by extensive comparison with experimental data.

Chapter 4

Structural and Biochemical Characterisation of a Potential Phosphorylation-Dependent Interaction Between PLK1 and Rad50

4.1 Introduction

Phosphorylation is highly redundant and relatively non-specific, and the function of specific sites is often unclear. The work described in this chapter is based on the results of a novel and direct quantitative proteomics approach to identifying phospho-interactors with selected, conserved phosphomotifs by the the Hořejší lab who identified potential novel phosphorylation-dependent interactors of casein kinase 2 pS/pT-E/D-x-E motifs, located within 4 DDR proteins, namely Rad50-pT690, breast cancer type 1 susceptibility protein (BRCA1)-pS1336, MutL protein homolog 1 (MLH1)-pS477, and *Xeroderma pigmentosum* group C-complementing protein (XPC)-pS883/pS884 (*Dullovi et al., 2023*). Our interest has focussed on Rad50 pT690 since 3 of the top 10 proteins, PLK1, FOXK1, Bub3, with the largest significant fold-change in binding between phospho and non-phospho Rad50 peptides were known to have phospho-binding activities with vastly different specificities. Of these three, the Smerdon lab has a longstanding interest in Plk1 structure and function and the MRN complex (*Elia et al., 2003; Yata et al., 2012*). We therefore focussed on trying to understand how Rad50 pT690 interacts with Plk1 with a view to investigating the physiological significance and implications of this phosphorylation event for regulation of Rad50, and thus the MRN complex.

4.1.1 Rad50 and the Mre11/Rad50/Nbs1 (MRN) DNA repair complex

The MRE11-Rad50-Nbs1 (MRN) complex is one of the primary sensors of DSB-mediated DNA damage through recognition and binding of broken DNA, whereupon it stimulates and amplifies the DNA damage response cascade through NBS1-mediated recruitment of ATM. Typically, the MRN complex exists as either a 2:2:1 heteropentameric or 2:2:2 heterohexamic complex, which is conserved throughout eukaryotes, however the catalytic core -comprised of the M_2R_2 tetramer- is

found in all domains of life. The essentiality and interdependent nature of the MRN complex is absolute, as underscored by embryonic lethality induced by knockdown of individual subunits.

The global structure of the MRN is composed of a globular catalytic head core with two flexible 60-nm long Rad50 coiled-coil domain arms that extend away to the apex of the structure (Figure 1.1 top left panel; *Rotheneder et al., 2023*). At the apex of the structure is located the Rad50 zinc-hook motif that chelates a single Zn^{2+} ion through highly conserved CxxC motifs to facilitate homo-dimerisation (*Park et al., 2017*). The catalytic head core is further derived of an ATP-dependent DNA binding and nucleolytic processing centre, contributed by the Rad50 Walker A/B ATPase domain halves and MRE11. NBS1 is located at the catalytic head through two C-terminal interaction sites with the MRE11 N-terminal nuclease domain (Figure 1.1 top right panel; *Park et al., 2011; Schiller et al., 2012*).

Meiotic recombination 11 homolog 1 (MRE11) is a 708-amino acid protein (80kDa) that is a core component of the MRN complex (Figure 4.1 bottom panel). In its dimeric forms, MRE11 acts a structural scaffold through independent Rad50 and Nbs1 interaction surfaces. The N-terminal region of MRE11, consisting of amino acids amino acids 1-411, is conserved throughout all species and consists of an N-terminal calcineurin-like phosphoesterase domain and accessory DNA binding cap domain (Figure 4.1 bottom panel). The phosphoesterase or nuclease domain, formed by residues 1-300, catalyses Mn^{2+} or Mg^{2+}-dependent 3'-5' exonuclease activity during short-range DNA-end back-resection (Figure 4.1 bottom panel); *Paull and Gellert, 1998*). The MRE11 phosphoesterase also possesses single-stranded and double-standed DNA endonuclease activity. These nuclease activities can occur independently of Rad50; however, endonuclease activity is further stimulated by Rad50 binding in the presence of ATP (*Paull and Gellert, 1998; Trujillo and*

Sung, 2001). In addition, amino acids 80-121 contribute to binding of the C-terminal region of Nbs1 (Figure 4.1 bottom panel); *Park et al., 2011*). The nuclease domain is followed by a cap domain, which provides DNA substrate selectivity and is involved in binding to a central portion of the Rad50 Walker B ATPase motif (*Lammens et al., 2012*). In the intervening region between N- and C-terminal halves is located a DNA-binding domain (residues 407-421; Figure 4.1 bottom panel). The C-terminal half of MRE11, corresponding to residues 421-708, is unique to eukaryotes. This region is involved in binding to the N-terminal side of the Rad50 Walker A/B ATPase motifs (residues 447-490) and contains a glycine-arginine-rich (GAR) domain (Figure 4.1 bottom panel); residues 556-600) (*Déry et al., 2008; Lammens et al., 2012*). In particular, the 6 R residues are targeted by protein arginine methyltransferases (PRMTs) to promote MRE11-mediated DNA-binding and nuclease activity, prior to assembly of the MRN (*Déry et al., 2008; Yu et al., 2012*). A second DBD is located at the MRE11 C-terminus (Figure 4.1 bottom panel); residues 643-692).

Nijimegen breakage syndrome 1/Nibrin (Nbs1) is a 754-amino acid (85kDa) functional adaptor module of the MRN complex (Figure 4.1 bottom panel). Nbs1 is characterised by an N-terminal 'super-modular' architecture (residues 1-316), derived from an FHA domain (Figure 4.1 bottom panel; residues 1-111), followed by two tandem BRCT repeats (Figure 4.1 bottom panel; BRCT1 residues 112-196 and BRCT2 residues 225-316) (*Lloyd et al., 2009*). This region is involved in binding to a central cluster of di-phosphorylated SDTD motifs, located in MDC1 that are constitutively phosphorylated by casein kinase 2 (CK2) (*Chapman et al., 2008; Spycher et al., 2008*). This interaction facilitates constitutive association of MDC1 with the MRN complex, which facilitates retention and further recruitment of the MRN following initial DSB-dependent recruitment (*Chapman et al., 2008; Spycher et al., 2008*). The aforementioned MRE11-binding site is located at

the C-terminus of Nbs1 (residues 547-695), followed by the ATM FxF/Y recruitment motif (Figure 4.1 bottom panel; residues 740-749) (*You et al., 2005; Schiller et al., 2012*).

Rad50 is a 1312-amino acid (153kDa) protein that structurally resembles the structural maintenance of chromosome (SMC) family of scaffolding proteins that acts an ATP-dependent molecular machine, due to the presence of N-terminal (residues 1-180) and C-terminal (residues 1155-1312) walker A/B nucleotide-binding motifs (Figure 4.1 bottom panel). As mentioned, these motifs form a symmetric globular ABC-type ATPase domain that is involved in hydrolysis of ATP and is essential for MRE11-mediated nuclease activity. The ATPase halves are separated from one another by a set of heptad repeats that form a 60nm long anti-parallel CC domain. At its apex is located a hook domain (Figure 4.1 bottom panel; residues 674-690) that contains the invariant CxxC motif (residues 681-684), which is involved in Zn^{2+}-coordinated chelation-mediated Rad50 homodimerisation. The CxxC motif and local Zn-hook residues are essential for MRN structural integrity, since hypomorphic Zn-hook residues mutants impair *S. cerevisiae* Rad50 dimerisation, ATM signalling, and DNA damage tolerance (*Hohl et al., 2015*). This intramolecular Rad50 complex rod-shaped dimer is the predominant structural Rad50 conformation, however a zinc-hook dependent intermolecular tetrameric complex has been proposed to tether DNA strands of up to 120nm during siter chromatid-exchange mediated DSB repair (*Park et al., 2017*) (Figure 4.1B). This hypothesis is supported by genetic analyses, transmission electron microscopy, and cryo-EM structures of homologous Rad50 proteins (*Hopfner et al., 2002; Kissling et al., 2022; Rotheneder et al., 2023*).

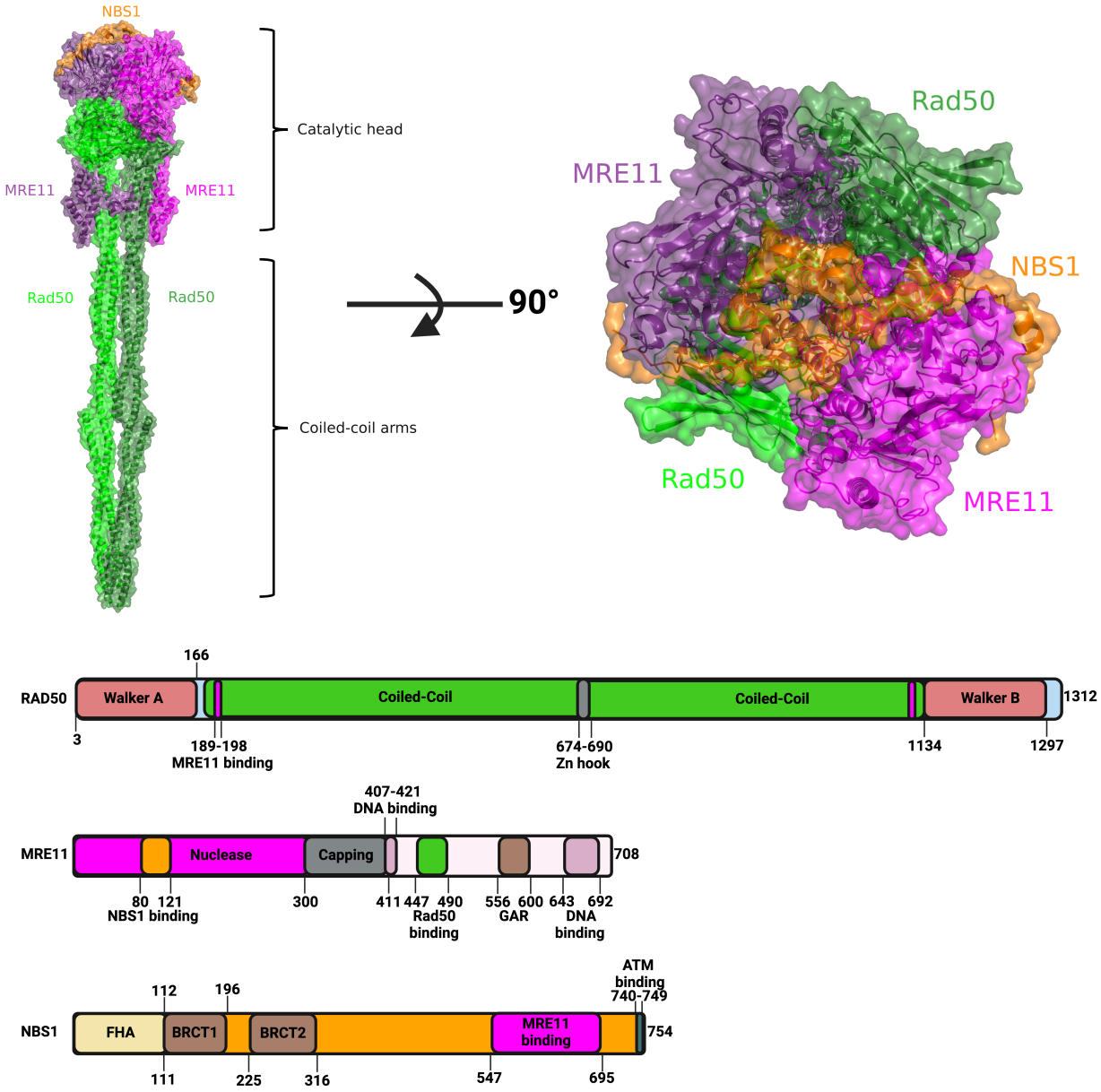


Figure 4.1 – The MRN complex. The global cryo-EM structure of the *Chaetomium thermophilum* heteropentameric (2:2:1) MRN complex (top panel). MRN complex members secondary structures are presented in the cartoon format and surface area representation with 40% transparency (top left panel). The catalytic head and coiled-coil arms are also indicated. The structure was rotated about the x-axis by 90° to highlight the catalytic head structural composition (top right panel). The Rad50 region comprising the extended Zn-hook coiled-coil, the MRE11 C-terminus and NBS1 N-terminal FHA-BRCT super-module are absent in this structure. Figure created using the PyMOL Molecular Graphics System v3.1.1 (Schrödinger, LLC). The domain structure of individual MRE11, Rad50, and NBS1 members (bottom panel). Figure created with Biorender.com.

The focus of this chapter, Rad50-T690, is located in a structured region that is adjacent to the zinc-

hook (Figure 4.2). The effect of T690 phosphorylation alone on Rad50 zinc-hook stability, tertiary and quaternary structurisation, and zinc coordination has been previously explored in the context of short 42-amino acid (residues 665-706) and longer 183-amino acid (residues 585-766) peptides (*Tran, Padjasek, and Krężel, 2022*). Interestingly, these analyses demonstrated that Rad50 T690 phosphorylation, or substitution with a phosphomimetic glutamatic acid residue, has a major destablising effect on ZCC secondary structure in the context of smaller peptides, whereas there was no effect on ZCC secondary structure in extended peptides (*Tran, Padjasek, and Krężel, 2022*). Moreover, phosphorylation and the T690E phosphomimetic exhibited a 0.7-fold decrease in small ZCC dimers, while the T690E phosphomimetic had a minor 0.2-fold destabilising effect on the extended ZCC dimer. However, in all cases the T690E phosphomimetic adopted a rod-shaped dimer with a tetrathiolate-coordinated zinc-ion alike the non-phosphorylated counterpart (*Tran, Padjasek, and Krężel, 2022*).

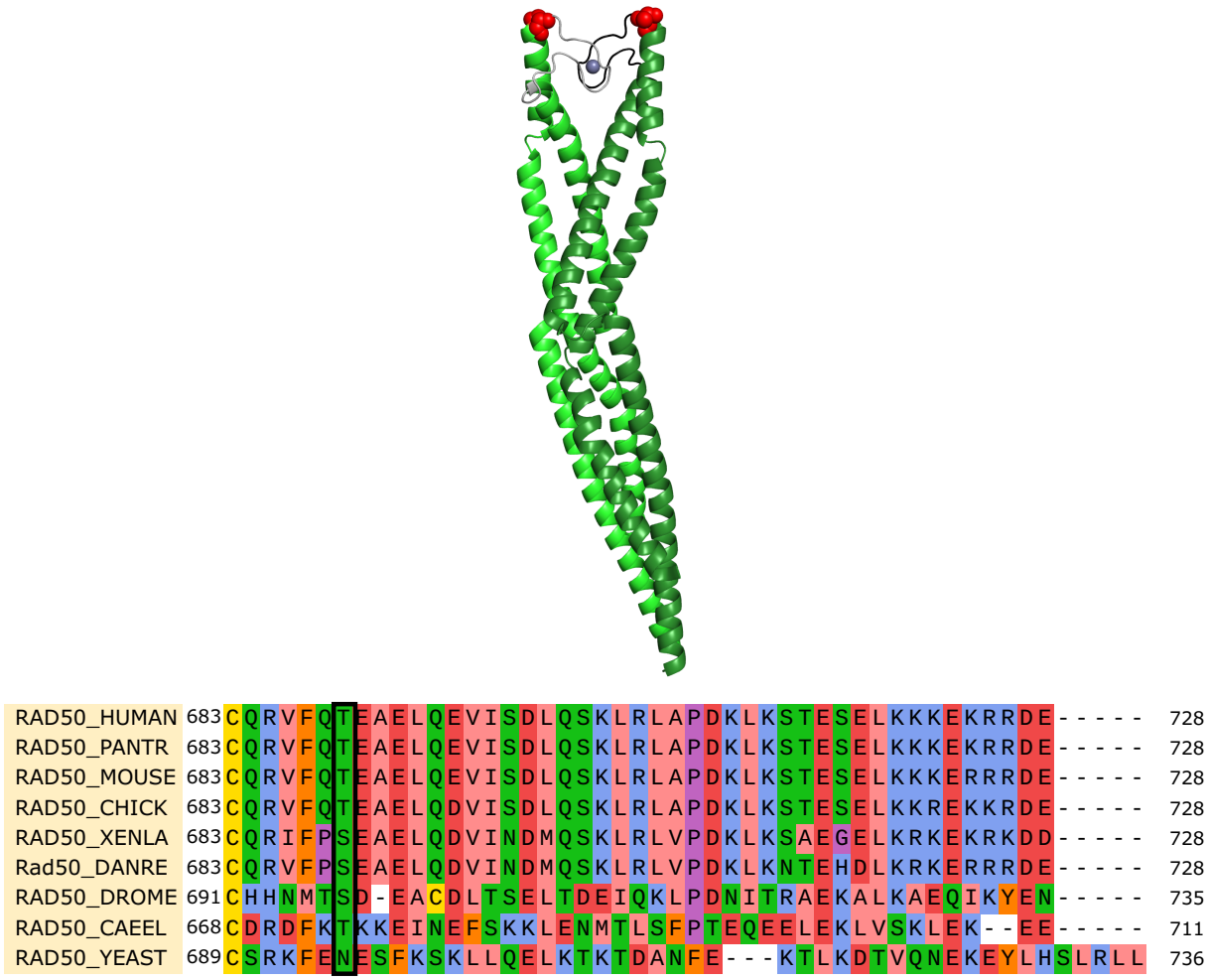


Figure 4.2 - The Rad50 Zinc hook coiled-coil domain (ZCC) and T690 site are highly conserved. (Top panel) The crystal structure of the homodimeric ZCC fragment. Accession code: 5GOX. Both ZCC monomer secondary structures are represented in the cartoon format and individual monomers are coloured in light green and dark green, respectively. Each Zn hook are coloured in light grey and dark grey with the chelated Zn^{2+} ion represented as a dark grey sphere. All atoms for the T690 target phosphorylation site from each monomer are represented as red spheres. **(Bottom panel) Evolutionary conservation of Rad50-pT690.** Individual amino acids are coloured by physicochemical properties.

4.1.2 Polo-like kinase

Polo-like kinase 1 (PLK1) is a highly conserved serine/threonine kinase that orchestrates cell cycle progression (*Hartwell et al., 1973; Sunkel and Glover, 1988, Lake and Jelinek, 1993; Goldsteyn et al., 1994*). PLK1 consists of an N-terminal bilobal kinase domain and a unique non-catalytic C-terminal phospho-binding domain, which is crucial for the spatiotemporally regulated and phosphorylation-

dependent targeting of PLK1 to specific cellular structures (*Goldsteyn et al., 1994; Lee et al., 1998*). As a critical cell cycle regulator, PLK1 kinase activity carefully coordinates several key processes involved in safeguarding the genome. Therefore, PLK1 expression and activity is tightly regulated through molecular phospho-recognition and the opposing activities of inhibitory phosphatases and stimulatory kinases (*Cheng et al., 2003; Elia et al., 2003; Macůrek et al., 2008*). PLK1 mRNA levels oscillate in a cell cycle-dependent manner, with low levels in G_1, rising in S phase, and then peaking at the G_2/M checkpoint to stimulate mitotic entry (*Goldsteyn et al., 1994*). In early S phase, PLK1 localises to centrosomal and chromatin structures, wherein PLK1 activity couples centrosomal duplication and DNA replication (*Song et al., 2011; Shen et al., 2013*). As cells transition to late S phase and G_2, PLK1 activity mediates DDR activation and then later redistributes from chromatin to the nuclear envelope in preparation for mitosis (*Yata et al., 2012; Ghegiani et al., 2017; Kagami et al., 2017 Linder et al., 2017*). During G_2 checkpoint recovery, PLK1 and cyclin-dependent kinase 1 (CDK1)-mediated phosphorylation of DDR components inhibit HR and NHEJ DNA damage repair pathways to prioritise accurate chromosomal segregation (*Esashi et al., 2005; Li et al., 2017*). Interestingly, recent evidence has emerged that demonstrate PLK1 activity can stimulate mitotic alternative-end joining-mediated DSB repair in HR-deficient cells (*Gelot et al., 2023*). During prophase and metaphase, PLK1 relocalises to kinetochores and mitotic spindle structures to ensure amphitelic spindle attachments and proper alignment of sister chromatids at the metaphase plate (*Martin et al., 2011; Qiao et al., 2016; Zhang et al., 2016; Lera et al., 2019; Singh et al., 2021*). Following spindle assembly checkpoint satisfaction, PLK1 localises to the spindle midzone to regulate accurate sister chromatid segregation, and then localises during telophase to cytokinetic midbody structures to coordinate successful daughter cell division (*Hu et al., 2012; Adriaans et al., 2019*). Historically, the role of PLK1 has predominantly focused on its mitotic function, where it

interacts with >600 substrates in a phosphorylation-dependent manner and phosphorylates >2000 substrates (*Lowery et al., 2007; Grosstessner-Hain et al., 2011*). Consequently, dysregulation of PLK1 activity or PLK1 overexpression is intimately linked with genomic instability, cancer metastasis, chemotherapeutic resistance, and ultimately, poor patient prognosis (*Liu, Sun, and Wang, 2017*).

4.1.3 PLK1: The Polo Box Domain

PLK1 is a 63kDa globular protein, encoded by 603 amino acids, that features an N-terminal kinase domain and a non-catalytic C-terminal phospho-binding module (Figure 4.3A). Early evidence found that the C-terminal region was highly conserved across PLK1 orthologues, and was composed of two regions that share a moderate level of sequence homology (*Clay et al., 1993*). The C-terminal region was subsequently termed the Polo Box domain (PBD), and each region was designated PB1 and PB2 (*Lee et al., 1998*). Further investigation elucidated that despite presence of a functional kinase domain, knockout of the PLK1-PBD or a W414F mutation was sufficient to abrogate the ability of PLK1 to localise to the spindle poles and midbodies, and rescue mitotic defects in Cdc5-deficient *S. cerevisiae* cells (*Lee et al., 1998; Seong et al., 2002*). Similarly, mitotic localisation of human PLK1 and murine PLK4 was disrupted upon knockout of individual PB regions, which suggested that the presence of both PBs were required for localisation of mitotic PLK1 (*Hudson et al., 2001; Seong et al., 2002*). Proteomic screens using synthetic oriented pT-P peptide and phosphopeptide libraries found that PLK1 contained a phosphopepitde binding domain, which interacted independently of the kinase domain (*Elia, Cantley, and Yaffe, 2003*). Through a series of truncation mutants, *Elia et al.,* found that the minimal pT-P binding motif was comprised of both PB1 and PB2 and a linker region (residues 326-603) that connects the kinase domain and C-terminal PBD. Additionally, degenerate phosphopeptide libraries containing a fixed pT-P motif found that the PBD specifically selects for a S residue in the pT -1 peptide position and moderately selects for a +1 P residue (*Elia, Cantley, and Yaffe, 2003*). These findings

were crucial for the definition of the PBD interaction consensus sequence, [P/F]- [Φ/P]-[Φ]-[T/Q/H/M]-S-[**pS/pT**]-[P/X], and subsequent identification of an optimal phosphopeptide, MQS**pT**PL, which tightly interacted with PBD (K_D: 280nM) in a 1:1 stoichiometry (*Elia, Cantley, and Yaffe, 2003*). Additionally, preincubation with the optimal phosphopeptide ablated centrosomal localisation of recombinant PLK1-PBD. The minimal PBD structured region was proteolytically defined (residues 367-603), and this contained the last 45 residues of the kinase domain-PB1 linker, termed the Polo-cap (Pc), although it was not required for phosphopeptide binding (*Elia et al., 2003*). In 2003, the crystal structures of apo-PBD and PBD in complex with the optimal phosphopeptide were solved (*Cheng et al., 2003; Elia et al., 2003*) (Figure 4.3B left panel). These structures elucidated that the PBD adopts a unique fold composed of two symmetric PB folds composed of a 6-stranded anti-parallel β sheet and α-helix motif that form a central 12-stranded β-sandwich with 3 outer α-helical segments (Figure 4.2B left panel). Additionally, the N-terminal Pc contributes to the structure through a loop, 3_{10}-helix, and a short linker loop (L1) that connects to the PB1 β_1 strand (Figure 4.3B left panel). The Pc was found to encircle PB2 to tether it to PB1 through the antiparallel arrangement of Pc α_1 and PB2 α_3 and packing of Pc α_1 and Pc 3_{10} helix against a short PB2 α segment. PB1 and PB2 are connected by a second 32-amino acid linker (L2), which runs in an antiparallel fashion to L1 between the PBD β-sandwich (*Cheng et al., 2003; Elia et al., 2003*) (Figure 4.3B left panel). Interestingly, the PBD does not undergo any major conformational change upon phosphopeptide binding, however the L2 loop adopts a structured conformation (*Cheng et al., 2003*). The phosphopeptide binds in an extended conformation in a groove between PB1 and PB2 and contacts residues from PB1 β_1, PB1 α_2, the N-terminal region of L2, PB2 β_7, and the PB2 loop between β_9-β_{10}. The β_9-β_{10} loop residues H538 and K540 interact directly with the pT γ-phosphate group oxygen atoms through 3 side chain hydrogen-bonding contacts that form a pincer-like arrangement (Figure 4.3B right panel; *Elia et al., 2003*). Importantly, mutation of H538 to a M residue and K540 to an A residue abrogated phosphopeptide interactions, PBD-mediated targeting of PLK1 to centrosomes, and PLK1 overexpression-mediated aneuploidy (*Elia et al., 2003*). A further 5 hydrogen bonds are formed by indirect interactions, via a lattice of water molecules, through side chains from V411, S412,

W414 N533, F534, and R557 (Figure 4.3B right panel; *Elia et al., 2003*). The absolute requirement for a pT -1S residue and functional importance of W414 is derived from two -1S side chain hydrogen bonds with W414 main chain atoms, an indirect water molecule-mediated hydrogen bond with the main chain of L491, and favourable van der Waals interactions between the -1 S Cβ atom with the W414 side chain Cδ1 atom (Figure 4.3B right panel, *Cheng et al., 2003; Elia et al., 2003*).

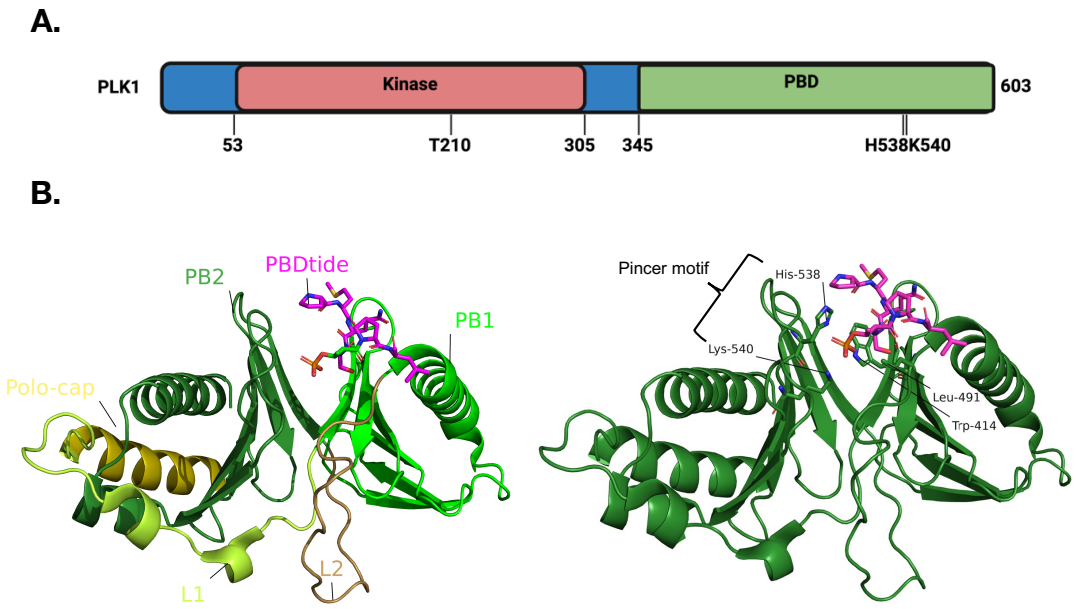


Figure 4.3 - The PLK1 Polo Box Domain. (A) The domain structure of PLK1. Kinase domain (salmon) and PBD (green) boundaries, and key residues involved in kinase activation (T210) and phospho-ligand binding are indicated below the domain map. **(B) The crystal structure of PLK1-PBD bound to its optimal synthetic phosphopeptide.** Accession code: 1UMW. In the left panel, the PBD domain secondary structure is represented in the cartoon format and individual fragments that comprise the domain are coloured. In the right panel, PBD residues that comprise the Pincer motif, involved in direct binding to pSer and pThr epitopes, are indicated, as well as PBD-W414 and L491 residues that directly contact the absolutely conserved pThr/pSer -1Ser residue.

These structures collectively revealed the molecular basis by which PBD-mediated pT/pS substrate recognition is responsible for the intracellular localisation of PLK1. Moreover, these findings and biochemical experiments suggested that unbound PBD might act as a negative regulator of kinase domain activity. This regulation likely occurs through autoinhibitory intramolecular interactions between both domains, since C-terminally truncated PLK1 mutants exhibit higher levels of kinase

activity than full-length PLK1 (*Elia et al., 2003*). Structural analyses of PBD crystal packing interactions have proposed an additional layer of PBD-mediated substrate recognition (*Śledź et al., 2011*). The proposed model involves a cryptic hydrophobic tyrosine pocket formed by highly conserved Y417, Y421, Y481, Y485 residues, located adjacent to the pT/pS-binding interface, which enables recognition of extended phosphorylated PBD substrates that contain hydrophobic residues (*Sharma et al., 2019*). Indeed, overexpression of an N-terminally GFP tagged Y421A-L478A-Y481D Y pocket mutant with an intact pT/pS binding interface disrupted phosphorylation-dependent interactions with Polo-box-interacting protein 1 (PBIP1), induced mitotic arrest due to mitotic spindle defects, and increased mitotic catastrophe-mediated cell death events (*Sharma et al., 2019*). Similarly, treatment with Polytryin, a Y pocket small-molecule inhibitor, disrupted PBIP1-derived phosphopeptide interactions and induced mitotic arrest in a dose dependent manner (*Sharma et al., 2019*).

4.1.4 PLK1: The Kinase Domain

As mentioned in the previous section, PLK1 recruitment to mitotic structures is dependent on recognition of phosphorylated motifs substrates by the C-terminal PBD, and PBD-mediated recognition likely relieves autoinhibitory interactions with the kinase domain (*Elia et al, 2003*). The substrate consensus motif for PLK1 kinase domain-mediated phosphorylation, [D/E]-x-S/T-[ϕ]-x-[D/E], was elucidated by systematic alanine and glycine scanning mutagenesis of peptides derived from the putative mitotic entry substrate, cell division cycle 25C phosphatase (Cdc25C) (*Nakajima et al., 2003*). The major details of the mechanism of activation of the kinase domain were later resolved by the crystal structure of the PLK1 kinase domain in complex with a nonhydrolysable ATP analog (accession code: 2OU7), and comparison with structures of inactive and activate state

Hanks kinases (*Hanks and Hunter, 1995; Kothe et al., 2007; Taylor and Kornev, 2011*). The N-terminal kinase domain is formed by residues 13-345, and adopts a typical bilobal kinase fold, wherein ATP and Mg^{2+} cofactors bind in the cleft between each lobe. The N-terminal lobe, corresponding to residues 37-131, is derived from a loop, β-strand, and core 5-stranded antiparallel β-sheet that is split by a 3_{10}-helix and α-helix insertion between β4-β5 (*Kothe et al., 2007*). In particular, the N-terminal region of the α-helix, or so-called C-helix, interacts with the T activation loop for poised catalysis upon kinase activation. The N-lobe contains two highly conserved AxK and GxG motifs within the first 3 β-strands of the core β-sheet. The K82 residue of the AGK motif is involved in coupling the ATP phosphate groups with the C-helix residue E101, to such an extent that the K82M mutant is sufficient to inhibit PLK1 kinase activity (*Goldsteyn et al., 1995; Lee et al., 1995; Kothe et al., 2007; Taylor and Kornev, 2011*). The G-rich loop connects the adenine-accommodating β_2-β_3 strands, and throughout protein kinases the G-rich loop is highly flexible and involved in positioning γ-phosphate for catalysis (*Kothe et al., 2007; Taylor and Kornev, 2011*). Residues 138-330 form the C-terminal lobe, which is predominantly alpha-helical due to 8 α-helices, 3 3_{10}-helices and 2 small anti-parallel sheets. The C-lobe of PLK1 features the HRD motif-containing catalytic loop, DFG motif and T210-containing T-loop activation segment, and the substrate-binding P+1 loop (*Hanks and Hunter, 1995; Kothe et al., 2007*). Residues 174-181 between β_8 and 3_{10}-helix 2 comprise the catalytic loop with the HRD motif and K178 located at the N-terminus. The HRD motif D residue acts as a catalytic base, accepting protons released from phosphorylated substrates, while K178 is involved in hydrogen-bonding with the P+1 substrate recognition loop residue T214 in active state kinases (*Kothe et al., 2007; Taylor and Kornev, 2011*). In the PLK1 kinase domain crystal structure, the P+1 loop is rotated away from K178. The PLK1 T-loop activation segment is formed by residues 194-230, and contains an N-terminal DFG anchor and the phosphorylation-mediated T210

activation residue (*Kothe et al., 2007*). The PLK1 structure adopts an inward-facing DFG conformation, such that D174 is positioned ready to interact with active site Mg^{2+} and ATP phosphate atoms. The T-loop is highly dynamic or disordered in inactive kinases, therefore the T210 residue was mutated to V during crystallisation of the PLK1 kinase domain (*Kothe et al., 2007*). The γ-phosphate of the pThr residue in active Hanks kinases interacts with a conserved K residue in β_9 (PLK1- β_{10} K191), and the R residue of the catalytic loop HRD motif, to stabilise the activation loop (*Kothe et al., 2007*). T210 is phosphorylated upon mitotic entry through CDK1 and Bora-mediated activation of Aurora A kinase (*Macůrek et al., 2008*). Specifically, CDK1 phosphorylates the Aurora A co-activator Bora at S112, which in turn, stimulates the interaction between Bora and Aurora A, leading to Aurora A-mediated phosphorylation of PLK1-T210 (*Vigneron et al., 2018*). T210 is also phosphorylated by Chk1 in response to ionising radiation, and this triggers PLK1-mediated phosphorylation of Rad51 during HR activation (*Yata et al., 2012; Peng et al., 2021*). Recently methyltransferase G9a-mediated monomethylation of K209 was shown to antagonises PLK1-T210 phosphorylation-mediated activation under DNA damage conditions, in order to finetune the DDR (*Li et al., 2019*).

4.1.5 The PLK Family and PBD function

The human serine/threonine PLK family consists of 4 catalytic members (PLK1-4), and a fifth non-catalytic member, PLK5 (*Goldsteyn et al., 1994; Li et al., 1996; Karn et al., 1997; Liby et al., 2001; Andrysik et al., 2010*). Canonical PLK1-4 are defined as a distinct group of serine/threonine kinases that are composed of an N-terminal Hanks kinase domain, and at least one copy of the unique $\beta_6\alpha$ polo box region, involved in binding to substrates containing S-pSer/pThr motifs (*Hanks and Hunter, 1995; Elia et al., 2003*). Members of the PLK family share relatively low sequence identity, exhibit

distinct expression patterns and cellular functions, thereby demonstrating a lack of functional redundancy across members (*Winkles and Alberts, 2005; de Cárcer et al., 2011*).

The PLK family can be divided into 3 subfamilies based on sequence conservation and evolutionary conservation, with PLK1 and PLK4 in subfamilies of their own (*de Cárcer, Manning, and Malumbres, 2011*). PLK1 is evolutionarily conserved throughout amorphea and archaeplastida eukaryotic supergroups, except across embryophyta lineages (*Okamura et al., 2017; Burki et al., 2020*). In particular, prototypical PLK1 is the most conserved member of the PLK family with orthologues in unicellular yeasts (*Saccharomyces cerevisiae:* Cdc5, *Schizosaccharomyces pombe*: Plo1) and phycophytes (*Cyanidoschyzon merolae*: PLK), as well as complex metazoans (*Drosophila melanogaster*: Polo, *Caenorhabditis elegans:* Plk1, *Xenopus laevis*: Plx1, *Mus musculus*: PLK, and *Homo sapiens*: PLK1) (*Hartwell et al., 1973; Sunkel and Glover, 1988; Lake and Jelinek et al., 1993; Ohkura et al., 1995; Kumagai and Dunphy, 1996; Chase et al., 2000; Okamura et al., 2017*).

PLK2/SNK, PLK3/FNK, and PLK5 represent a separate subfamily that share 38.1%, 38.4%, and 26.5% sequence identity with PLK1, respectively (*de Cárcer, Manning, and Malumbres, 2011*). All members of this subfamily are only found in bilaterian animals and feature two C-terminal Polo Box motif copies, however PLK5 contains no N-terminal kinase domain due to a premature stop codon in exon 6 (*de Cárcer, Manning, and Malumbres, 2011; Andrysik et al., 2010*). Mice deficient in PLK2 or PLK3 are viable, which suggests that neither members are essential or are replaceable by other members or kinases (*Ma, Charron, and Erikson, 2003; Myer et al., 2011*). PLK2 mRNA levels peak during the G_1/S phase transition, at which point PLK2 is targeted to centrosomes in a PBD-dependent manner to indirectly facilitate centriole duplication (*Winkles and Alberts, 2005; Cizmecioglu et al., 2012*). PLK2 phosphorylates F-box protein WD repeat-containing protein 7 (Fbxw7), leading to its

subsequent degradation by the 26S proteasome (*Cizmecioglu et al., 2012*). Fbxw7 is an E3 that targets cyclin E for proteasomal degradation, therefore PLK2-mediated destablisation of Fbxw7 results in stabilisation of cyclin E. Stabilised cyclin E then associates with its cognate binding partner, CDK2, to activate CDK2-mediated phosphorylation of centriole duplication regulators, such as nucleophosmin (*Okuda et al., 2000; Cizmecioglu et al., 2012*). PLK3 expression is induced by mitogen stimulation while its mRNA levels remain constant throughout the cell cycle (*Winkles and Alberts, 2005*). PLK3 kinase activity is activated in response to nocodazole-mediated G_2/M phase arrest, where it associates with depolymerised tubulin, but also in response to ionising-radiation induced DDR activation (*Bahassi et al., 2002*). Genotoxic stress-mediated activation of PLK3 is driven by ATM phosphorylation, enabling PLK3 to function as an effector kinase that phosphorylates Chk2 and p53, thereby activating G_1/S and G_2/M checkpoints arrests (*Bahassi et al., 2002*). PLK5 expression is restricted to the central nervous system, whereby it is predominantly expressed in differentiated brain tissue cells, such as neurons and the glia (*de Cárcer et al., 2011*). PLK5 appears to function as a tumour suppressor in brain tissue, such that its overexpression activates pro-apoptotic signalling in glioblastoma cells (*de Cárcer et al., 2011*). Interestingly, PLK5 is not involved in cell cycle regulation, however its overexpression arrests cells at the G_0/G_1 boundary (*de Cárcer et al., 2011*).

PLK4/SAK forms a more divergent subfamily due to the presence of a unique 3 PBD architecture (*Leung et al., 2002; Slevin et al., 2012*). PLK4 mRNA levels are low in early G_1, but rise in late G_1 and continue to rise during S, G_2, and M phases (*Winkles and Alberts, 2005*). PLK4 is an essential member, since PLK4-deficient mice die post gastrulation due to mitotic defects associated with aberrant mitotic spindle structures (*Hudson et al., 2001*). In turn, aberrant mitotic structures arise

because PLK4 is predominantly a regulator of centriole biogenesis, and thus, centrosomal duplication (*Habedanck et al., 2005; Kim et al., 2013*). PLK4-mediated centrosomal duplication occurs through a so-called cryptic PBD, derived from PB1 and PB2, which homodimerises and targets PLK4 to centrioles through interactions with either of the centrosomal scaffold proteins, centrosomal proteins 152 (Cep152) and 192 (Cep192) (*Slevin et al., 2012; Kim et al., 2013*). Interestingly, PB1 and PB2 regions are structurally distinct from one another, and their homodimerisation later regulates PLK4 stability by kinase domain-mediated transautophosphorylation of a phosphodegron motif that is recognised and ubiquitylated by $SCF^{\beta\text{-}TRCP}$ (*Guderian et al., 2010; Slevin et al., 2012*). In turn, $SCF^{\beta\text{-}TRCP}$-mediated ubiquitylation targets PLK4 for 26S proteasomal degradation, thereby restricting centrosomal duplication to once per cell cycle (*Guderian et al., 2010*). Moreover, the cryptic PBD adopts a canonical PBD fold, however, features several unique structural features that diverge from PLK1 PB1 (*Elia et al., 2003; Slevin et al., 2012*). These include an extended PB1 N-terminal region, a so-called 'stirrup' 16-amino acid insertion between β5 and β6, and a hairpin extension between β3 and β4 (*Slevin et al., 2012*). Additionally, the PLK4 PB2 diverges from PLK1 PB1 in the positioning of β2-β3 loops and an ordered loop between β5 and α1 that replaces the PLK1 PB1 β6 strand (*Slevin et al., 2012*). Instead, β6 is located C-terminal to α1 and occludes the phospho-binding site present in PLK1 PB1, suggesting that the PLK4-PBD does not function as a phospho-binding module (*Slevin et al., 2012*).

4.1.6 PLK1 signalling in the DDR

PLK1 was first identified in *S. cerevisiae* through genetic screens of temperature-sensitive cell division cycle (Cdc) mutants with delayed cell cycle progression (*Hartwell et al., 1973*) and a plethora of crucial functional roles in every cell-cycle stage have subsequently been extensively

elucidated (*Reviewed in Song, Liu, and Liu, 2012; Iliaki, Beyaert, and Afonina, 2021; Kalous and Aleshkina, 2023*). Of specific interest here is the involvement of Plk1 in DDR regulation that appears intimately associated with its cell-cycle related activities (Figure 4.4). For example, PLK1 phosphorylation of Rad51 recombinase at S14 primes casein kinase 2 (CK2)-dependent phosphorylation at T13, which stimulates an interaction between the NBS1-FHA domain and Rad51 (Figure 4.4; *Yata et al., 2012*). Consequently, S-phase PLK1 activity is indirectly involved in NBS1-mediated recruitment of Rad51 to DSB sites (*Yata et al., 2012*). In turn, the local concentration of Rad51 at DSBs increases, thereby driving Rad51-mediated nucleoprotein filament formation, which catalytically promotes HR activation through homology recognition and strand exchange (Figure 4.4; *Ogawa et al., 1993; Baumann, Benson, and West, 1996; Yata et al., 2012*). During G_2 DNA damage checkpoint recovery, PLK1 is crucial in regulating the MRE11 subunits of the MRN complex (Figure 4.4; *Li et al., 2017*). The PBD has previously been shown to interact with MRE11, although the mechanism is poorly defined (*Li et al., 2017*). This interaction enables kinase domain-mediated phosphorylation of MRE11 at S649, which subsequently primes CK2-dependent phosphorylation of MRE11 at S688 (Figure 4.4; *Li et al., 2017*). Both S649 and S688 phosphorylation sites are located in the second of two MRE11 C-terminal DNA-binding domains (*Usui et al., 1998; D'Amours and Jackson, 2001; Li et al., 2017*). In turn, phosphorylation of these two sites disrupts the formation of MRE11 and DNA complexes, thereby inhibiting the loading of the MRN complex at DSBs (Figure 4.4; *Li et al., 2017*).

Upon mitotic entry, PLK1 and CDK1 inhibit the 53BP1 and BRCA2 DDR axes of HR and NHEJ-mediated DDR pathways (Figure 4.4; *Esashi et al., 2005; van Vugt et al., 2010; Benada et al., 2015*). Mitotic CDK1 phosphorylates 53BP1 at S380, which primes PBD-dependent PLK1 recruitment (*van*

Vugt et al., 2010). Once recruited, PLK1 phosphorylates 53BP1 at S1618 within the ubiquitin-dependent recruitment (UDR) motif (Figure 4.4; *Fradet-Turcotte et al., 2013; Benada et al., 2015*). Phosphorylation of S1618 prevents 53BP1 recruitment to H2A-Ub chromatin sites, which serves to inhibit NHEJ-mediated DSB repair and subsequent fusion of mitotic telomeres (Figure 4.4; *Orthwein et al., 2014; Benada et al., 2015*). Addtionally, PLK1 phosphorylates the Chk2 FHA phospho-binding module to inactivate Chk2 kinase activity during mitosis (Figure 4.4; *van Vugt et al., 2010*). In parallel, CDK1 phosphorylates the BRCA2 C-terminal Rad51-interaction domain at S3291, which ablates interactions between BRCA2 and Rad51, and in turn, inactivates mitotic HR (Figure 4.4; *Esashi et al., 2005*). While unable to drive mitotic HR, BRCA2 appears to drive accurate amphitelic spindle-mediated chromosomal segregation in a PLK1-dependent manner (Figure 4.4; *Ehlén et al., 2020*). BRCA2 is phosphorylated in an N-terminal ST-rich region at S193 and T207 by PLK1 (*Ehlén et al., 2020*). In particular, T207 phosphorylation creates a PBD docking site that drives formation of a heterotetrameric complex with PLK-1 phosphorylated BUB1-related protein kinase (pBubR1) and serine/threonine protein phosphatase 2A (PP2A) (*Ehlén et al., 2020*). The pBRCA2:PLK1:pBUBR1:PP2A complex is crucial for proper mitotic spindle functionality, since phosphonull BRCA2-T207A expression results in chromosomal misalignment and missegregation, and subsequent aneuploidy (Figure 4.4; *Ehlén et al., 2020*).

Strikingly, PLK1 activity does not appear to inhibit all DDR pathways in mitosis, especially alternative end-joining in HR-deficient cells (Figure 4.4; *Gelot et al., 2023*). Recent evidence has emerged that demonstrates PLK1 phosphorylation of DNA polymerase theta (Polθ) at ten sites (S1289, S1321, S1436, S1444, S1482, S1486, S1488, S1493, S1551, S1555, S1563, T1584, S1594, S1628, S1635, S1639, and T1755) (Figure 4.4; *Gelot et al., 2023*). The phosphorylated S1482-S1493 cluster

interacts with TOPBP1 C-terminal BRCT repeats 7-8 domain. The function of this interaction is to facilitate mitotic alternative end-joining repair of DSBs, as disruption of the interaction leads to an increased number of mitotic γH2AX foci and G_1 micronuclei (Figure 4.4; *Gelot et al., 2023*).

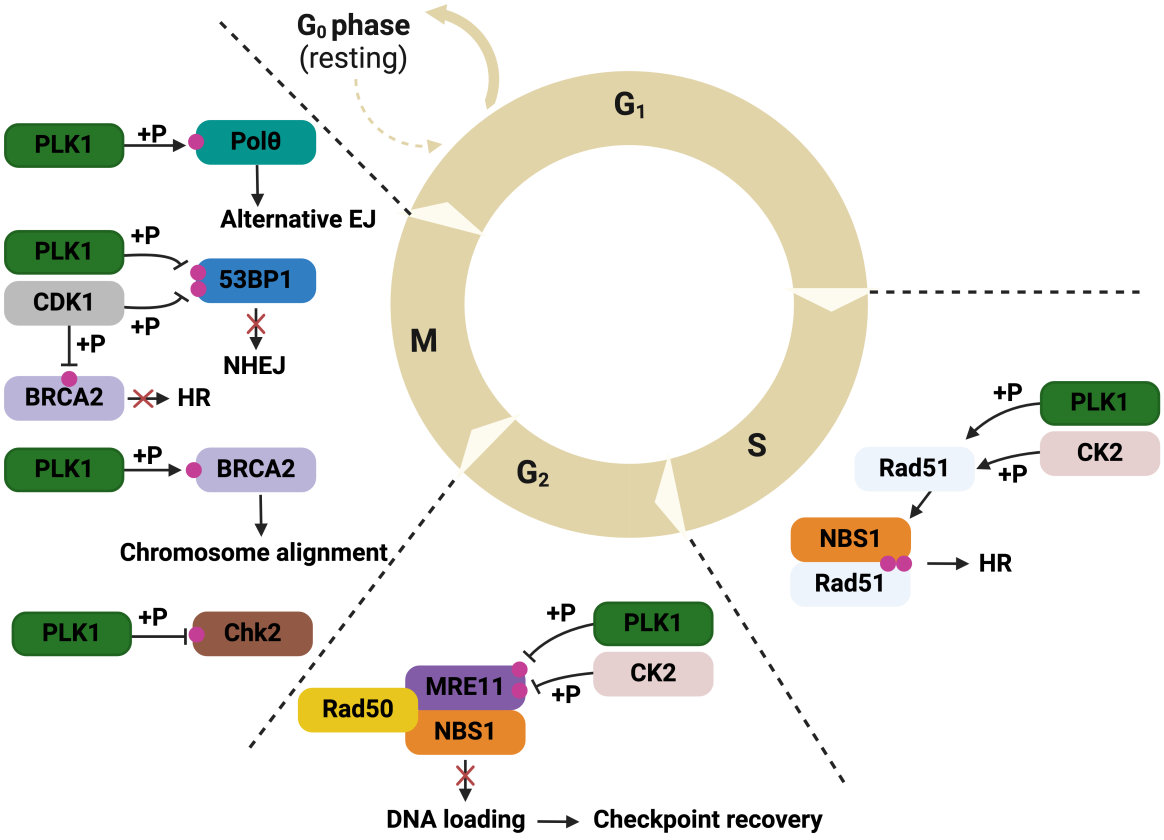


Figure 4.4 - PLK1-mediated phospho-regulation of DNA damage signalling throughout the cell cycle. In S-phase (S), PLK1 (dark green) catalyses phosphorylation (+P; magenta circles) of Rad51 (light blue), which primes CK2 (light)-mediated phosphorylation of Rad51. The latter event triggers NBS1 recruitment and subsequent Rad51-mediated nucleoprotein filament formation, required for homologous repair (HR). During G_2 checkpoint recovery, inhibitory (flat arrow) PLK1-mediated phosphorylation of MRE11 (purple) primes CK2-mediated phosphorylation of MRE11, which prevents (crossed arrow) loading of the MRE11-Rad50 (yellow)-NBS1 complex onto DNA. Mitotic PLK1 activity stimulates alternative end-joining (alternative EJ) through phosphorylation of Polθ (teal). CDK1 (grey)-mediated phosphorylation of 53BP1 (dark blue) primes PLK1 phosphorylation of 53BP1, which leads to inhibition of mitotic non-homologous end joining (NHEJ). CDK1 also inhibits HR through phosphorylation of BRCA2 (light purple), while PLK1 phosphorylation of BRCA2 stimulates formation of a pBRCA2:PLK1:pBUBR1:PP2A heterotetrameric complex (not shown), involved in proper chromosomal alignment. In addition, PLK1-mediated phosphorylation of Chk2 (brown) inhibits mitotic Chk2 activity. Figure created with Biorender.com.

Recent studies in *Drosophila* and human cells have hinted at a direct interaction between PLK1 and MRN complex members (*Xu et al., 2018; Landmann et al., 2020*). For example, *Drosophila* PLK1

(Polo) is recruited to DSB sites in an apparent kinase domain-dependent manner, however Polo was recruited independently of the mitotic phase in which DSBs were induced (*Landmann et al., 2020*). Both dMRE11 and dRad50 transiently associate with DSBs, prior to Polo recruitment, while hypomorphic dRad50 mutant cells led to depletion of MRE11 and Polo DSB recruitment (*Landmann et al., 2020*). Finally, a mitosis-specific MRN (mMRN) complex appears to recruit PLK1 via a so-called mitotic MRN associated protein (MMAP) bridge (*Xu et al., 2018*). MMAP/WDCP is a member of the mMRN complex, and a poorly characterised WD repeat and coiled-coil containing protein (*Yokoyama and Miller, 2015; Xu et al., 2018*). PLK1 interacts with MRE11 and MMAP and phosphorylates both respective C-termini at S688 and S689 and at S686 and S695, respectively. Consequently, the phosphorylation of MRE11 and MMAP stabilises the mMRN-PLK1 complex, and enhances its interactions with the microtubule depolymerase, KIF2A (*Xu et al., 2018*). Additionally, PLK1 phosphorylates KIF2A to activate its microtubule depolymerising activity (*Jang et al., 2009*). This activity is essential for the proper and dynamic capture of chromosomes during mitotic chromosomal alignment and segregation, which is driven by spindle catastrophe and rescue mechanisms (*Jang et al., 2009*).

4.2 Aims & Objectives

The Smerdon lab solved the first crystal structure of phosphopeptide-bound PBD, and defined the phosphorylation-dependent interaction consensus motif, which was crucial at the time to understand how PLK1 is targeted to centrosomes during normal mitotic progression (*Elia et al., 2003*). Our lab, therefore, in collaboration with the Hořejší lab further investigated the potential phosphorylation-dependent interaction between Rad50-T690 and PBD. As discussed previously, PLK1 contains a C-terminal PBD that possesses the following phosphorylation-dependent

interaction consensus sequence: [P/F]- [Φ/P]-[Φ]-[T/Q/H/M]-S-[**pS/pT**]-[P/X], whereby an invariant S residue is located in the pT-1 position, in contrast to the corresponding Rad50-pT690 -1Q residue (*Elia, Cantley, and Yaffe, 2003; Elia et al., 2003*). The original oriented peptide library screens that defined the PBD interaction consensus sequence focused on identifying phospho-binding modules that bind to pT-P motifs, and this core motif was fixed, while the pT -1 position was changed to only S and A residues (*Elia, Cantley, and Yaffe, 2003; Elia et al., 2003*). Consequently, the Rad50-pT690 peptide local sequence diverges from the PBD phosphorylation-dependent interaction consensus sequence suggesting that a distinct set of PBD target proteins remain to be identified.

Together these observations presented an exciting prospect for expanding our current understanding of PBD phosphorylation-dependent interactions both in terms of the apparently novel but uncharacterised specificity of PBD binding and the potential effects on Rad50 and MRN function arising from phosphodependent (and independent) interactions with RAD50 pT690 and potentially other sites within this large DNA-repair and signalling complex.

4.3 Results

4.3.1 PDB expression and purification

As a starting point, the region encompassing the proteolytically-defined PBD (residues 367-603) was cloned in-frame with an N-terminal GST-tag and succeeding 3C protease cleavage site (LEVLFQ↓GP) into the pGEX6P-1 bacterial expression vector. A phospho-binding deficient construct, termed Pincer, was also generated through mutation of 'Pincer' residues, H538 and K540, to A and M residues, respectively. The recombinant Pincer protein was included to serve as a negative control for binding assays, with the provision that further mutations would be made in the

PBD gene fragment should the synthetic Rad50 phosphopeptide bind phosphorylation-dependently to an unexpected PBD surface. Both proteins were purified by glutathione-Sepharose affinity chromatography, overnight 3C protease GST-tag cleavage, and gel filtration to homogeneity at >90% purity (Figure 4.5). The typical yields were 10mg/L culture.

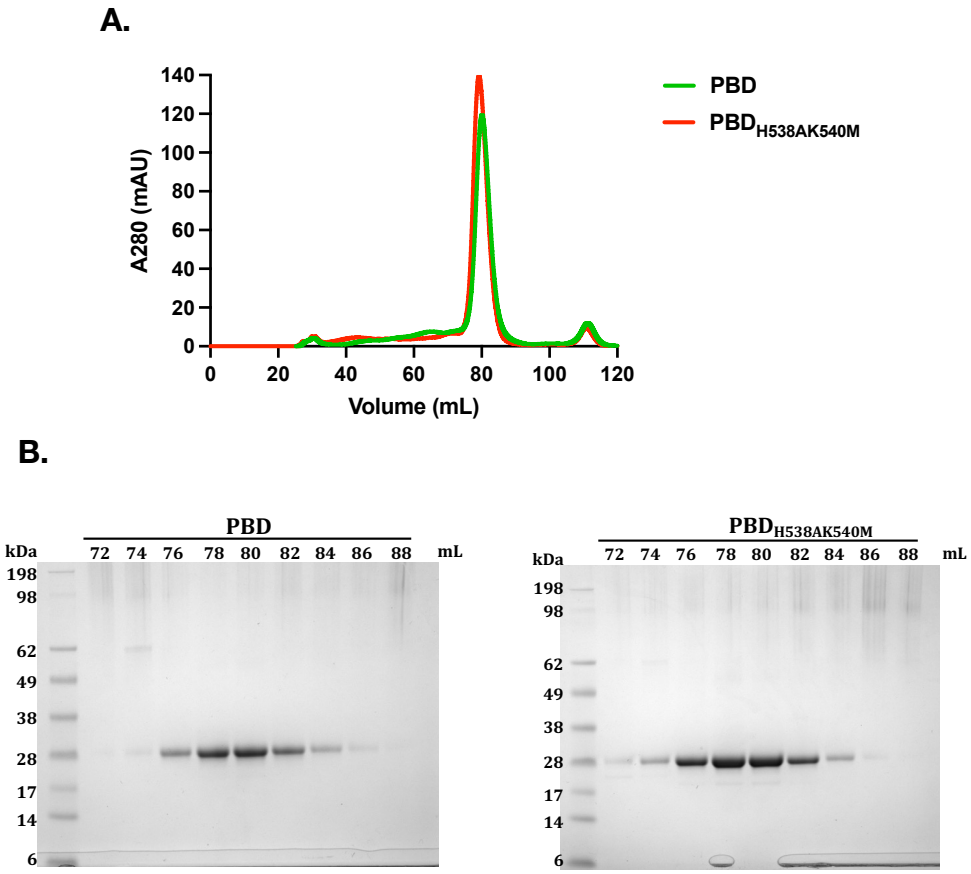


Figure 4.5 - Expression and purification of wild-type and phospho-binding deficient recombinant PLK1 PBD domains. (A) Size-exclusion chromatography (SEC) elution profiles of purified PBD (left panel) and Pincer (right panel) and domains. The overall elution volume (mL) and ultraviolet (UV) absorbance profile at 280nm (mAU) is indicated for wild-type PBD (green), $PBD_{H538AK540M}$ (Pincer; red) from a Superdex 200 16/60 column. **(B) SDS-PAGE analysis of SEC elution peaks for PBD (left panel) and Pincer (right panel) proteins.** The fraction elution volume is indicated at the top of each gel.

4.3.2 Peptide synthesis

In total, 3 Rad50-pT690 phosphopeptides and 1 Rad50-T690 peptide were synthesised by 2 separate companies during the course of this study. The first peptide, Ac-RVFQ**pT**EAELAYKK-Am, failed to bind to the recombinant PBD during initial MST-binding trial experiments, and mass spectrometry post-translational modification analyses indicated that the T690 residue was not phosphorylated. The second peptide was initially designed identically to the phosphopeptide from the published peptide-pulldown experiments: Biotin-eahx-CPVCQRVFQ**pT**EAELQEVISD, but it continuously failed synthesis quality controls. A year later, an N-terminally truncated variant of the biotinylated Rad50 phosphopeptide was successfully synthesised, Biotin-eahx-RVFQ**pT**EAELQEVISD. The latter phosphopeptide was not detectable by UV absorption-based quantification, and was insoluble in water and buffer, however pre-solubilisation of 1mg with 30μL 0.1M ammonium bicarbonate aided solubility in 50mM Hepes pH 8, 150mM NaCl for binding experiments. Due to an inability to accurately quantify the biotinylated phosphopeptide, a rough estimate of the peptide concentration (μM) was made based on the mass of the peptide within a given volume (mg/mL) divided by the observed molecular weight (2182.4Da). Finally, both non-phosphorylated and phosphorylated variants of the first peptide were synthesised, Ac-RVFQ**(p)T**EAELAYKK-Am. An AYKK C-terminal linker was added to aid solubility in buffer and UV absorption-based quantification using the Beer-Lambert law, with the observed molecular weight as 1661.4Da and a predicted extinction coefficient of $1490 M^{-1} cm^{-1}$. For simplicity, peptide 2 will be referred to as biotin-pT690 peptide, and the final two non-phosphorylated and phosphorylated peptides as T690-AYKK peptide and pT690-AYKK peptide, respectively.

4.3.3 Phosphopeptide Binding I – Microscale Thermophoresis

MST binding experiments were conducted by fluorescently labelling recombinant PBD and Pincer proteins, because the biotin-pT690 phosphopeptide contained only a single primary amine group that would react with our NHS-ester reactive dye. In the first instance, binding experiments were conducted by titrating the positive control PBDtide, Ac-MAGPMQS**pT**PLNGAYKK-Am, a well-established PBD interactor with an optimised binding motif. Moreover, all experiments were validated by quality control checks via pre- and post-binding fluorescence scans of all individual MST capillaries. Previous binding analyses by ITC have demonstrated that PBDtide binds to PBD in a Pincer motif-dependent manner with a K_D of 280±27 nM, and the crystal structure of PBD bound to PBDtide was solved and extensively characterised by the Smerdon lab. Our MST binding experiments indicate that our PBD fragment binds to PBDtide with a similar K_D of 291.7nM (95% CI: 188.9 - 447.9nM) (Figure 4.6), while the phospho-binding-deficient Pincer did not bind PBDtide. Interestingly, our transformed normalised fluorescence change plots (Figure 4.6B) indicate that PBD exhibits a PBDtide concentration-dependent decrease in thermophoretic behaviour. This observation does not appear to be a result of ligand-induced conformational changes, because previous structural analyses have demonstrated that PBD undergoes minimal conformational changes upon PBDtide binding. Instead, the reduced thermophoretic behaviour is more likely due to a smaller hydrodynamic radius, which results from the displacement of water molecules at the PBDtide interface and the formation of a lattice of water-mediated hydrogen bonds between PBD and PBDtide.

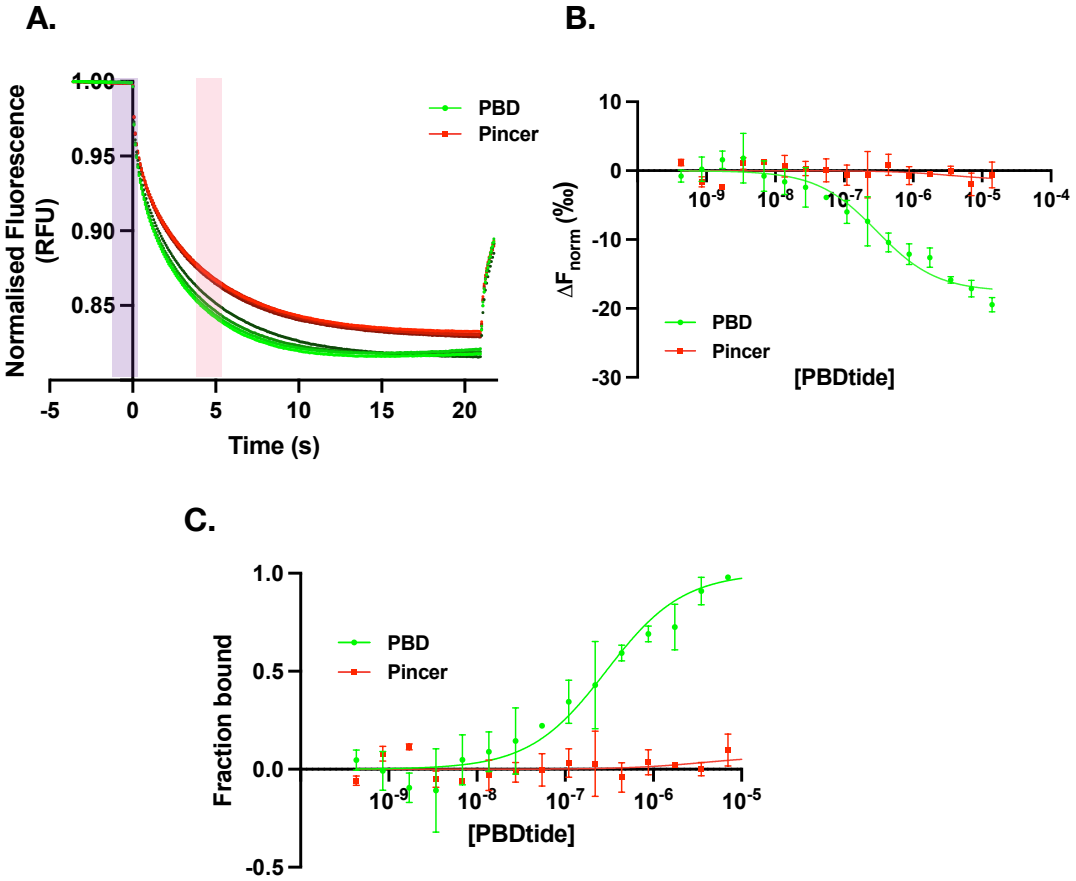


Figure 4.6 - MST binding analysis indicates that the recombinant PBD protein functions as a phospho-binding module. (A) Raw fluorescence MST traces of PBDtide binding to PBD and Pincer proteins. Each coloured line represents the average normalised fluorescence values quantified from 4 capillaries for PBD (green) and Pincer (red) binding experiments. A colour gradient scheme is used, whereby brighter colours denote capillaries containing maximal PBDtide phosphopeptide concentration and darker colours represent decreasing in PBDtide concentration. The cold (purple) and hot (pink) region of the experiment used to calculate ΔF_{norm} and fraction bound values is indicated. **(B) Representative change in normalised fluorescence (ΔF_{norm}) plots indicates that PBD exhibits negative thermophoretic behaviour upon binding to PBDtide, whereas Pincer remains unchanged.** The concentration of PBDtide phosphopeptide (x-axis) is indicated on a $\log_{10}$ scale. **(C) A representative fraction bound plot is used to calculate the affinity (K_D) of the interaction between PBD and PBDtide phosphopeptide.** Individual data points represent the mean values and error bars are the standard deviation values from three independent experiments.

Our results, therefore, clearly suggested that our recombinant PBD/Pincer proteins were functional, and MST binding analysis was able to accurately capture dissociation constants that are consistent with previous PBD interactions deteremined by ITC and SPR. Following this validation step, we

probed the interaction between PBD/Pincer and biotin-pT690 (Figure 4.7). The biotin-pT690 phosphopeptide interacted with PBD with a K_D of 2.58μM (K_D 95% CI: 1.45 – 4.58μM), whereas no interaction was detected between Pincer and biotin-pT690. Clearly, our MST binding curves were not fully saturated (Figure 4.7C), therefore more accurate K_D values could be determined with a higher biotin-pT690 phosphopeptide concentration, but this was not investigated further. Importantly, our K_D between biotin-pT690 and PBD were ~20-fold lower than was previously observed by preliminary SPR binding experiments. This difference could be a result of several factors, including the 5 amino acid N-terminal extension present within the original biotin-pT690 phosphopeptide, a buffer mismatch associated with our ammonium bicarbonate peptide solubilisation strategy, or the contrast between peptide immobilisation in SPR experiments and fluorescence and solution-based measurements in MST experiments.

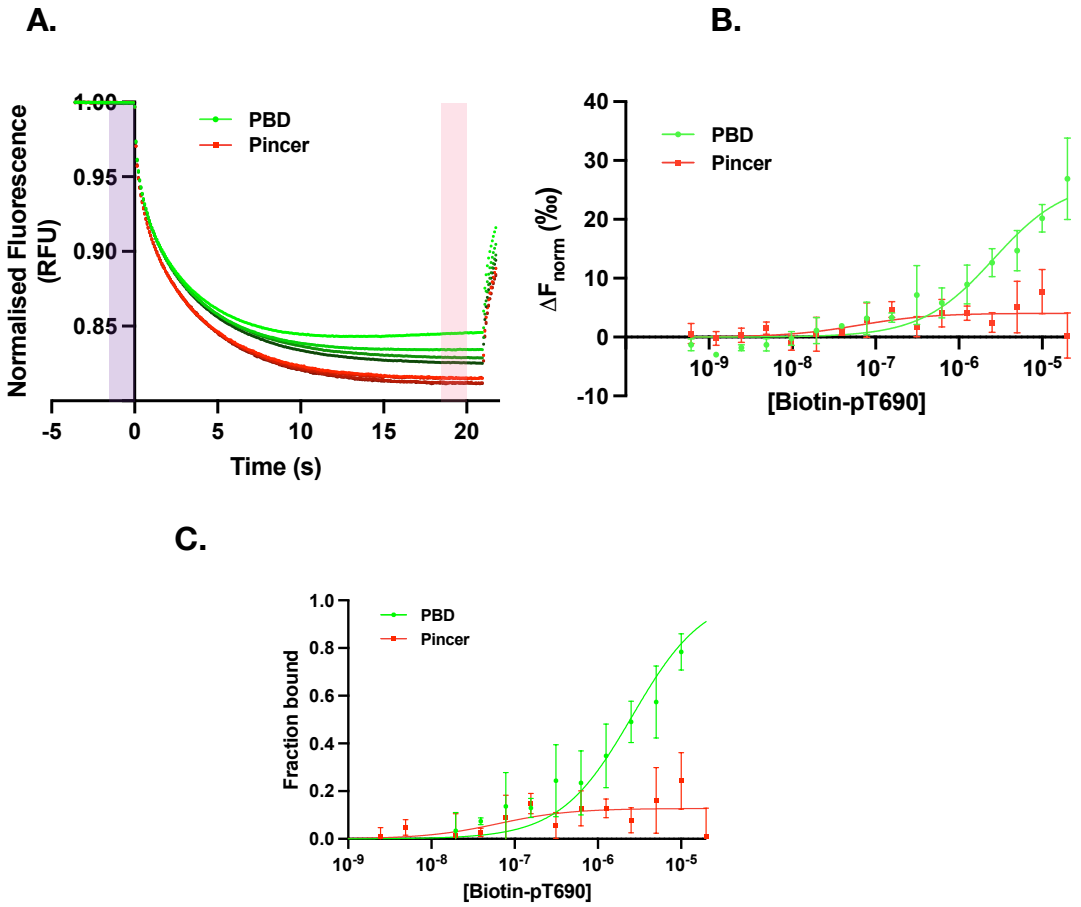


Figure 4.7 - MST binding analysis suggests that the recombinant PBD protein binds to the biotin-Rad50-pT690 phosphopeptide. (A) Raw fluorescence MST traces of PBDtide binding to PBD and Pincer proteins. (B) Representative change in normalised fluorescence (ΔF_{norm}) plots suggests that PBD exhibits positive thermophoretic behaviour upon binding to PBDtide, whereas Pincer remains unchanged. (C) A representative fraction bound plot is used to calculate the affinity (K_D) of the proposed interaction between PBD and the biotin-Rad50-pT690 phosphopeptide. Individual data points represent the mean values and error bars are the standard deviation values from three independent experiments.

4.3.4 Structural analysis of the PBD-Rad150 pT690 complex

Based on the primary binding results, it was determined that the interaction between PBD and biotin-pT690 was dependent on an intact H538K540 Pincer motif, and the interaction was sufficiently high affinity for structural analysis of the complex by X-ray crystallography. We decided,

therefore, to pursue the possibility of solving a high-resolution structure of the PBD:biotin-pT690 complex to determine the molecular basis of the interaction.

Initial PBD:biotin-pT690 phosphopeptide co-crystallisation screens were conducted by vapour diffusion with the biotin-pT690 phosphopeptide in 2-molar excess of PBD and ~280-fold higher than our experimentally determined K_D. Strikingly, ~20% of screened conditions (84 hits/384 conditions) resulted in sufficient crystal nucleation and growth after 24 hours at 18°C with at least 3 different morphologies. This implied that either our recombinant PBD fragment readily crystallised in the absence of biotin-pT690 phosphopeptide or vice versa, or binding of the phospho-ligand stabilised the PBD structure. To address all these possibilities, PBD only and biotin-pT690 phosphopeptide screens were conducted with all of the previously tested conditions. From these analyses, we determined that ~90% (75/84) of hits induced crystal formation in PBD alone screens, while the remaining 9 hits were specific to co-crystallisation with the biotin-pT690 phosphopeptide. X-ray diffraction datasets were collected at Diamond Light Source, Oxford for 3 hits without phosphopeptide soaking, while a further dataset was collected at European Synchrotron Radiation Facility (ESRF), Grenoble after crystal soaking with 10μM biotin-pT690. Xtriage crystal analysis indicated that 75% of datasets exhibited a >95% data completeness with no outlier warnings, thereby demonstrating the high quality of our datasets. In total, the x-ray crystal structure of the C-terminal PLK1 fragment corresponding to PBD residues 373-594 was solved 4 times in 2 different space groups (P1 $2_1$1 and P2 2_12_1) (Figure 4.8A and Tables 4.1-4.2) using molecular replacement with the previously deposited optimal phosphopeptide-bound PBD structure (PDB accession code: 1UMW) as a search model (*Elia et al., 2003*). In all cases, maps phased with initial molecular replacement models failed to show clear residual density corresponding to the biotin-pT690

phosphopeptide at any PBD surface (Figures 4.8B-C). Nonetheless, near complete models of the PBD structure (residues 373-498, 506-594) were manually built and refined at 2.21-2.83Å. All of our crystal structures adopted the classical PBD-fold containing a central 12-stranded β-sandwich bordered by 3 α-helical segments, derived from 2 $\beta_6\alpha$ Polo-box regions (PB1 and PB2). Additionally, we observed the Polo-cap (Pc), comprised of α_1 followed by a loop, 3_{10}-helix, and a short linker loop that connects to β_1 of PB1. The Pc encircles PB2 and tethers it to PB1 through the antiparallel arrangement of Pc α_1 and PB2 α_3 and structural packing of Pc α_1 and the 3_{10}-helix against the short PB2 α helical segment. The L2 linker connecting PB1 and PB2 was mostly absent from our structures due to a lack of electron density, which indicated that our structure was in the unbound state (*Cheng et al., 2003*).

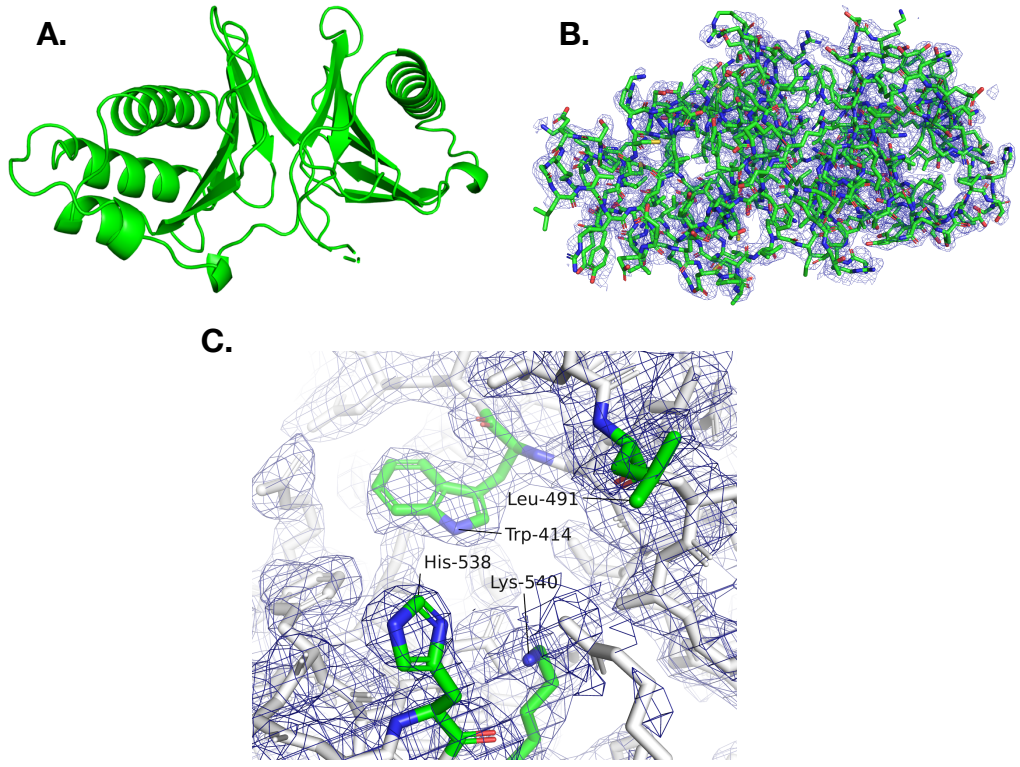


Figure 4.8 - The 2.58 Å crystal structure of unbound PBD. (A) The PBD secondary structure is represented in the cartoon format. **(B)** PBD atoms are represented in the stick format and coloured by element with the supporting electron density at a 1σ contour level coloured in blue. **(C)** PBD residues involved in phospholigand binding are coloured in green with the supporting electron density in blue.

PLK1 (373-594)		
Data Collection ESRF MASSIF-3		
Space group	P2 2_12_1	
Cell dimensions	a, b, c 32.8 Å 75.3 Å 98.3 Å α, β, γ 90° 90° 90°	
Wavelength	0.61992 Å	
Resolution limit	2.58 Å	
	Overall (30.22-2.58 Å)	**Outer Shell (2.74-2.58 Å)**
R_{merge}	0.063	0.364
R_{meas}	0.090	0.514
R_{pim}	0.063	0.364
Total n° of observations	15,902	1,258
Total n° of unique observations	8,085	687
<I/σ>	12.9	5.5
Completeness	98.4 %	87.3 %
Redundancy	2.0	1.8
Refinement		
Resolution limits	30.22-2.58 Å (2.74-2.58 Å)	
N° of reflections	7,818 (1141)	
R-factor (R_{work})	0.28 (0.42)	
R-factor (R_{free})	0.30 (0.42)	
Ramachandran favoured	94.79 %	
Ramachandran outliers	0.95 %	
Rotamer outliers	1.54 %	
Clashscore	7.15	
Average B-factor	38.19 $Å^2$	
Model		
Protein	1 X PLK1 (373-594)	
	373-498, 506-594 aa	
	26 water molecules	
Total n° of atoms	1,778	
R.M.S deviations		
Bond lengths	0.003 Å	
Bond angles	0.65°	

Table 4.1 - Summary of crystallographic statistics for the unbound PBD structure in the P2 2_1 2_1 space group.

PLK1 (373-594)		
Data Collection Diamond I24		
Space group	P1 $2_1$1	
Cell dimensions	a, b, c 33.2 Å 92.2 Å 36.3 Å α, β, γ 90° 100° 90°	
Wavelength	0.99001 Å	
Resolution limit	2.82 Å	
	Overall (28.26-2.82 Å)	**Outer Shell (2.87-2.82 Å)**
R_{merge}	0.195	0.959
R_{meas}	0.214	1.043
R_{pim}	0.085	0.406
Total n° of observations	33,637	1,551
Total n° of unique observations	5,205	240
<I/σ>	7.6	2.1
Completeness	99.8 %	96.8 %
Redundancy	6.5	6.5
Refinement		
Resolution limits	28.26-2.82 Å (3.11-2.83 Å)	
N° of reflections	5,187 (1,282)	
R-factor (R_{work})	0.33 (0.38)	
R-factor (R_{free})	0.35 (0.40)	
Ramachandran favoured	94.79 %	
Ramachandran outliers	0.95 %	
Rotamer outliers	1.54 %	
Clashscore	7.43	
Average B-factor	56.62 $Å^2$	
Model		
Protein	1 X PLK1 (373-594)	
	373-498, 506-594 aa	
	0 water molecules	
Total n° of atoms	1,752	
R.M.S deviations		
Bond lengths	0.003 Å	
Bond angles	0.65°	

Table 4.2 - Summary of crystallographic statistics for the unbound PBD structure in the P1 2_1 1 space group.

4.3.5 Interaction modelling

As an alternative approach, we implemented AlphaFold Multimer and AlphaFold 3 prediction models because of the previous success predicting the Peli 1_{FHA}:$MDC1_{pT765}$ complex (*Jumper et al., 2021; Mirdita et al., 2022; Abramson et al., 2024*). In the first instance, the primary sequence of PLK1 corresponding to the PBD region was inputted with the 9-mer Rad50-T690 peptide sequence in either its non-phosphorylated form or with a phosphomimetic T690E residue. Again, the T690 peptide model output map was manually phosphorylated by addition of the γ-phosphate moiety to the T690 position using COOT software. In AF3, the input T690 peptide was phosphorylated at the 690 position, prior to model prediction. For ease, these models will be described as AF-M-T690, AF-M-E690, and AF3-pT690, respectively. As described earlier, all 3 models were then ranked based on intrinsic model accuracy estimates, including pLDDT, PAE, pTM, and ipTM (Figure 4.9). Despite accurately predicting the overall PBD fold accurately, the top-ranking AF-M-T690 model exhibited an ipTM score of 0.58. In general, ipTM scores below 0.6 are indicative of a failure to accurately predict the relative positions of PBD and the T690 peptide, and so no further analysis of the AF-M-T690 model was conducted. Both AF-M-E690 and AF3-pT690 models formed a consensus interface prediction that was dependent on hydrogen bonding contacts between the E/pT residue and the Pincer motif (Figure 4.10). The AF-M-E690 model predicted only hydrogen bonding and salt bridge contacts between the E690 residue carboxylate side chain with the Pincer motif K540 residue ε-amine side chain, and R557 guanidinum side chain group. Moreover, a further 9 hydrogen bonds were detected between pT -4R and PBD-E488, pT -1 and PBD-L491, pT +1E and pT +2A and PBD-W414, pT +2A and +4L and PBD-D416, and 3 hydrogen bonds between pT +3E and PBD-R516. The majority of these peptide-contacting residues are shared in the crystal structure of PBD with its

optimal phosphopeptide, but the corresponding residue in modelled peptides diverge significantly (*Elia et al., 2003*). For example, W414 main chain atoms are involved in hydrogen bonding with the invariant pT -1S side chain hydroxyl group. The fundamental importance of W414 hydrogen bond contacts with -1S hydroxyl groups is underscored by the fact that mutation to F ablates PBD-mediated centrosomal localisation of PLK1f. In AF-M-E690, an alternative W414 hydrogen-bonding network is predicted in the absence of a pT -1S residue (Figures 4.10 and 4.11). W414 main chain carboxyl atoms hydrogen bond with pT +2A α-amine atoms, while a hydrogen bond is predicted between the pT +1E carboxyl side chain and the W414 phenyl side chain N atom (Figures 4.10 and 4.11). In contrast, AF3-pT690 positioned the pT690 γ-phosphate group more similar to previously solved phosphopeptide-bound PBD crystal structures, wherein the oxygen atoms are poised for hydrogen-bonding contacts with both Pincer H538 and K540 residues, however a further hydrogen bond was predicted between the pT main chain carboxyl group and PBD-L491 α-amine group (Figures 4.10 and 4.11). The AF3-pT690 model predicted only one further hydrogen bond between the pT +2A residue α-amine group and the PBD-489 main chain carboxyl group (Figures 4.10 and 4.11). Strangely, AF-M-E690 and AF3-pT690 model predictions positioned the respective peptides in distinct configurations: AF-M-E690 positioned the phosphomimetic peptide in a C-to-N configuration, while AF3-pT690 positioned the phosphopeptide in an N-to-C configuration (Figure 4.11). From these analyses, the AF-M-E690 prediction model seemed more likely, because of the shared phosphorylation-dependent Pincer motif, PBD-414, and hydrogen-bonding contributions between prediction models and established crystal structure could adequately explain the binding response observed in MST experiments. Additionally, the alternative C-to-N peptide configuration provided a hypothetical basis for how the Rad50-pT690 phosphopeptide was not previously

identified from the PBD interaction consensus sequence derived from previous oriented peptide library screens.

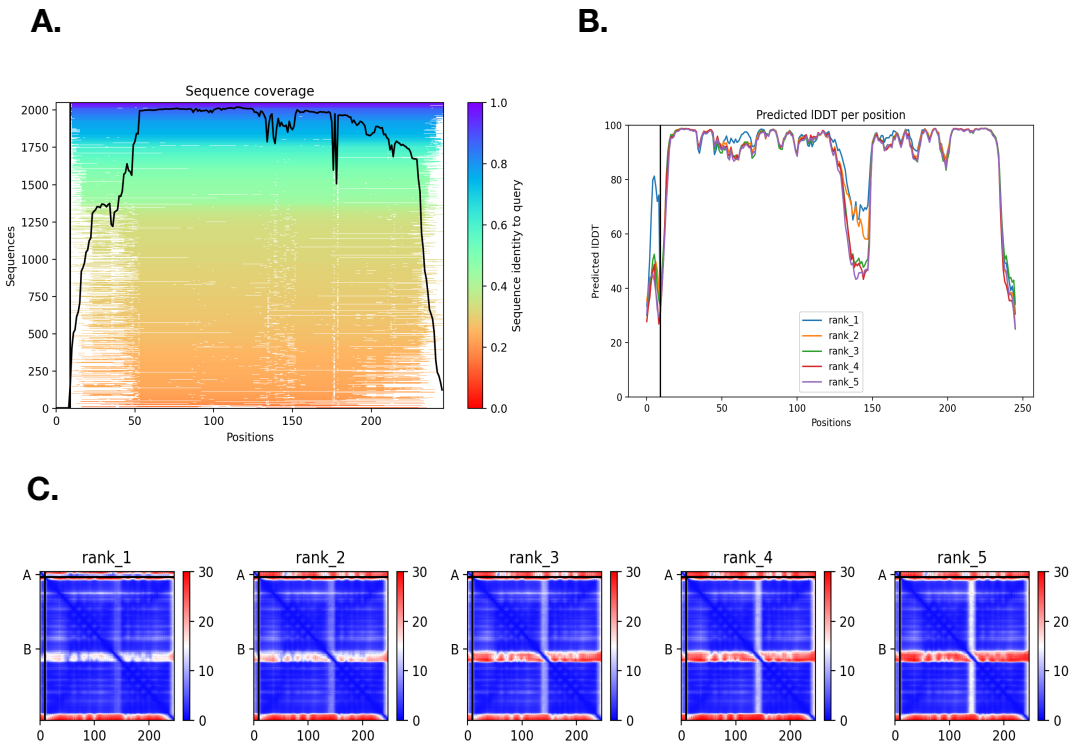


Figure 4.9 - AlphaFold 2 Multimer (AF-M) metric outputs for the PBD:Rad50$_{E690}$ phosphopeptide complex structural models. (A) The AlphaFold 2 multiple sequence alignment identifies ~1750-2000 homologous primary amino acid sequences that share ~80-95% sequence identity to the PBD query and 0 homologous Rad50$_{E690}$ 9-mer peptide sequences. The left-hand side x-axis indicates each amino acid position within the the 9-mer Rad50$_{E690}$ peptide query sequence, while the right-hand segment indicates the sequence within the PBD. The y-axis indicates the number of homologous sequences identified. The graph is coloured according to the sequence identity to both input sequences (0-100%). The number of homologous sequences identified, and the level of sequence identity per amino acid residue is depicted by the solid black line. **(B) The AlphaFold 2 IDDT metric identifies a single model with a high level of Rad50$_{E690}$ local residue confidence.** The predicted local distance difference test (pLDDT) is a per-residue confidence measurement of the expected distance deviation of a predicted model versus a true experimental structure expressed on a scale of 0 to 100. Each AlphaFold 2 output model is indicated by a coloured line. **(C) The AlphaFold 2 PAE metric identifies a single model with a high confidence level of oriented interacting residues.** The predicted aligned error (PAE) for a pair of residues is given on a scale of 0-30Å is depicted using a blue-to-red gradient colour scheme. For each individual ranked model, the PAE per residue for intra-Rad50$_{E690}$ paired residues is depicted by the top left square, intra-PBD paired peptide residues on the main square, and inter- Rad50$_{E690}$:PBD and PBD: Rad50$_{E690}$ paired residues are indicated on the left and top off diagonal rectangular plots, respectively.

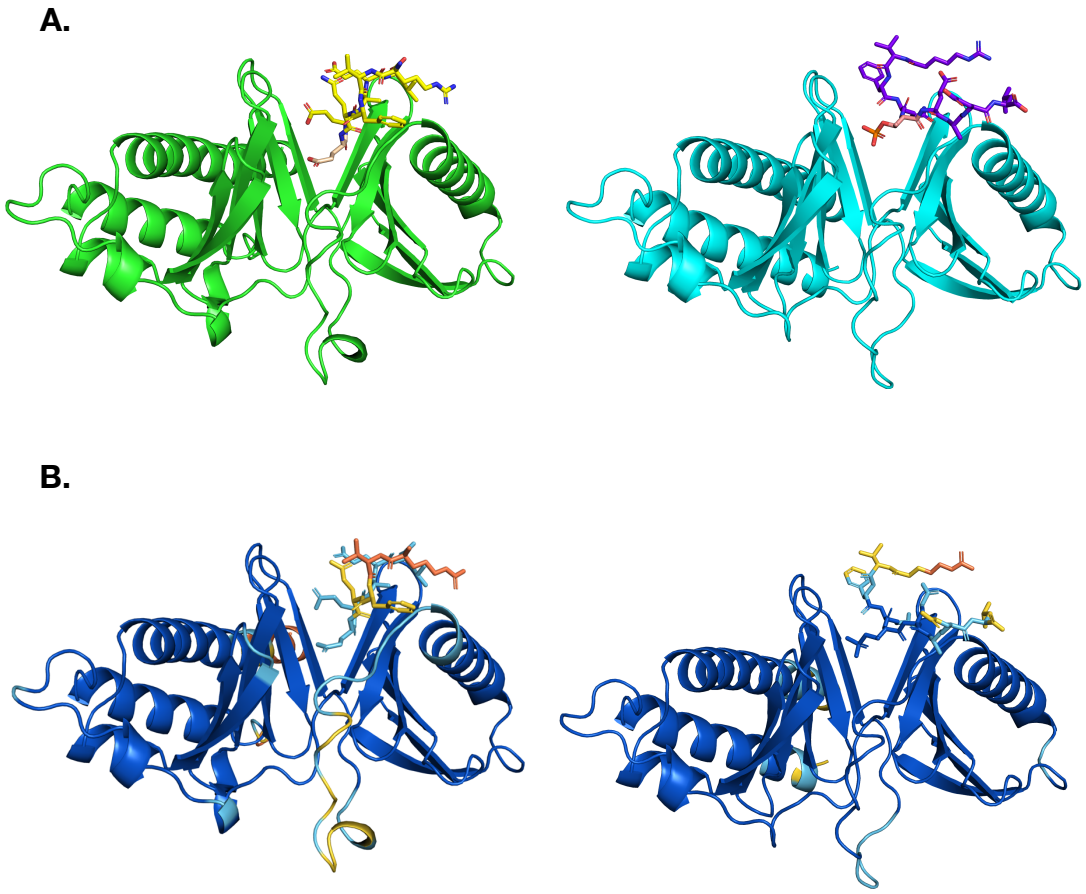


Figure 4.10 - AlphaFold 2 Multimer (AF-M) and 3 (AF3) models predict an interaction between PBD and the Rad50-pT/E690 peptide. (A) The AF-M PBD:Rad50$_{E690}$ phosphomimetic peptide (left panel) and AF3 predicted PBD: Rad50$_{E690}$ phosphopeptide (right panel) complex models. The AF-M predicted model is depicted with PBD in green and Rad50$_{E690}$ carbon backbone atoms are all coloured in yellow, except the E690 residue, which is coloured in wheat. The AF3 predicted PBD domain is coloured in cyan, and the Rad50$_{E690}$carbon backbone atoms are coloured in magenta, except the pThr, which is coloured in salmon. **(B) Both AF models are coloured according to per residue local confidence (pLDDT) score.** Dark blue indicates very high confidence (pLDDT >90), light blue indicates high confidence (90 > pLDDT > 70), yellow indicates low confidence (70 > pLDDT > 50), and orange indicates very low confidence (pLDDT < 50).

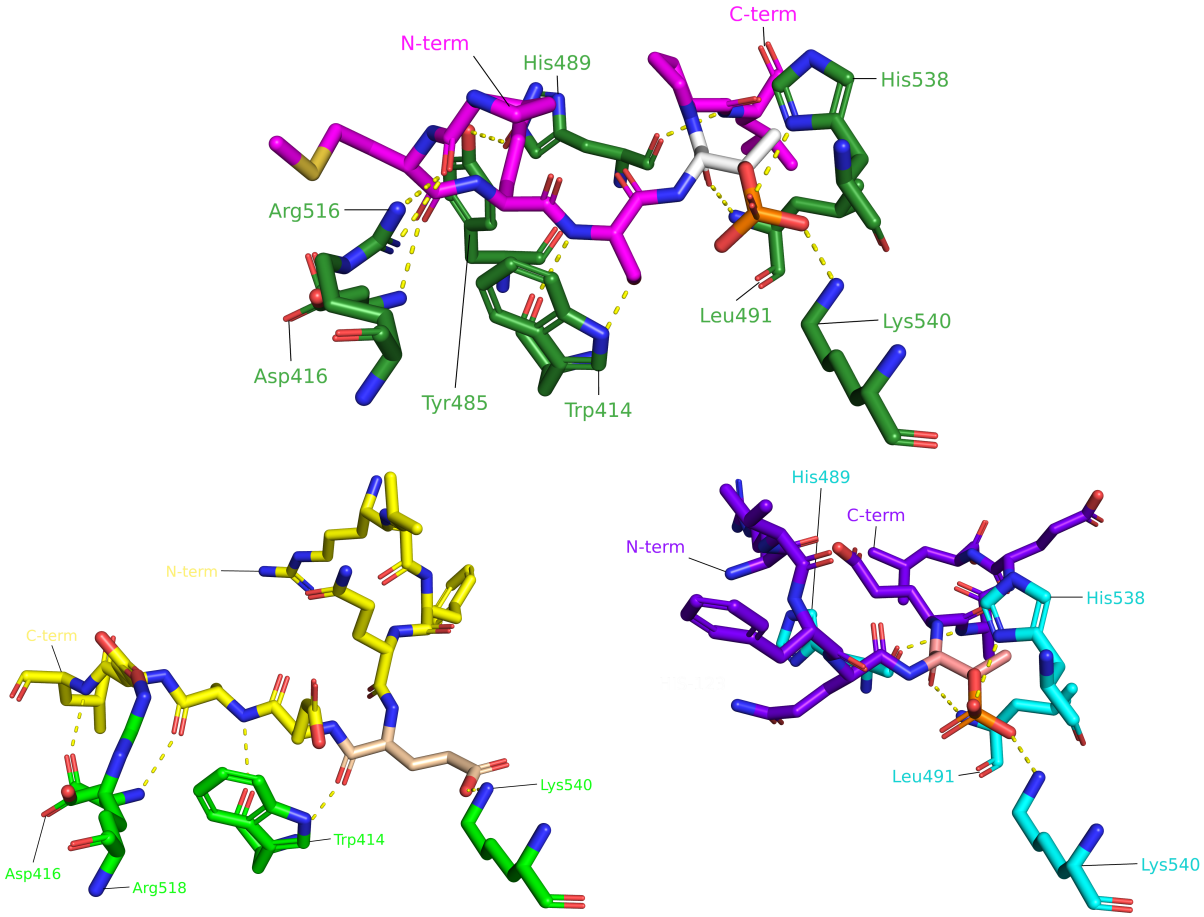


Figure 4.11 – Structural alignment of AlphaFold 2 Multimer (AF-M) and 3 (AF3) models to the crystal structure of PBD bound to PBDtide indicate alternate predicted Rad50 peptide binding configurations and differential PBD binding sites. (Top panel) PBD residues (forest) involved in PBDtide (magenta) binding from the crystal structure of PBD (accession code: 1UMW; Elia et al., 2003) are indicated in the stick format, with the pThr residue highlighted (white). Dashed lines highlight hydrogen-bonding contacts between PBD and PBDtide. (Bottom panels) Predicted PBD (AF-M in green; left panel, AF3 in cyan; right panel) residues involved in binding to the Rad50-E690 (AF-M; yellow) or Rad50-pThr690 (AF3; purple) phosphopeptide are indicated with the E690 (wheat) and pThr690 (salmon) residues highlighted. Dashed lines indicate predicted hydrogen-bonding contacts between PBD proteins and Rad50T690E (yellow) or Rad50pT690 (purple) peptides. Figure created using the PyMOL Molecular Graphics System v3.1.1 (Schrödinger, LLC).

4.3.6 Phosphopeptide Binding II – Isothermal Titration Calorimetry

Since these observations were based on prediction models and the MST data, we decided to pursue thermodynamic characterisation of the interaction using ITC. Again, we included the PBDtide

phosphopeptide as a positive control, and purified Pincer as a negative control for PBD binding experiments to further demonstrate that our PBD fragment was functional. PBDtide binding to PBD was direct and enthalpically driven, and our K_D of 286nM ±16.3nM closely resembles our MST binding data, and the binding affinity determined previously by ITC (Figure 4.12). Importantly, our raw thermograms exhibited PBD saturation of the PBDtide at early injections, therefore the parameters determined from these analyses could be improved by repeating with a smaller molar ratio of PBDtide titrant. Recombinant Pincer failed to elict any heat-exchange with PBDtide, thus our recombinant PBD protein was a functional phospho-binding module that interacted with substrates by the Pincer motif (data not shown). Unfortunately, we were consistently unable to detect any interaction between PBD and the biotin-pT690 phosphopeptide due to buffer mismatch (Figure 4.12A, right panel). This was likely due to the ammonium bicarbonate phosphopeptide solubilisation strategy, which was not mitigated by removal of small molecule contaminants by extensive dialysis using 1kDa molecular weight cutoff membranes. Based on these issues, we repeated ITC binding experiments with the aforementioned pT690-AYKK phosphopeptide, which was freely soluble in PBD binding buffer and quantifiable by UV-absorption measurements. These experiments also failed to detect any heat-exchange between PBD and pT690-AYKK (Figure 4.12A, bottom panel), indicative of a lack of interaction. The ITC analyses, therefore, implied that our previous MST analysis of biotin-pT690 and PBD were potentially a result of buffer mismatch. However, it remained possible that the C-terminal extension in the pT690-AYKK phosphopeptide disrupted interaction with PBD, or the 5-mer N-terminal Rad50 extension or peptide immobilisation was required.

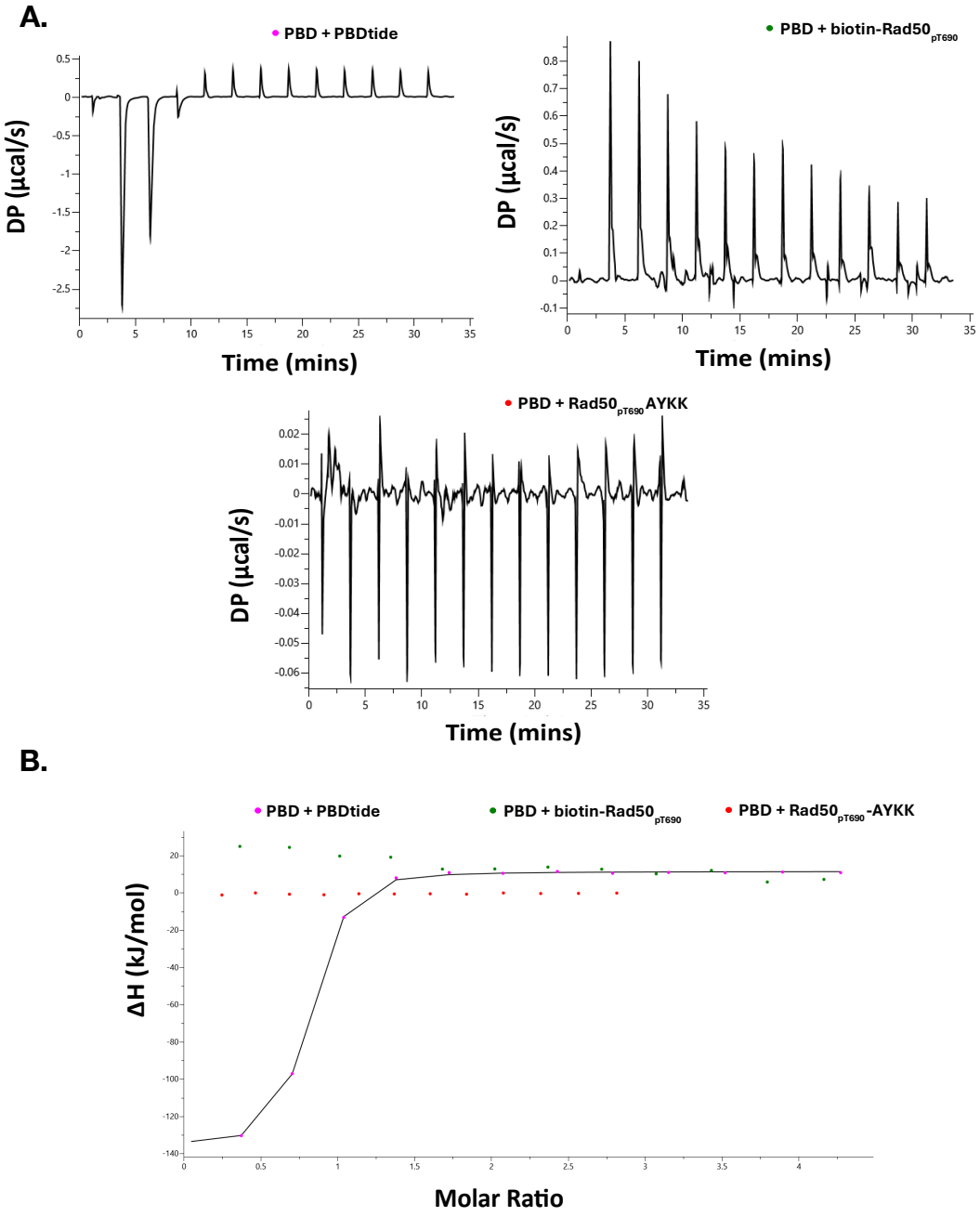


Figure 4.12 - ITC binding experiments demonstrate that the PBD directly binds to the PBDtide control phosphopeptide (pink), but no detectable binding was detected with biotin-Rad50$_{pT690}$ (green) and Rad50$_{pT690}$-AYKK (red) phosphopeptides. (A) Representative thermogram plots for PBDtide (left panel), biotin-Rad50$_{pT690}$, and Rad50$_{pT690}$-AYKK binding experiments. **(B)** A representative heat curve/binding isotherm of ITC PBD phosphopeptide binding experiments demonstrates that only the PBDtide phosphopeptide binds to the PBD. Data is representative of two independent experiments.

4.3.7 pT690 RAD50 Production by Amber Codon Suppression

Since we were unable to further characterise the interaction between Rad50-pT690 peptides and recombinant PBD, we sought to investigate the interaction at the protein-protein level. To determine the specific dependency of the interaction on T690 phosphorylation, we were presented with the technical challenge of expression and purification of recombinant ZCC protein that was phosphorylated only at the T690 position. The most physiologically relevant approach to tackle this issue was implementation of a recently published amber-suppression based pT OTS to site-specifically incorporate a pT residue at the T690 position (*Zhang et al., 2017; Moen et al., 2022*). Here, we provide a brief description of the theory that underpins pS and pT OTSs (*Park et al., 2011; Rogerson et al., 2015; Zhang et al., 2017; Moen et al., 2022; Mohler et al., 2023*).

pS/pT-OTSs are fundamentally similar to the mechanisms that underlie polypeptide chain growth during mRNA translation (*Noren et al., 1989*) (Figure 4.13). The codon that encodes the amino acid of interest (e.g. Rad50-T690) is mutated to a UAG amber stop codon within an expression vector that is compatible with the orthogonal tRNA:tRNA synthetase pair expression vector, which contains the L-threonine kinase gene *PduX* in the pT-OTS expression (*Noren et al., 1989; Zhang et al., 2017*). The orthogonal phosphoseryl-tRNA and phosphothreonyl-tRNA synthetases (pSerRS and pThrRS) from *Methanococcus maripaludis* catalyse a two-step ATP-dependent reaction: initial formation of a phosphoserine/threonine-adenylate (pS/pT-AMP), and subsequent esterification of phosphoserine/threonine to the 3’ end of an engineered *Methanococcus jannaschii* tRNA (pSer/pThr-$tRNA_{CUA}$) that decodes the UAG codon (*Rogerson et al., 2015; Zhang et al., 2017; Gomez and Ibba, 2020*). Consequently, pSer-$tRNA_{CUA}^{pSer}$ or pThr-$tRNA_{CUA}^{pThr}$ is generated. Importantly, the source of O-phospho-L-serine is derived from exogenous supplementation of the media, while O-

phospho-L-threonine is obtained through overexpression of the *Salmonella enterica* L-threonine kinase, PduX, which phosphorylates some of the intracellular L-threonine pool (*Rogerson et al., 2015; Zhang et al., 2017*). An engineered tRNA-binding elongation factor (EF-pSer/pThr) binds and efficiently delivers pSer-$tRNA_{CUA}^{pSer}$ or pThr-$tRNA_{CUA}^{pThr}$ to the A site of the large ribosomal subunit, thereby enabling pSer/pThr addition to the growing polypeptide chain (*Rogerson et al., 2015; Moen et al., 2022*).

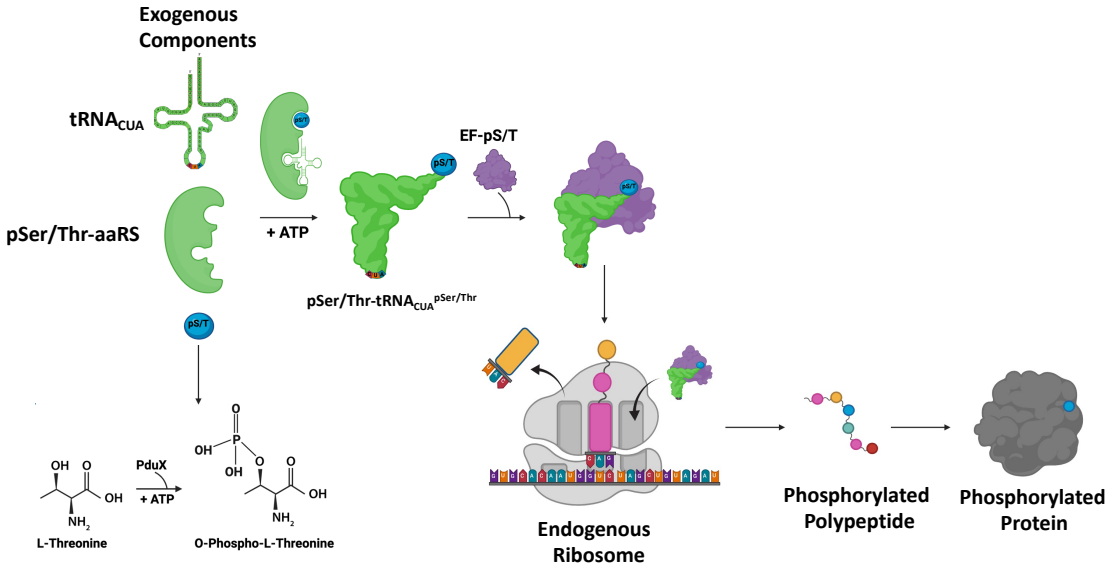


Figure 4.13 – pSer/Thr orthogonal translation systems (OTSs). Exogenous components of the pSer/Thr OTSs include an expression plasmid containing a pSer/Thr cognate $tRNA_{CUA}$:aminoacyl tRNA synthetase (aaRS), elongation factor (EF), and the *S. enterica* L-threonine kinase, PduX. O-phospho-L-serine is supplemented to the bacterial liquid growth medium. A separate bacterial overexpression plasmid contains the gene encoding the phosphoprotein of interest with an amber (TAG) stop codon that specifies the site of pSer/Thr incorporation. Upon ATP-dependent activation in cells, the complex is delivered to endogenous *E. coli* ribosomal aminoacyl A sites where the mRNA of a transcribed amber stop codon is decoded. Ribosomal peptidyl P site catalyse peptidyl transferase-mediated peptide bond formation between the pSer/Thr amino and the growing polypeptide chain, which is later folded to form a phosphorylated protein. Figure created using Biorender.com.

To our knowledge, the only utilisation of the pT OTS has focused on site-specific incorporation of pT into six recombinant proteins, namely GFP, Ub, cyclin-dependent kinase 2, a fusion reporter protein, consisting of an elastin-like polypeptide and GFP (MS-READ reporter), as well as a Checkpoint kinase 2 (Chk2) fragment encompassing the kinase domain region (Chk2-KD) (*Zhang et al., 2017; Moen et al., 2022*). Incorporation of pT into Ub at residue T12 proved successful in generating sufficient yields of homogeneous phosphoprotein for x-ray crystallography structure determination, but the small size and high stability of Ub make it an optimal model protein for pT OTS studies (*Zhang et al., 2017*). The successful incorporation of pT at the T160 residue in the CDK2 activation loop enabled the characterization of CDK2-loop phosphorylation-mediated activation, as demonstrated

by small-scale in vitro kinase assays that evaluated histone 1 phosphorylation by CDK2 (*Zhang et al., 2017*). The GFP and MS-READ reporter protein were employed in small-scale experiments to quantify OTS-mediated pT incorporation efficiency using fluorescence-based measurements (*Zhang et al., 2017; Moen et al., 2022*). In contrast, recombinant Chk2-KD was utilised for small-scale luminescence and Western blot-based experiments, which explored the effects of site-specific Chk2 phosphorylation on its *in vitro* kinase activity, with itself and the dual-specificity phosphatase, Cdc25C serving as substrates (*Moen et al., 2022*).

Expression and purification of recombinant Rad50-ZCC with pT incorporated at T690 (ZCC_{pT690}) involved three key steps: i) development of a rapid detection method for 690 phosphorylation (Figures 4.14-4.15), ii) optimisation of pT OTSs for T690 site-specific incorporation (Figure 4.16), iii) validation of pT690 incorporation and quantification of incorporation efficiency (Figures 4.17-4.21).

4.3.7.1 Detection

Firstly, a rapid T690 phosphorylation detection method was developed, because previous attempts at generation of a Rad50-pT690 phospho-specific polyclonal antibody failed (Zuzana Hořejší personal communication). As an alternative, the CK2-substrate antibody was identified as a potentially suitable pT690-specific marker for Western blotting, because the minimal consensus sequence for CK2-mediated phosphorylation is S/T-x-x-D/E (*Litchfield, 2003*). Previous analysis of alterations in the HEK-293T phosphoproteome in response to quinalizarin-mediated CK2 inhibition suggests that CK2 phosphorylates Rad50-T690 (*Franchin et al., 2015*). Moreover, the kinase library project recently defined the substrate specificity of 303 human serine/threonine kinase, and both CK2 catalytic subunits of the $CK2\alpha_2$:$CK2\beta_2$ heterotetramer, $CK2\alpha$ and $CK2\alpha$', scored as the most

biochemically favourable Rad50-T690-specific kinases (CK2α log_2: 2.488 and CK2α' log_2: 2.180) (*Johnson et al., 2023*). To validate whether the CK2-substrate antibody would indeed detect ZCC-T690 phosphorylation, *in vitro* CK2 kinase assays involving recombinant CK2A1 and ZCC were necessary. In the first instance, the *Rad50* gene encoding for the ZCC fragment (residues 585-766) was cloned in-frame with the N-terminal GST-3C tag into the pGEX6P-1 vector for recombinant purification. Further phosphonull (T690A), phosphomimetic (T690E), and phospho acceptor-swap (T690S) mutants were made to test the specificity of the CK2-substrate antibody and the effect of CK2-mediated T690 phosphorylation on phosphorylation of other Rad50 sites. A bacterial pDB1-*CK2A1* expression vector was gifted by David Litchfield, which contains the *CK2A1* gene in-frame with an N-terminal GST tag and factor Xa protease cleavage site (I-E/D-G-R↓) *(Litchfield, 2003)*. Recombinant $ZCC_{A/E/S/T690}$ proteins and GST-CK2α were purified by affinity and size-exclusion chromatography as described previously, however GST-CK2α was eluted from glutathione sepharose beads upon addition of 25mM reduced L-glutathione pH 8, to retain the N-terminal GST-tag (Figure 4.12A). Purified $ZCC_{A/E/S/T690}$ and GST-CK2α proteins were slightly heterogenous at ~80% and ~60% purity, due to the presence of potential degradation products. *In vitro* GST-CK2α kinase assays followed by Western blotting with the anti CK2 substrate antibody confirmed that the ZCC fragment was an *in vitro* CK2 substrate (Figure 4.14B). Mutation of the T690 site to either A or S residues had a minor effect, while the phosphomimetic reduced ZCC phosphorylation by as much as 50% (Figure 4.14B). Mass spectrometric analysis of the CK2-phosphorylated ZCC fragment identified 3 highly confident phosphorylated S/T residues: S640 and T690, and one of T663/Y666/S667/T671/T674/S679, with only T674 containing a +3 E residue that is consistent with the CK2 phosphorylation consensus (Figure 4.15). Rad50-S640 phosphorylation was previously identified in a proteomic screen investigating early S-phase/double thymidine-arrested HeLa cells

(*Dephoure et al., 2008*). Again, CK2α and CK2α' scored as the most biochemical favourable S640-specific kinases (CK2A1 log_2: 4.686 and CK2A2 log_2: 5.075). S640 is located at the base of helix 2 within each monomer of the zinc-hook coiled-coiled apex dimer (*Park et al., 2017*). T674 was a newly identified *in vitro* CK2α substrate, however CK2α is the 6th most favourable kinase (CK2A1 log_2: 0.514).

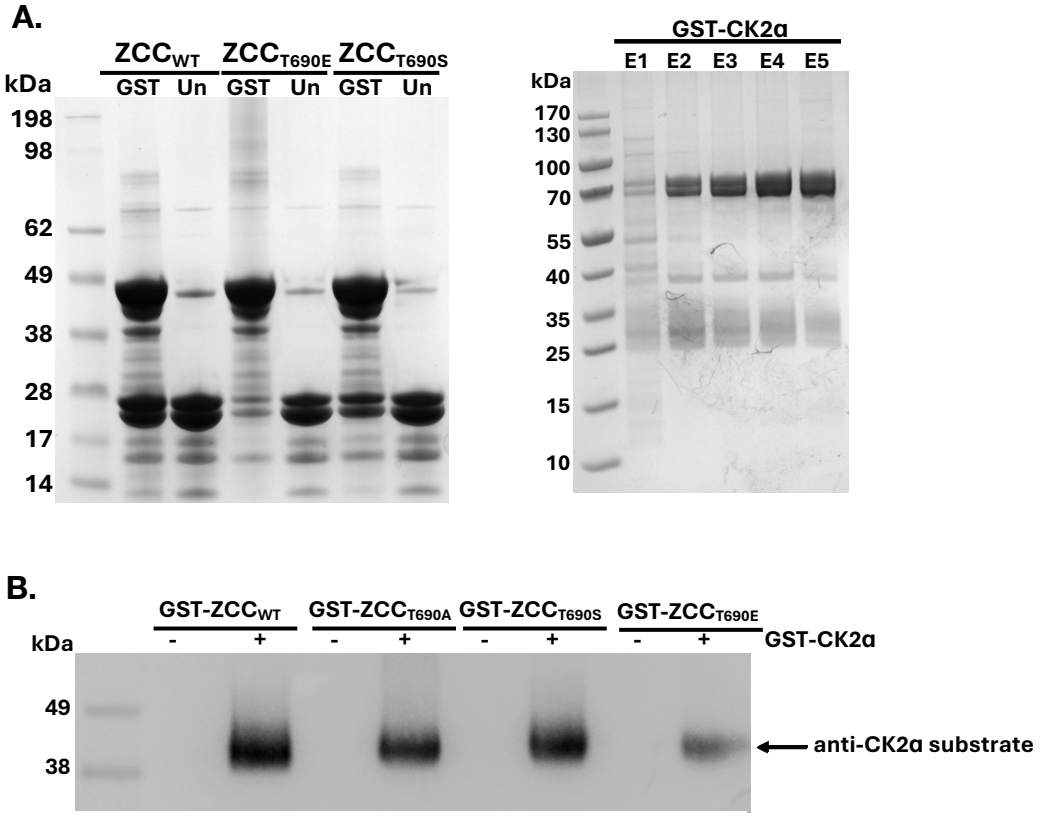
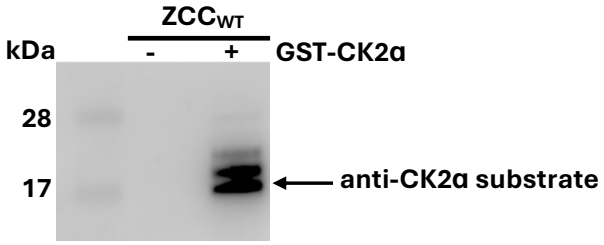


Figure 4.14 – *In vitro* kinase assays suggest that the recombinant CK2 catalytic subunit (CK2α) phosphorylates the Rad50-ZCC fragment independently of the presence of an intact T690 site. (A) SDS-PAGE analysis of purified Rad50-ZCC proteins (left panel) and a CK2 catalytic subunit (right panel). (B) Western blots of *in vitro* kinase assays suggest that CK2α phosphorylates the Rad50-ZCC fragment independently of the presence of an intact T690 site. Samples that were incubated in the presence of GST-CK2α are indicated by a **+** symbol. All samples were incubated at 30°C in the presence of 500μM ATP for 1 hour. Representative image of Western blots conducted using an anti-CK2α substrate antibody from three independent experiments.



Annotated Sequence	Amanda Score	Peptide Modification	PSM	Confidence (Sequest)
[K].LFDVCGSQDFESDLDR.[L]	485.29	pS640	3	High
[R].AMLAGATAVYSQFITQLTDENQSCCPVCQR.[L]	192.14	pT663/Y666/S667/T671/T674/S679	14	High
[R].VFQTEAELQEVISDLQSK.[L]	164.74	pT690	1	High
[R].AMLAGATAVYSQFITQLTDENQSCCPVCQR.[L]	154.29	pT663/Y666/S667/T671/T674/S679	5	Low

Figure 4.15 – *In vitro* kinase assays coupled with mass spectrometry demonstrate that the anti-CK2α antibody is a suitable biomarker for recombinant Rad50-T690 phosphorylation. Western blots of *in vitro* kinase assays (top panel) suggest that CK2α phosphorylates the Rad50-ZCC fragment at multiple sites. Bottoms-up proteomics with post-translational modification analysis (bottom panel) of the sample in the top panel demonstrates that CK2α phosphorylates Rad50-S640,T690, and either T663/Y666/S667/T671/T674/S679 in *in vitro* experiments. Sample preparation and proteomics analysis were conducted by the University of Birmingham Advanced Mass Spectrometry Facility. (PSM: Peptide-Spectrum Match).

4.3.7.2 Rad50 pT690 Production

Following successful development of a T690 phosphorylation detection method, we proceeded with site-specific incorporation of pT residues into the ZCC-T690 position. Firstly, the *ZCC* gene was cloned in-frame with a C-terminal His_6 tag into a pRSF1030C-derivative bacterial expression vector gifted by the Cronan Lab (*Chakravartty and Cronan, 2015*). pRSF1030C is an arabinose-inducible pBAD-derivative expression vector containing a RSF1030 origin of replication, which is compatible with ColE1 and p15A origin-containing vectors, such as the IPTG-inducible pThrOTS$_{Zeus}$ and pThrOTS$_{Zeus}$-Rop$^-$ pThr OTS component expression vector (*Chakravartty and Cronan, 2015; Moen et al., 2022*). Moreover, pRSF1030C contains a chloramphenicol-resistance cassette that differs from

the kanamycin-resistance cassette-containing pThrOTS$_{Zeus}$, and pThrOTS$_{Zeus}$-Rop^{-}, thereby enabling dual antibiotic selection during bacterial cotransformations. Of note, pThrOTS$_{Zeus}$ and pThrOTS$_{Zeus}$-Rop differ from one another due to removal of the repressor of primer (Rop) gene, which reduces plasmid copy number. Later, the *ZCC* gene was cloned in-frame with an N-terminal MBP and 3C cleavage site tag, and a C-terminal His$_6$ tag into pRSF1030C-derivative vectors, termed W40 and AC6, respectively. The major difference between W40/AC6 and pRSF1030C is choice of affinity tags and the presence of an ampicillin-resistance cassette. The ampicillin resistance cassette also enables dual antibiotic selection during bacterial cotransformations, whie the major advantage of this dual N-terminal and C-terminal tag system is the ability to separate amber codon-mediated protein truncates from full-length pT-incorporated ZCC recombinant proteins. W40 contains a slightly stronger ribosome binding site sequence than AC6, which was proposed to increase mRNA levels, and in turn, OTS-mediated cytotoxicity. The bacterial strains used in this study for the pThr OTS, the C321.ΔAΔycdxΔserCΔpphA and BL21(DE3)ΔycdxΔserCΔpphA *E. coli* strains were used, and both were gifted by the Rhinehart Lab (*Moen et al., 2022*). C321.ΔAΔycdxΔserCΔpphA is also an MG1655 derivative with all 321 amber stop codons recoded to ochre stop codons, and *prfA* deletion, while BL21(DE3)ΔycdxΔserCΔpphA retains all 321 amber stop codons and *prfA* (*Moen et al., 2022*). The *ycdx* gene was deleted, because it encodes for threonine phosphatase, and *pphA* was deleted due to encoding for a potential lambda phosphatase homologue (*Moen et al., 2022*). The *serC* gene was deleted to reduce intracellular phosphoserine levels, and thus, phospho-aminoacylation competition, since *serC* encodes a transaminase that is involved in the conversion of 3-phosphohydroxypyruvate to 3-phosphoserine during serine biosynthesis (*Shimizu and Dempsey, 1978; Moen et al., 2022*). Extensive optimisation experiments were conducted at 1L scale, because initial small-scale experiments failed to detect ZCC-pT species. Large-scale

optimisation experiments included expression vector and OTS component vector combination pairs in all aforementioned strains, changes in undefined rich media (LB, TB, and 2TY) and autoinduction media (PASM5052, MDA5052, and NZY-AIM), supplemented with L-arabinose at mid-log phase inoculation density (A_{600nm}: 0.05 and 0.1), L-arabinose concentration for protein expression (0.05%, 0.1%, and 0.2%), and post-induction temperature (18°C and 37°C) and duration (4 hours and 18 hours). Across both OTSs and all strains studied, the best performing growth expression strategy was the following: inoculation into 1L TB in 2.5L Tunair half-baffle shake flasks at A_{600} of 0.1, growth at 37°C, 250rpm until an A_{600} of 0.4 was reached, whereupon 1mM IPTG and 2% (w/v) L-arabinose induced expression for a further 4 hours at 37°C, 250rpm. The pThrOTS-Rop$^-$ plasmid combinations resulted in a ~250-fold increase in ZCC yield, but strangely no pT incorporation was detected.

There was negligible difference in pT-incorporated ZCC-T690 yields between the C321.ΔAΔycdxΔserCΔpphA and BL21(DE3)ΔycdxΔserCΔpphA pT strains. In general, C321.ΔAΔycdxΔserCΔpphA was the preferred pT-incorporation strain, because all 321 amber stop codons in the BL21-derivative strain were not recoded to ochre stop codons, and the release factor 1 gene was present (*Moen et al., 2022*). After optimisation, the average yield of recombinant MBP-3C-ZCC-pT690-His_6 was ~5μg/L culture (Figure 4.16).

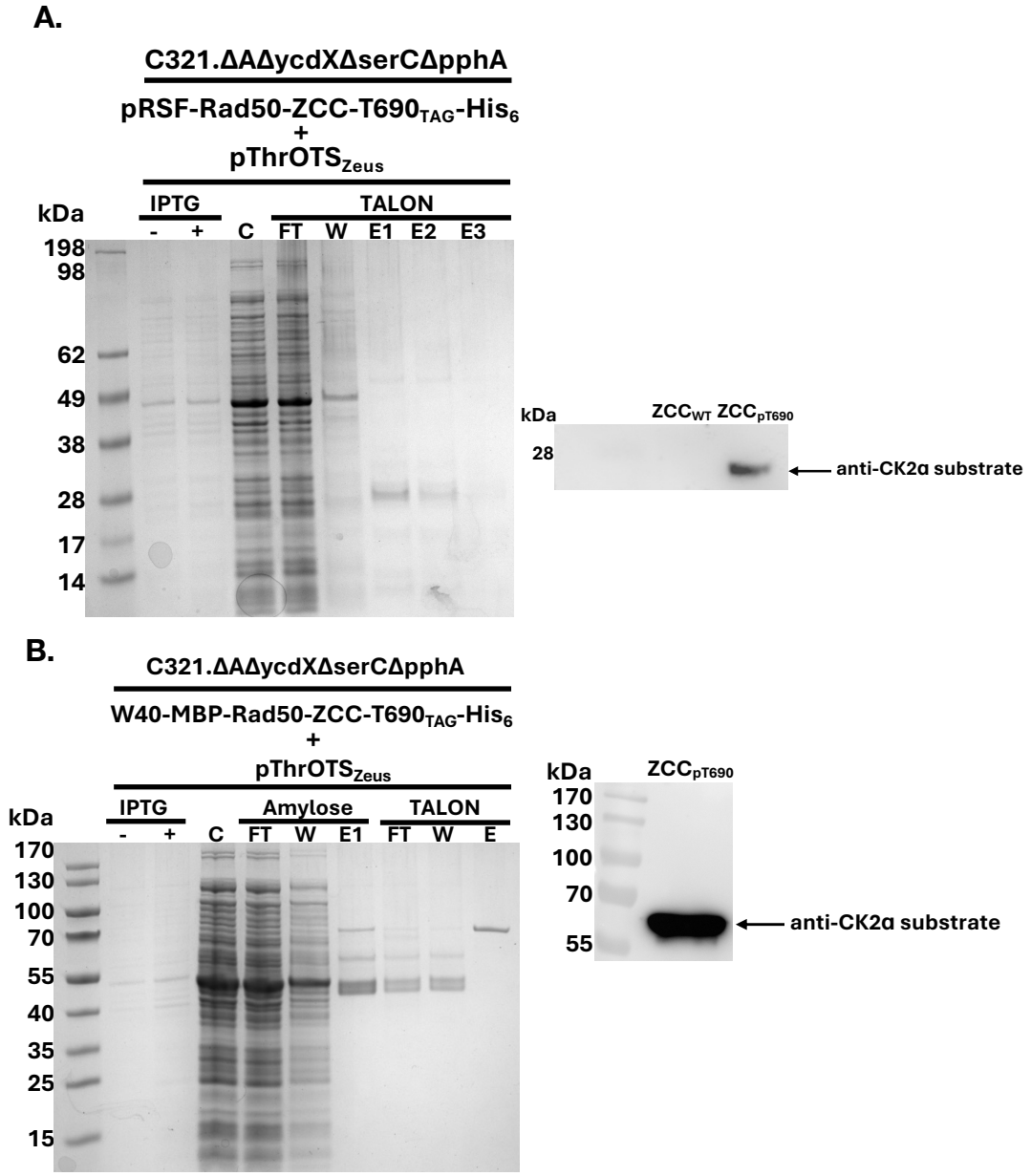
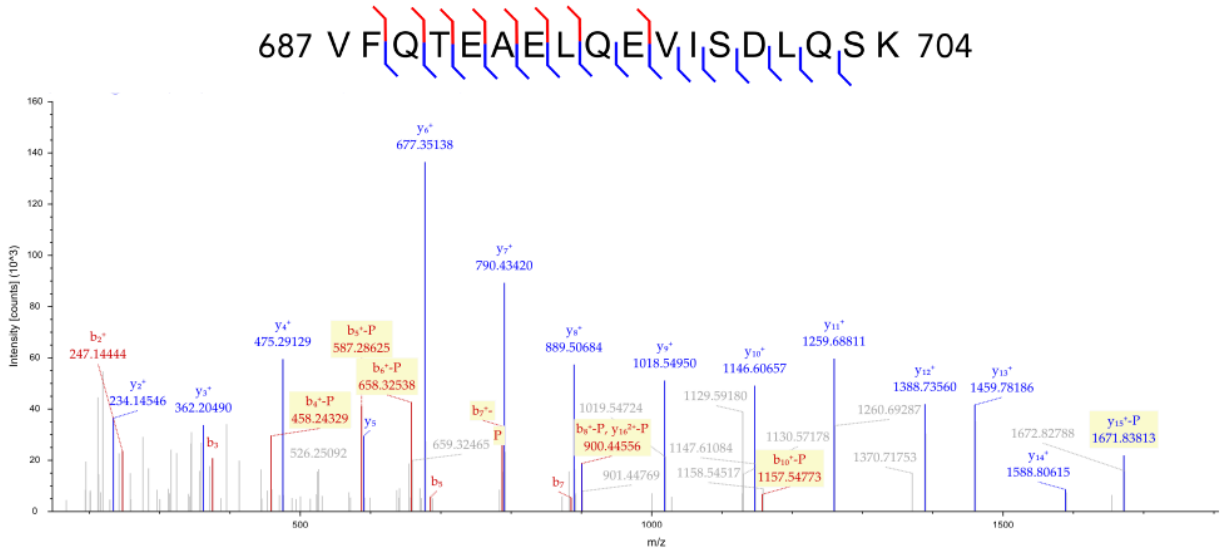


Figure 4.16 – Optimisation of OTS-dependent site-specific incorporation of pThr into recombinant Rad50-690 protein. (A) Purification and Western blot validation of recombinant Rad50-pT690-His$_6$ using the Rinehart laboratory pThr OTS with the pRSF1030C vector. SDS-PAGE analysis of Rad50-pT690-His$_6$ purification are shown in the left panel. Clarified lysate (C), flow through (FT), TALON bead washes (W), and elutions (E) are indicated. Western blot using the anti-CK2α substrate of the purified protein compared to a non-phosphorylated control is provided in the right panel. **(B) Purification and Western blot validation of recombinant MBP-3C-Rad50-pT690-His6 using the Rinehart laboratory pThr OTS with the W40 vector.**

4.3.7.3 Validation/quantification of pT690 incorporation

Successful pT incorporation T690 in MBP-3C-ZCC-pT690-His_6 was validated by Western blot detection, followed by mass spectrometry analysis of a positive anti-CK2-substrate band (Figures 4.16-4.17). No phosphopeptide enrichment strategies, such as titanium dioxide, were utilised because it was not provided by the University of Birmingham Advanced Mass Spectrometry Facility. While this resulted in reduced MS1 and MS2 spectrum signal intensities, it was an unbiased approach that confirmed site-specific phosphothreonine incorporation. Specifically, our MS2 spectrum data of a fragmented VFQpTEAELQEVISDLQSK phosphopeptide exhibited mass-to-charge peaks corresponding to collision-induced dissociation of a neutral H_3PO_4 group in $b4^+$-P, b_5^+-P, b_6^+-P, b_8^+-P, y_{16}^{2+}-P, b_{10}^+-P, and y_{15}^+-P (TEAELQEVISDLQSK) ions that closely resembled predicted masses (Figure 4.17). Taken together, the mass-to-charge ratio of these fragmented ions demonstrate that there is a 100% possibility that our recombinant Rad50 protein was phosphorylated at T690.

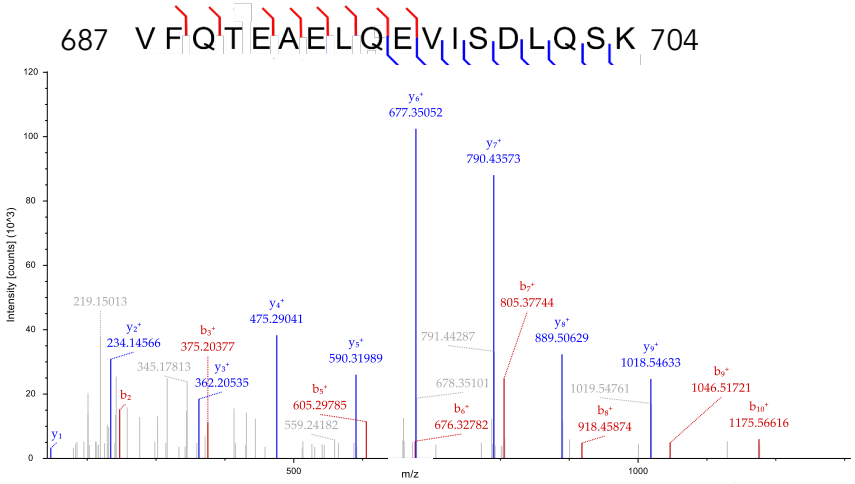


Ion	Theoretical mass (Da)	Observed mass (Da)
b_4-H_3PO_4	458.2398	458.2433
b_5-H_3PO_4	587.2824	587.2863
b_6-H_3PO_4	658.3195	658.3254
b_8-H_3PO_4	900.4462	900.4456
y_{16}^{2+}-H_3PO_4	900.4547	900.4456
b_{10}-H_3PO_4	1157.5473	1157.5478
y_{15}-H_3PO_4	1671.8436	1671.8381

Figure 4.17 - Tandem mass spectrometry (MS/MS) validation of site-specific phosphothreonine incorporation at Rad50-T690 in recombinant MBP-3C-Rad50-pT690-His6 protein. Mass-to-charge ratio (m/z) relative to ion abundance (intensity) of ions identified from MS2 analysis of a Rad50 VFQpTEAELQEVISDLQSK phosphopeptide is indicated. Neutral loss (-) of H_3PO_4 (P) during collision-induced dissociation is indicated. B^+ (blue) and y^+ (red) ions are mapped to the phosphopeptide sequence (top). Experimentally-observed phospho-specific ion masses are compared to theoretical masses calculated using the UCSF Protein Prospector MS-Digest webserver (https://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest). **Sample preparation, MS/MS and data analysis was performed by the University of Birmingham Advanced Mass Spectrometry Facility.**

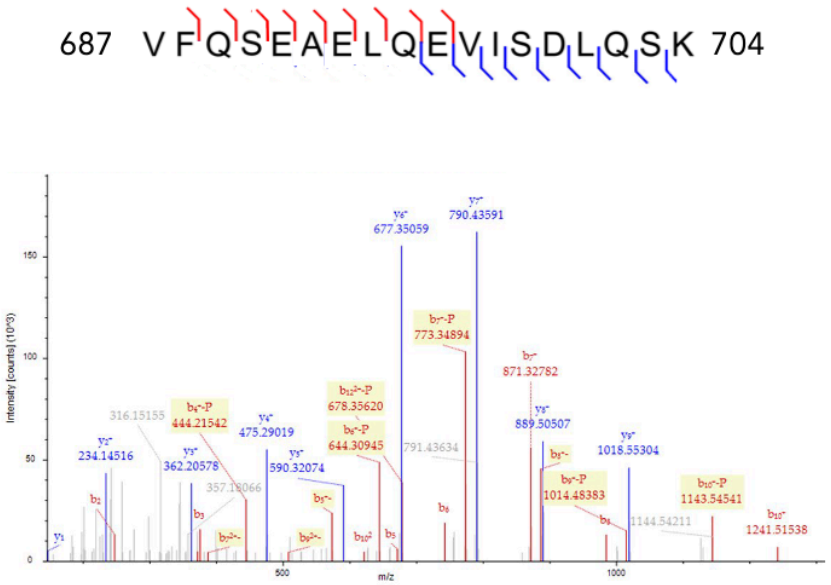
Despite this success, further verification of the pThr OTS fidelity was crucial, as native amino acid and phosphoserine misincorporation at target sites has been observed in up to 33% of previously

tested MS-READ reporter proteins (*Moen et al., 2022*). Extensive mass spectrometry analyses of the ZCC-pT690 species indicate that T, pS, S, and G misincorporation events at the 690 position also occurred (Figures 4.18-4.20).



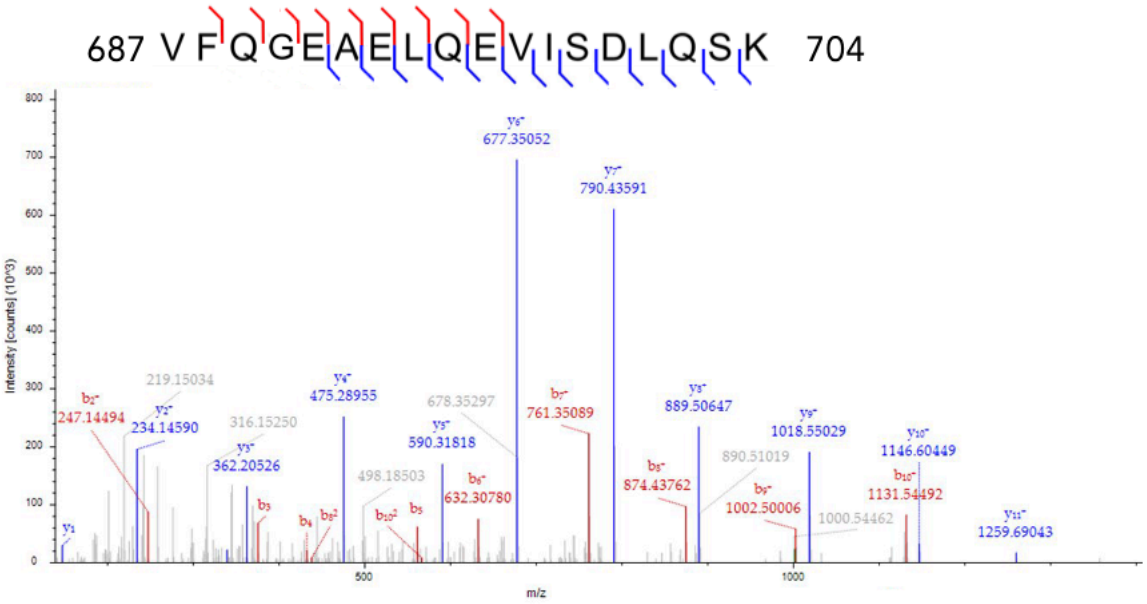
Ion	Theoretical mass (Da)	Observed mass (Da)
b_3	375.2027	375.2038
b_5	605.2930	605.2979
b_6	676.3301	676.3278
b_7	805.3727	805.3774
b_8	918.4567	918.4587
b_9	1046.5153	1046.5172
b_{10}	1175.5579	1175.5662

Figure 4.18 - Tandem mass spectrometry (MS/MS) validation of site-specific threonine misincorporation at Rad50-T690 in recombinant MBP-3C-Rad50-pT690-His6 protein. Mass-to-charge ratio (m/z) relative to ion abundance (intensity) of ions identified from MS2 analysis of a Rad50 VFQTEAELQEVISDLQSK peptide is indicated. B^+ (blue) and y^+ (red) ions are mapped to the phosphopeptide sequence (top). Experimentally-observed of four non-phospho-specific ion masses are compared to theoretical masses calculated using the UCSF Protein Prospector MS-Digest webserver (https://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest). **Sample preparation, MS/MS and data analysis was performed by the University of Birmingham Advanced Mass Spectrometry Facility.**



Ion	Theoretical mass (Da)	Observed mass (Da)
b_4-H_3PO_4	444.2241	444.2154
b_6-H_3PO_4	644.3039	644.3095
b_7-H_3PO_4	773.3464	773.3489
b_9-H_3PO_4	1014.4891	1014.4838
b_{10}-H_3PO_4	1143.5454	1143.5317

Figure 4.19 - Tandem mass spectrometry (MS/MS) validation of site-specific phosphoserine misincorporation at Rad50-T690 in recombinant MBP-3C-Rad50-pT690-His6 protein. Mass-to-charge ratio (m/z) relative to ion abundance (intensity) of ions identified from MS2 analysis of a Rad50 VFQpSEAELQEVISDLQSK phosphopeptide is indicated. B^+ (blue) and y^+ (red) ions are mapped to the phosphopeptide sequence (top). Experimentally-observed of phospho-specific ion masses are compared to theoretical masses calculated using the UCSF Protein Prospector MS-Digest webserver (https://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest). Sample preparation, MS/MS and data analysis was performed by the University of Birmingham Advanced Mass Spectrometry Facility.



Ion	Theoretical mass (Da)	Observed mass (Da)
b_6	632.3039	632.3078
b_7	761.3464	761.3509
b_8	874.4305	874.4376
b_9	1002.4891	1002.5001
b_{10}	1131.5317	1131.5449

Figure 4.20 - Tandem mass spectrometry (MS/MS) validation of site-specific glycine misincorporation at Rad50-T690 in recombinant MBP-3C-Rad50-pT690-His6 protein. Mass-to-charge ratio (m/z) relative to ion abundance (intensity) of ions identified from MS2 analysis of a Rad50 VFQGEAELQEVISDLQSK peptide is indicated. b^+ (blue) and y^+ (red) ions are mapped to the peptide sequence (top). Experimentally-observed of non-phospho-specific ion masses are compared to theoretical masses calculated using the UCSF Protein Prospector MS-Digest webserver (https://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest). **Sample preparation, MS/MS and data analysis was performed by the University of Birmingham Advanced Mass Spectrometry Facility.**

The mass spectrometry validated ZCC-pT690 proteins were then quantified by phos-tag SDS-PAGE to determine the extent of ZCC-690 misincorporation events (Figure 4.21). Briefly, phos-tag SDS-PAGE is an electrophoretic technique that involves a 594Da divalent zinc or manganese ion-based phosphoryl-binding ligand complex, phos-tag (*Kinoshita et al., 2006*). Phos-tag is co-polymerised

with acrylamide in the resolving gel, where it specifically binds to phosphorylated proteins (*Kinoshita et al., 2006*). In turn, phos-tag binding retards the migration of phosphorylated proteins during electrophoresis, thereby enabling separation of individual proteins based on phosphorylation status (*Kinoshita et al., 2006*). Similar to published data, our analyses appear to indicate that approximately 48-67% of the recombinant ZCC-pT690 protein was phosphorylated, although our phosphorylated protein contained misincorporated pS residues (Figure 4.21). These analyses should be repeated with a non-phosphorylated control using commercial phos-tag pre-cast gels. Due to time constraints and technical complexities, we were unable to determine the extent of phosphoserine misincorporation from our phosphorylated ZCC-690 species, but it noinetheless remains a more physiological representation of endogenous Rad50-T690 phosphorylation than all other aforementioned approaches.

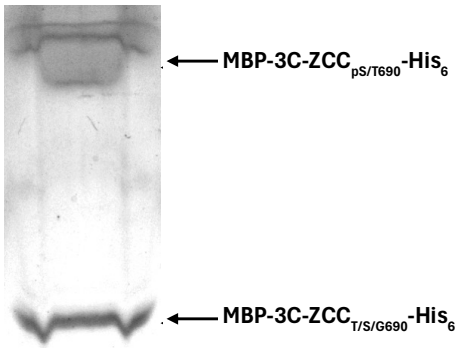


Figure 4.21 - Phos-tag SDS-PAGE analysis suggests that ~48-67% of the recombinant MBP-3C-Rad50-pT690-His6 protein is phosphorylated. 50μM phos-tag and 100μM $MnCl_2$ was copolymerised with 8% (w/v) acrylamide/bis-acrylamide in resolving gels. The upper band represents pSer and pThr incorporated recombinant MBP-3C-ZCC$_{690}$-His$_6$ phosphoprotein, while the lower represents Thr, Ser, and Gly-incorporated MBP-3C-ZCC$_{690}$-His$_6$ protein. Representative image of a single experiment.

4.3.8 ZCC Binding - MST

Our verified MBP-3C-ZCC-pT690-His_6 recombinant phosphoprotein was then fluorescently labelled using our NHS-reactive dye for MST-binding experiments with either PBD or Pincer as titrated substrates. Given the low yield from the pT OTS, we only purified enough MBP-3C-ZCC-pT690-His_6 from 30L bacterial culture growths for one shot PBD/Pincer binding experiments. All experiments were validated by quality control checks via pre- and post-binding fluorescence scans of all individual MST capillaries. Our MST binding experiments appear to suggest that MBP-3C-ZCC-pT690-His_6 did not interact with either PBD or Pincer upon titration of 18.9nM-175μM of each protein (Figure 4.22).

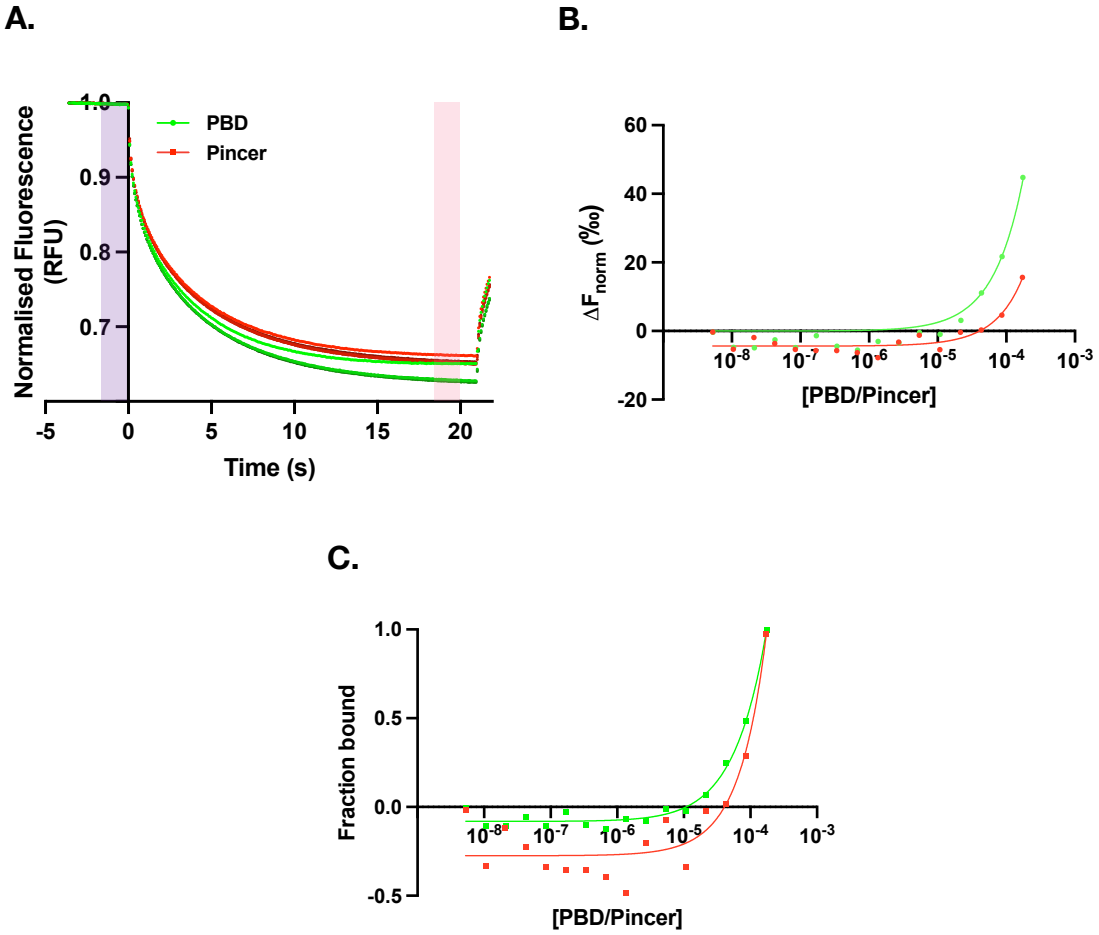


Figure 4.22 - MST binding analysis suggests that recombinant ZCC_{pT690} protein does not interact with either PBD proteins. (A) Raw fluorescence MST traces of ZCC_{pT690} binding to PBD and Pincer proteins. Each coloured line represents the average normalised fluorescence values quantified from 4 capillaries for PBD (green) and Pincer (red) binding experiments. A colour gradient scheme is used, whereby brighter colours denote capillaries containing maximal PBD/Pincer concentration and darker colours represent decreasing in PBD/Pincer concentration. The cold (purple) and hot (pink) region of the experiment used to calculate ΔF_{norm} and fraction bound values is indicated. **(B) Representative ΔF_{norm} plots suggests that ZCC_{pT690} exhibits positive thermophoretic behaviour only above >10-50μM PBD/Pincer.** The concentration of PBD/Pincer is indicated on a $\log_{10}$ scale. (C) A representative fraction bound plot is used to calculate the affinity (K_D) of any potential interactions between ZCC_{pT690} and PBD proteins. Individual data points represent single values from one experiment.

Clearly, it was not possible to confidently conclude from a single experiment that there was no interaction between MBP-3C-ZCC-pT690 and PBD, and the effect of retaining ZCC N- and C-terminal affinity tags on interacting with PBD were unknown. We, therefore, conducted MST binding analyses

with wild-type (ZCC) and phosphonull (ZCC_{T690A}) MBP-3C-ZCC-T690 that was either non-phosphorylated or pre-phosphorylated by CK2 (pZCC and $pZCC_{T690A}$) in the absence or presence of 5mM EDTA (Figure 4.23). EDTA-based controls were included to chelate Zn^{2+} ions, thereby disrupting any zinc-hook mediated dimerisation, however previous structural analyses have demonstrated that the ZCC forms coiled-coil-mediated dimers even under Zn^{2+} free conditions (*Park et al., 2017*). Strangely, MST-binding experiments under these conditions suggested that ZCC bound with similar affinities to both PBD and Pincer in a phosphorylation-dependent and Zn^{2+}-dependent manner (Figure 4.23). Specifically, the K_D between CK2 phosphorylated MBP-3C-ZCC-T690-His_6 (MBP-3C-pZCC-His_6) with PBD was 1.01μM (95% CI: 0.775 - 1.31μM), while with Pincer it was 1.14μM (95% CI: 0.843 - 1.54μM) (Table 4.3). In contrast, no physiologically-relevant interaction was detected with the non-phosphorylated MBP-3C-ZCC-His_6 or upon EDTA-mediated removal of Zn^{2+} from the zinc-hook in both non-phosphorylated and phosphorylated ZCC (Figure 4.23). Similar to the previous PBDtide interactions, our normalised fluorescence change plots indicate that MBP-3C-pZCC-His_6 exhibits reduced thermophoretic behaviour in response to increasing concentrations of PBD and Pincer. Taken together, our initial MST data suggested that MBP-3C-ZCC-His_6 interacted with both PBD and Pincer proteins in a CK2 phosphorylation-dependent manner and was also dependent on zinc-hook-mediated zinc binding, but the interaction did not require phosphorylation of T690. The apparent T690 phosphorylation-independent binding mode was pursued by further MST-binding experiments between CK2-phosphorylated MBP-3C-ZCC-T690A with PBD or Pincer, wherein both proteins exhibited similar binding affinities of 150.7μM (95% CI: 49.1 - 485.6μM) and 132.0μM (95% CI: 37.7 - 624.5μM), respectively. In turn, this would imply that the binding of ZCC to PBD proteins is dependent on the phosphorylation status of T690, but independent of the Pincer motif. These data were clearly in stark contrast to the preliminary data generated by previous

Smerdon Lab members. There were many possibilities for these data. While inconceivable from all pre-existing PBD interaction analyses and our preliminary data, it was possible that ZCC did interact with PBD pT690-independently and at a site away from the Pincer motif. However, it was equally possible that phosphorylation of ZCC could induce local or global conformational changes in the zinc-hook motif and coiled-coil region, respectively, which was exacerbated upon addition by PBD/Pincer. It was also possible that any local or global conformational changes in ZCC could result in non-specific interactions with PBD/Pincer or altered thermophoretic behaviour. Alternatively, erroneous MST fluorescence signals could be a result of buffer mismatch associated with impurities from purification and *in vitro* kinase assays, despite dialysis of both proteins.

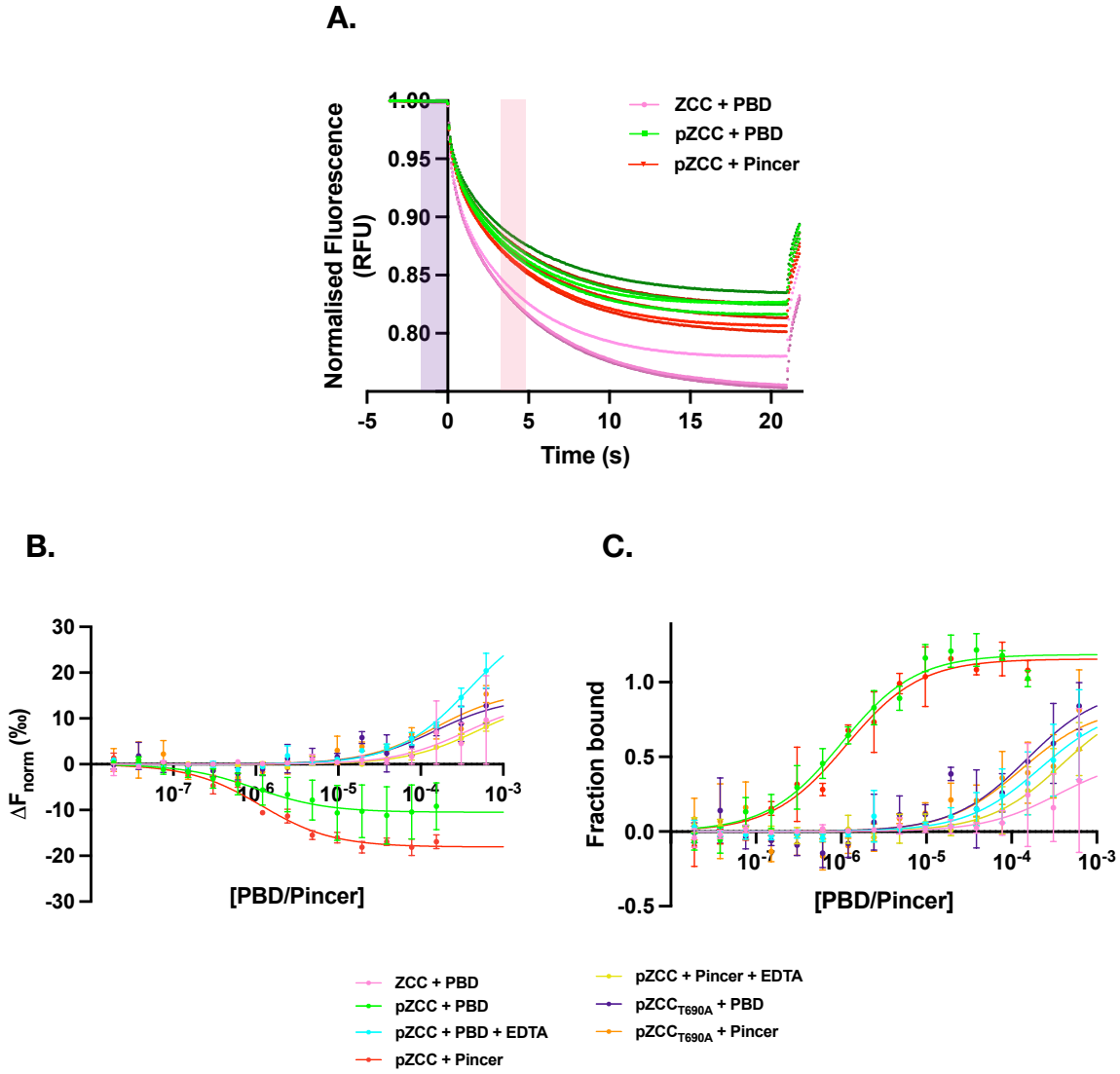


Figure 4.23 - MST binding analysis suggests that recombinant ZCC protein interacts with both PBD proteins in a CK2-mediated phosphorylation-dependent and oligomeric state-dependent manner. (A) Raw fluorescence MST traces of CK2-phosphorylated wild-type ZCC (pZCC) and phosphonull T690 ($pZCC_{T690A}$) binding to PBD and Pincer proteins, and non-phosphorylated ZCC not. Each coloured line represents the average normalised fluorescence values quantified from 4 capillaries for binding experiments involving non-phosphorylated ZCC with PBD (pink), and pZCC with either PBD (green), or with Pincer (red). A colour gradient scheme is used, whereby brighter colours denote capillaries containing maximal PBD/Pincer concentration and darker colours represent decreasing in PBD/Pincer concentration. The cold (purple) and hot (pink) region of the experiment used to calculate ΔF_{norm} and fraction bound values is indicated. **(B) Representative ΔF_{norm} plots suggest that pZCC exhibits negative thermophoretic behaviour upon binding to PBD and Pincer proteins.** The concentration of PBD/Pincer is indicated on a $\log_{10}$ scale. (C) A representative fraction bound plot is used to calculate the affinity (K_D) of interactions between ZCC and PBD proteins. Individual data points represent mean values and error bars are the standard deviation values from three independent experiments.

PBD	K_D (μM)
ZCC	352.1
pZCC	1.01
pZCC + EDTA	216.3
$pZCC_{T690A}$	150.7

Pincer	K_D (μM)
pZCC	1.14
pZCC + EDTA	392.2
$pZCC_{T690A}$	132.0

Table 4.3 - Summary of MST binding analyses conducted. The affinity/dissociation constant (K_D; in μM) is provided. All individual data points represent the best fit value from fitting a one site-specific binding non-linear regression model.

4.3.9 ZCC Binding – Size-Exclusion Chromatography

Based on the foregoing, we decided that purification of any potential CK2 phosphorylation-dependent PBD:ZCC complexes would resolve whether the apparent pT690-independent binding mode was possible. Alternatively, a failure to purify the complex would point towards MST buffer mismatch or altered pZCC conformational dynamics. To achieve this, we employed size-exclusion chromatography, because it can detect changes in the hydrodynamic radius of PBD and MBP-3C-pZCC-His_6 proteins under native conditions.

Initially, 92μM of untagged pZCC was analysed by gel-filtration to establish an elution baseline elution volume. This process was repeated using a 1:1 molar stoichiometric ratio of untagged 92μM PBD:pZCC, however of co-elution of both proteins was not observed (Figure 4.24). It was

hypothesised that during gel-filtration, the dilution of the samples may have lowered the protein concentration below the dissociation constant determined from MST-binding experiments.

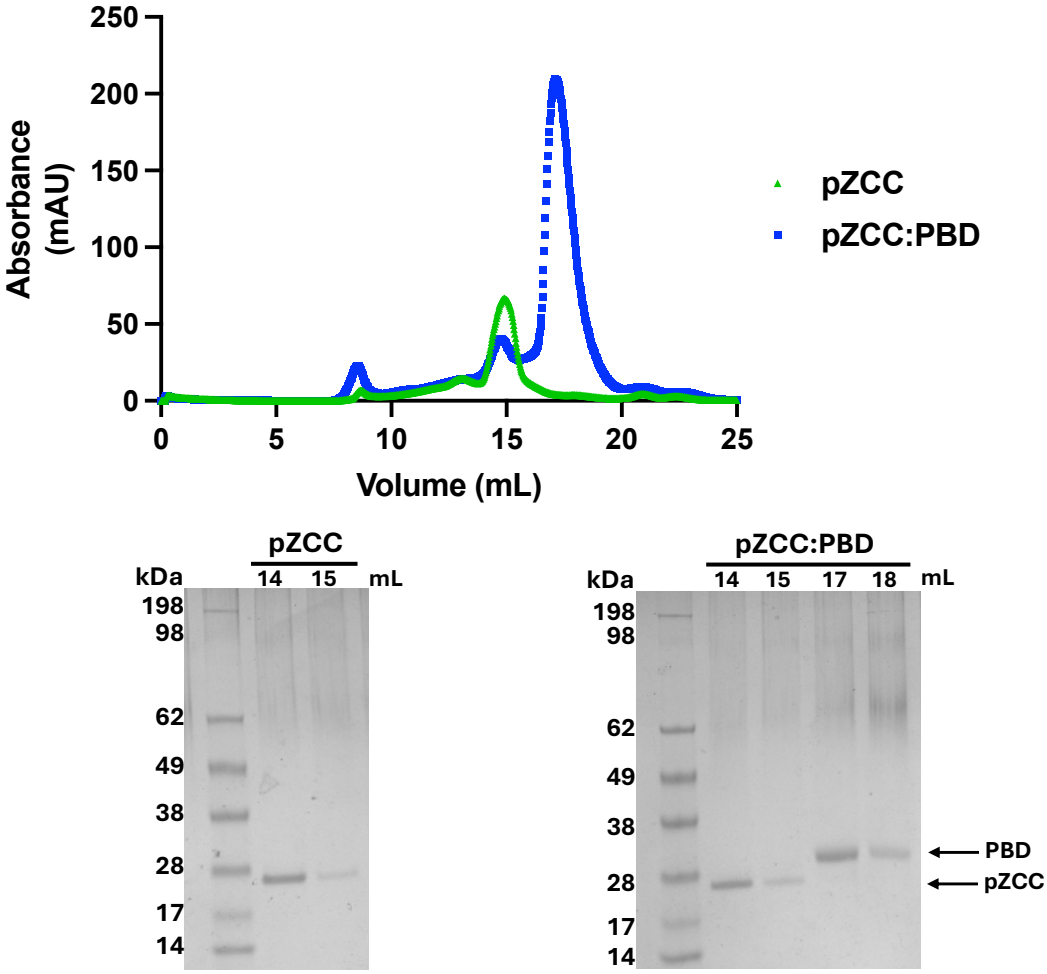


Figure 4.24 - Analytical SEC complex analysis suggests that pZCC and PBD do not form a stable complex at a 1:1 stoichiometric molar ratio of ~90µM loading concentrations. (Top panel) SEC analysis of ~90µM pZCC protein alone (green triangles) and ~90µM pZCC:PBD (top panel) is indicated with volume (mL) indicated on the x-axis and absorbance at 280nm (mAU) indicated on the y-axis. (Bottom panel) SDS-PAGE analysis of pZCC protein alone (left panel) and pZCC and PBD complex SEC analysis (right panel). The fraction elution volume is indicated at the top of each gel. Individual proteins are indicated with an arrow. Representative gel from three independent experiments.

To address this, the concentration of both the individual PBD and pZCC proteins and the potential 1:1 complex was increased to 248µM. Despite this adjustment, no definitive co-elution was

observed, and definitive confirmation was challenging due to the similar molecular weight of PBD and pZCC proteins, which limited differentiation of individual proteins (Figure 4.25).

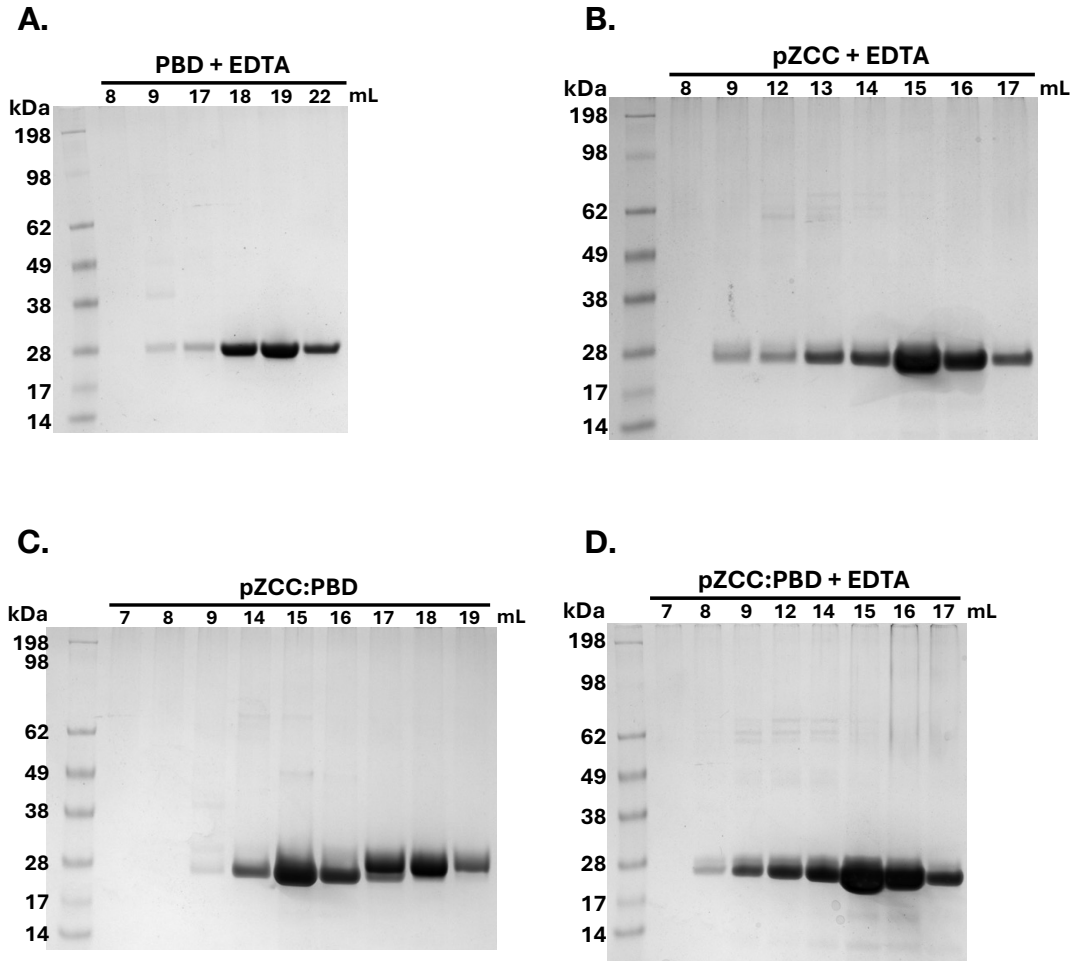


Figure 4.25 - SDS-PAGE analysis of analytical SEC complex analysis suggests that pZCC and PBD do not form a stable complex at a 1:1 stoichiometric molar ratio of ~250μM loading concentrations. (A) PBD protein alone in the presence of 5mM EDTA. (B) pZCC protein alone in the presence of 5mM EDTA. (C) Untreated PBD:pZCC. (D) PBD:pZCC in the presence of 5mM EDTA. The fraction elution volume is indicated at the top of each gel. Representative gels from three independent experiments.

To resolve this and mitigate any issues associated with binding saturation, a 43kDa N-terminal MBP tag was retained on our ZCC protein, and the stoichiometric ratio was raised to 1:13. Again, baseline elution volumes of 24μM MBP-3C-pZCC-His_6 and 308μM PBD were individually analysed by gel-filtration (data not shown) to establish baseline elution volumes for each protein. Retention of the

43kDa N-terminal MBP tag on our ZCC protein facilitated differentiation between both proteins, and PBD eluted 4mL later than MBP-3C-pZCC-His_6. Following this, 24μM MBP-3C-pZCC-His_6 and 308μM PBD, were mixed together for 10 minutes at 4°C and then separated by gel-filtration (Figure 4.26). The elution volumes of complex samples were consistently identical to those observed in the separate runs for each protein. These results clearly demonstrated that a ~13-fold molar excess of PBD was insufficient to interact with MBP-3C-pZCC-His_6, in contradiction to the tight K_D values obtained from MST experiments.

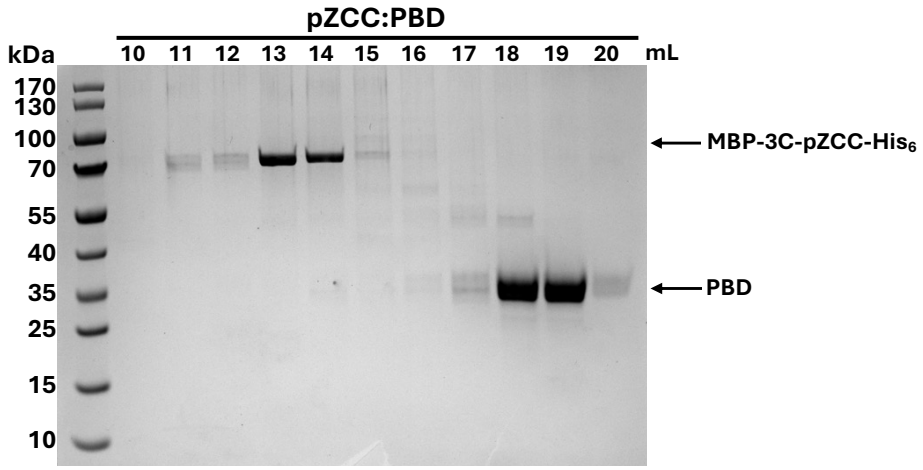


Figure 4.26 - SDS-PAGE analysis of analytical SEC complex analysis suggests that pZCC and PBD do not form a stable complex at a 1:13 stoichiometric molar ratio. The fraction elution volume is indicated at the top of each gel and individual proteins are indicated with an arrow. Representative gels from three independent experiments.

In summary, the SEC analyses did not align to the K_D measurements obtained by MST binding experiments (Figure 4.23). One possible explanation for this data was that the relative stoichiometry of ZCC phosphorylation was low - sufficiently high for detection of complex formation by sensitive MST experiments, but insufficiently low for analytical SEC. However, the possibility of an interaction between a small proportion of pZCC and PBD could not be formally excluded. Consequently, it

remained inconclusive whether the MST data reflected buffer mismatch, altered pZCC conformational dynamics or solvation energy changes induced by PBD/Pincer proteins.

4.3.10 ZCC binding - Chemical cross-linking

Therefore, we conducted homobifunctional amine crosslinker analysis using BS^3 and BS(PEG9), because both agents could irreversibly cross-link solvent exposed intermolecular α- and ε-amine groups in PBD and MBP-3C-pZCC-His_6. Moreover, any cross-links could be formed distance-dependently, since BS^3 possesses a moderate spacer arm length of 11.4Å while the long BS(PEG9) spacer arm length is 35.8Å. BS^3 experiments were conducted by combination of MBP-3C-pZCC-His_6 and PBD to final concentrations of 25μM and 40μM, respectively, in the presence of 2mM BS^3. The BS^3-treated solution was separated by gel filtration and there was no significant difference in elution volumes of either protein (data not shown). Interestingly, two larger molecular weight bands emerged from the peak corresponding to MBP-3C-pZCC-His_6, with approximate sizes of 170kDa and >170kDa, respectively (Figure 4.27). Given that the molecular weight of non-cross linked MBP-3C-pZCC-His_6 is ~64kDa and PBD is ~28kDa, these analyses suggested that BS3-induced intermolecular cross links induced formation of potential ~184kDa 2:2 heterotetramers and either ~224kDa PBD or ~256kDa MBP-3C-pZCC-His_6 homotetramers. Mass spectrometry analyses of these bands revealed intramolecular cross-linked peptides exclusively derived from self-crosslinking of the N-terminal MBP tag, between the MBP tag and pZCC protein, and self-crosslinking of pZCC (Tables 4.4-4.5). Surprisingly, PBD peptides containing single BS^3 modified lysines were also detected, despite the stark difference in molecular weight between the analysed cross-linked bands and single-modified PBD (Tables 4.4-4.5). In any case, these analyses suggested that BS^3 had induced formation of oligomeric MBP-3C-pZCC-His_6. These results were also observed

using similar concentrations of MBP-3C-pZCC-T690A-His_6 and PBD, which suggested that oligomer formation was not dependent on the phosphorylation status of T690. Finally, BS(PEG9) cross-linking analysis was performed in solution with 5μM untagged variants of ZCC-T690, pZCC, ZCC-T690A, pZCC-T690A in the presence of either 25μM PBD or Pincer and there was no noticeable difference in band intensity across all conditions (Figure 4.28). In summary, these data suggested that CK2-mediated phosphorylation of ZCC had no effect on the interaction with PBD proteins, nor on self-association, and that our MST binding data was a result of buffer mismatch. While this data was extensive, it still did not conclusively demonstrate a lack of interaction, because analysis of the ZCC crystal structure indicates a paucity of reactive amine groups proximal to the T690 site *(Park et al., 2017)* requiring a more targeted approach.

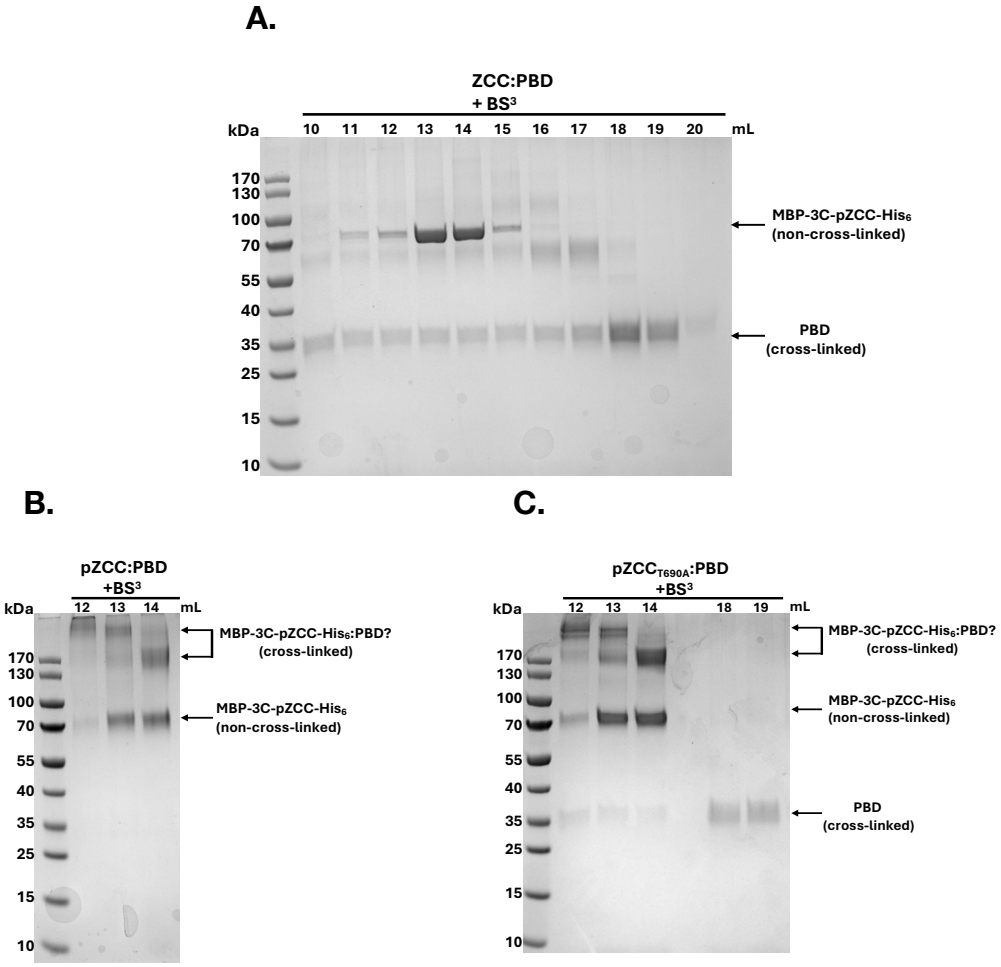


Figure 4.27 - SDS-PAGE analysis of analytical SEC from BS^3 cross-linking experiments suggests that ZCC and PBD may interact phosphorylation-dependently, but independently of T690 phosphorylation status. SDS-PAGE analysis from BS^3 cross-linking ZCC:PBD (A) pZCC:PBD (B) and pZCC$_{T690A}$:PBD (C) experiments. Potential cross-linked and non-cross-linked species are indicated. The fraction elution volume is indicated at the top of each gel. Representative gels from three independent experiments.

Annotated Sequence A (Rad50/MBP aa)	Annotated Sequence B (Rad50/MBP aa)	Max XlinkX Score	Xlink Type
[K]IEEGK MBP 2-7	SS[K]QR Rad50 652-656	31.72	Intra
GYNGLAEVG[K]K MBP 17-27	FE[K]DTGIK MBP 28-35	37.38	Intra
YENG[K]YDIK MBP 172-180	L[K]EINQTR Rad50 585-591	35.1	Intra
YENG[K]YDIK MBP 172-180	N[K]LQNVNR Rad50 752-759	35.1	Intra
SYEEELA[K]DPR MBP 307-317	N[K]LQNVNR Rad50 752-759	38.79	Intra
L[K]EINQTR Rad50 585-591	LA[K]LNK Rad50 594-599	28.09	Intra
R[K]EEQLSSYEDK Rad50 617-628	N[K]LQNVNR Rad50 752-759	102.22	Intra
[K]EEQLSSYEDK Rad50 618-628	E[K]EIPELR Rad50 744-751	49.03	Intra
L[K]EEIEK Rad50 645-651	SS[K]QR Rad50 652-656	44.79	Intra
STESEL[K]KK Rad50 714-722	L[K]EEIEK Rad50 645-651	59.82	Intra
STESEL[K]KK Rad50 714-722	SS[K]QR Rad50 652-656	40.95	Intra
QSIIDL[K]EK Rad50 737-745	L[K]EEIEK Rad50 645-651	32.13	Intra
QSIIDL[K]EK Rad50 737-745	N[K]LQNVNR Rad50 752-759	32.13	Intra
E[K]EIPELR Rad50 744-751	L[K]EEIEK Rad50 645-651	27.06	Intra
N[K]LQNVNR Rad50 752-759	LA[K]LNK Rad50 594-599	20.81	Intra
N[K]LQNVNR Rad50 752-759	E[K]EIPELR Rad50 744-751	50.9	Intra

Annotated Sequence (PLK1 aa)	XCore Score	Peptide Modification	PSM	Confidence (Sequest)
[R].DGTESYLTVSSHPNSLMKK.[I] 457-475	3.12	DSS Hydrolysed [K]	1	High
[R].NYMSEHLLKAGANITPR.[E] 484-500	4.52	DSS Hydrolysed [K9]	1	High
[R].NYMSEHLLKAGANITPR.[E] 484-500	4.02	DSS Hydrolysed [K9]	1	High

Table 4.4 - Tandem Mass spectrometry analysis (MS/MS) of the ~170kDa cross-linked species band from pZCC:PBD BS^3-cross linking experiments. Data indicate presence of only Rad50 intra cross-linked proteoforms (top table), however single BS^3-modified PLK1 species are also present (bottom table). (PSM: Peptide-Spectrum Match).

Annotated Sequence A (Rad50/MBP aa)	Annotated Sequence B (Rad50/MBP aa)	Max Xlink Score	Xlink Type
GYNGLAEVG[K]K MBP 17-27	FE[K]DTGIK MBP 28-35	26.11	Intra
STESEL[K]KK Rad50 714-722	SS[K]QR Rad50 652-656	40.95	Intra
L[K]EEIEK Rad50 645-651	SS[K]QR Rad50 652-656	31.92	Intra
[K]EEQLSSYEDK Rad50 618-628	N[K]LQNVNR Rad50 752-759	65.59	Intra

Annotated Sequence (PLK1 aa)	XCorr Score	Peptide Modification	PSM	Confidence (Sequest)
[R].DGTESYLTVSSHPNSLMKK.[I] 457-475	3.45	DSS Hydrolysed [K]	2	High
[R].NYMSEHLLKAGANITPR.[E] 484-500	2.85	DSS Hydrolysed [K9]	1	High

Table 4.5 - Tandem Mass spectrometry analysis (MS/MS) of the >170kDa cross-linked species band from pZCC:PBD BS^3-cross linking experiments. Data indicate presence of only Rad50 intra cross-linked proteoforms (top table), however single BS^3-modified PLK1 species are also present (bottom table). (PSM: Peptide-Spectrum Match).

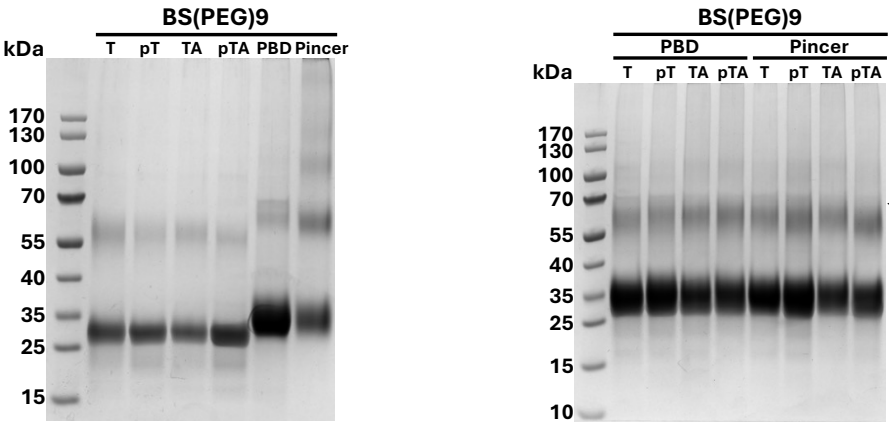


Figure 4.28 - SDS-PAGE analysis of analytical SEC from BS(PEG)9 cross-linking experiments suggests that pZCC and PBD may cross-link independently. SDS-PAGE analysis from BS(PEG)9 experiments of individual ZCC (T), pZCC (pT), ZCC_{T690A} (TA), $pZCC_{T690A}$, PBD, and Pincer proteins (left panel) and combinations (right panel) are indicated. Representative gels from three independent experiments.

4.3.11 ZCC Binding - Site-Specific Crosslinking Using Amber Codon Suppression

As a last attempt at non-amine-based cross-link analysis, we sought site-specific incorporation of the highly reactive unnatural lysine-derivative amino acid, N^6-[[2-(3-Methyl-3*H*-diazirin-3-yl)ethoxy]carbonyl]-L-lysine (AbK) into PBD. This was reasoned by the fact that upon exposure to UV light (330-370nm), the AbK diazirine adduct undergoes photolysis to form a highly reactive radical carbene intermediate (R-C:, where R-C depicts any atom bound to the carbon atom, and : denotes unpaired electrons on the carbon atom, respectively) (*Chen and Topp, 2019*). Consequently, the reactive carbene could form covalent bonds with molecules within a range of ~10-15Å of PBD-specific sites, through electrophilic insertion into C-C, C=C, C=O, C=N, and C=S nucleophilic bonds (*Chen and Topp, 2019*).

To attempt purification of AbK incorporated PBD, the *PBD* and *Pincer* genes were cloned in-frame of N-terminal MBP tag, subsequent 3C protease cleavage site, and C-terminal His_6 tag into a custom pCDF-PylT derivative vector. Both vectors, in conjunction with pBK-pylRS were constructed for *Methanosarcina barkeri* pyrrolysl-tRNA synthetase and pyrrolysine-$tRNA_{CUA}$-mediated site-specific incorporation of AbK at amber stop codons (*Polycarpo et al., 2004*). Based on the chemical properties of the diazirine side chain, AbK site-specific incorporation into interface-oriented K or R residues within PBD and Pincer could provide a proof-of concept basis for the identification of novel transient phosphorylation-dependent PBD interactors, such as pZCC. Moreover, the concept could conceivably be applied to any structurally characterised phospho-binding module, which could lead to the redefinition of pre-existing phosphorylation-dependent interaction consensus sequences.

All of the following analyses were structurally guided using an existing crystal structure of PBD bound to its optimal phosphopeptide (accession code: 1UMW) (*Elia et al., 2003*). R and K residue sites were selected based on relative structural similarity to AbK, in order to minimally perturb recombinant PBD folding and stability. Moreover, potential sites were then filtered if the sites were involved in intramolecular interactions, since disruption of these may also affect PBD folding and stability. The interface-oriented criterion was reasoned two-fold. Firstly, interface-oriented residues could potentially covalently capture transient and stable phosphorylation-dependent PBD interactors- within a range of ~10-15Å - from cells treated with DNA-damage and cell cycle arrest agents. Secondly, interface-oriented residues could minimise intramolecular AbK reactivity, thus reducing unintended structurally rigidity and protein instability. Following these analyses, the R557 residue was selected as the most ideal candidate, because the Cα atom is located 11.0Å and 10.5Å away from the phosphoserine γ-carbon phosphate ester linkage and the β-carbon alkane bond, respectively (Figure 4.29A).

Initial expression and purification trials of recombinant MBP-3C-PBD$_{AbK577}$-His$_6$ yielded a heterogeneous protein population of two proteins that possessed molecular weights of ~50kDa and ~65-70kDa, respectively. The ~65-70kDa protein roughly matched the corresponding predicted molecular weight of 70.3kDa, however the predicted molecular weight of a non-incorporated truncated PBD protein, i.e. MBP-3C-PBD$_{557*}$, would possess a similar predicted molecular weight of 63.8kDa (Figure 4.29B). A truncated non-incorporated protein should be included for future experiments, to determine the identity of the ~65-70KDa species, however due to time constraints this was not possible. Moreover, mass spectrometry analyses of the ~65-70kDa band only identified two PBD-specific peptides corresponding to residues 414-420 and 442-456, respectively, which

preceded the AbK557 incorporation site. As a result, the successful incorporation of AbK at PBD-R557 could not be confirmed, and therefore, further cross-linking analysis was not pursued.

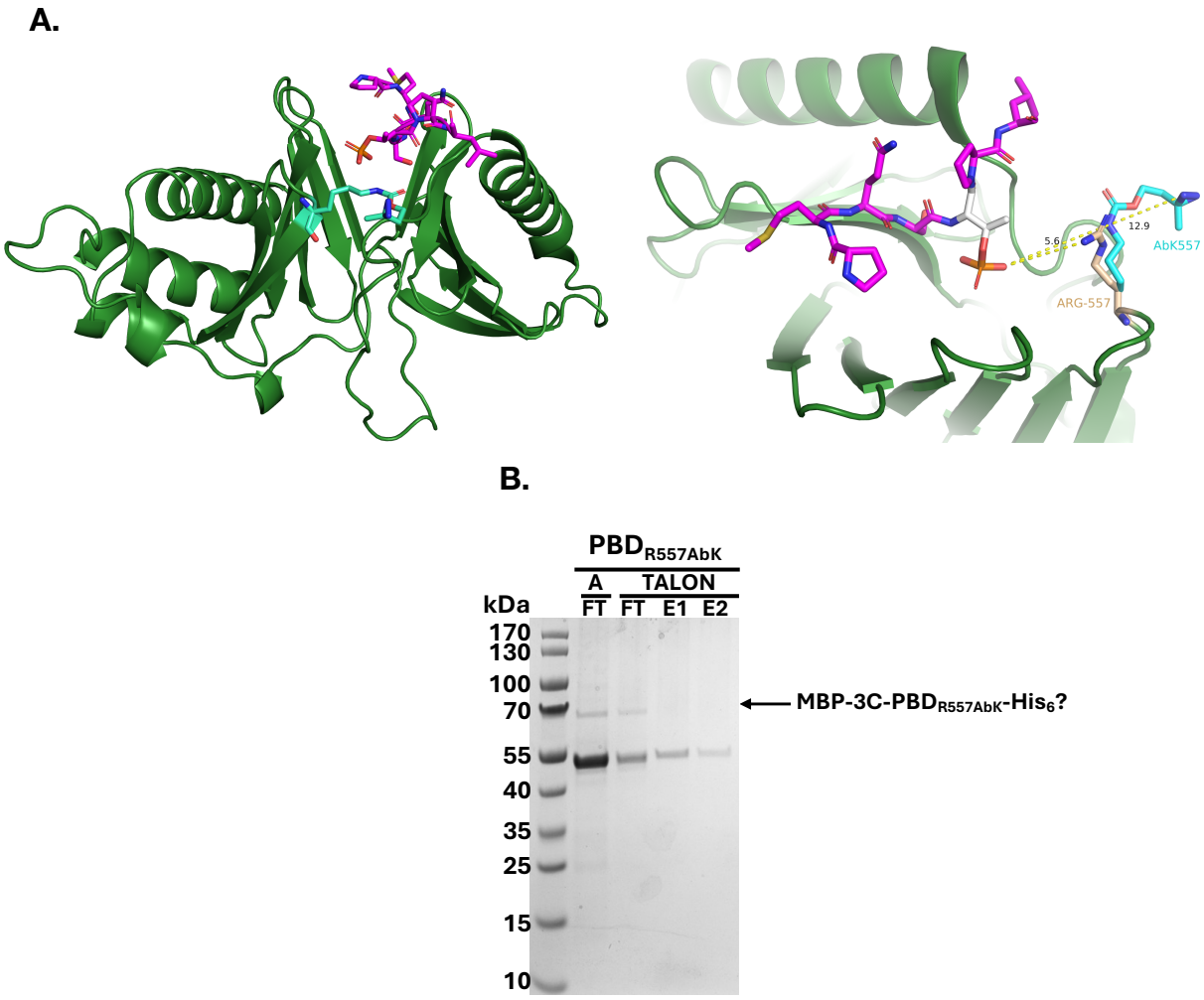


Figure 4.29 - Structure-guided design for site-specific incorporation of the AbK photoactivated cross-linker at the PBD phospho-ligand binding surface. (A) In silico insertion of the AbK residue (cyan) at PBDR557 using the crystal structure of PBD bound to its optimal phosphopeptide. Atoms comprising the AbK residue are represented in the stick format and coloured in cyan. A view of the overall structure (left panel) and the PBDtide-binding surface with comparison to Arg557 (wheat; right panel) are provided. The dashed lines indicate the distance (Å) between each terminal sidechain group and the O2 γ-phosphate oxygen atom. (B) SDS-PAGE analysis of an initial MBP-3C-PBDR557-His6 purification trial. Amylose (A) and TALON affinity chromatography flow through (FT) and elutions (E) are indicated. Protein bands corresponding to the predicted full-length molecular weight (~70.3kDa) are indicated.

4.4 Discussion

In summary, we have tried to prove extensively the phosphorylation-dependent interaction between ZCC-T690 and PBD with no reproducible success at the peptide or protein level. Initial MST binding experiments consistently detected a Pincer motif-dependent interaction with our biotin-pT690 phosphopeptide (Figure 4.7). We successfully produced 4 crystal structures of unbound PBD, demonstrating that our PBD was correctly folded (Figure 4.8), while ITC experiments indicated that PBD could interact with an optimal phosphopeptide as a functional phospho-binding module (Figure 4.11). Importantly, X-ray crystallography structural analyses (Figure 4.8) and ITC binding experiments with two phosphopeptide variants did not substantiate the MST phosphopeptide binding data (Figure 4.11). Moreover, AlphaFold predictions using both AF-M and AF3 models could not form a consensus about the potential binding interface between a 9-mer pT690 phosphopeptide and PBD, despite supposed high interface accuracy confidence metrics (Figures 4.9-4.10). Consequently, these data raised questions about whether buffer mismatch in our MST binding was responsible for erroneous false positive MTS fluorescence signals. These extensive biochemical and structural analyses, of course, suggest that PBD does not interact with pT690 phosphopeptide under any circumstance. Clearly, this data is at odds with the preliminary data generated by *Dullovi et al.,* and previous Smerdon Lab members that suggests the interaction does occur (*Dullovi et al., 2023*). The disparity between our study and the preliminary data could be explained due to two major differences. Firstly, the biotin-pT690 phosphopeptide used previously contained a 5-amino acid N-terminal extension, which contained the highly conserved CPVC motif, which chelates Zn^{2+} during ZCC dimerization (*Park et al., 2017; Dullovi et al., 2023*). In turn, this would suggest that the interaction between PBD and ZCC-pT690 is not solely phosphorylation dependent, but that an intact

zinc-hook mediates the interaction by an unknown mechanism. Secondly, the preliminary peptide-pull down and SPR experiments were underpinned by immobilisation of the extended biotin-pT690 phosphopeptide. Perhaps, it is possible that immobilisation of the extended biotin-pT690 phosphopeptide is necessary to induce a specific conformational state that elicits binding with PBD, but this remains to be investigated further using an identical phosphopeptide. It is also equally possible that either the newly synthesised phosphopeptides are defective in phosphorylation efficiency, stability, and solvent solubility. These possibilities at least provide some explanation for the unexplained ambiguity between both studies. Despite all of these issues, we still wanted to compare the previous peptide binding data with intact phosphoprotein.

In the second part of this study, phosphothreonine was site-specifically incorporated at the ZCC-T690 position. This was verified by Western blotting and mass spectrometry analyses (Figures 4.16-4.17), which confirmed successful incorporation, however phos-tag SDS-PAGE based quantification suggested that 48-67% of the protein was phosphorylated due to T, S, pS, and G misincorporation events (Figures 4.18-4.21) . Moreover, it was not possible to determine the relative proportion of pT or pS residues incorporated at the ZCC-690 position. Further optimisation of the pT OTS is clearly required to improve pT incorporation fidelity, and overall pT protein yield. In the first instance, phosphorylation-enrichment strategies should be employed to isolate only recombinant ZCC-pS/pT690, thereby reducing the proportion of non-phosphorylated ZCC-690 contaminants. The overall yield of recombinant ZCC phosphoprotein could be improved by two main strategies. For one, a systems wide analysis of the effects of pT OTS overexpression on global host cell fitness, metabolic resource allocation, and proteome dysregulation may identify components of the pT OTS that could be altered to reduce off-target cytotoxic effects (*Mohler et al., 2023*). This is a particularly

crucial avenue of research, because the pT bacterial strains were difficult to transform and exhibited extremely slow growth pre- and post-induction, which is indicative of poor host cell fitness. Additionally, screening of IPTG-inducible expression vectors, particularly those compatible with the pT OTS, such as the p15a origin-containing pRBC vector, could enhance the efficiency of screening autoinduction media expression systems (*Zhu, Mehl, and Cooley, 2022*). While this approach may not increase overall pT-OTS fidelity, the potential increased biomass could lead to higher pT-incorporated protein yields. In any case, we were able to conduct single MST binding experiments with our MBP-3C-ZCC-pT690-His_6 protein, which did not detect any interaction with PBD proteins (Figure 4.22). This appeared to address the apparent dependency of an intact zinc-hook proposed from the preliminary experiments, but these data were not conclusive without further experimental replicates. We then turned to MST binding experiments involving non-phosphorylated and CK2-phosphorylated MBP-3C-ZCC-His_6 and MBP-3C-pZCC-T690A-His_6 with PBD proteins either in the absence or presence of the Zn^{2+} chelating agent, EDTA (Figure 4.23). These experiments appeared to suggest that PBD and Pincer interact with only T690-containing pZCC proteins in a phosphorylation and Zn^{2+} manner. This was then pursued using analytical SEC and amine-based cross-linking experiments, which failed to substantiate the MST binding data (Figures 4.24-4.28). Moreover, the cross-linking experiments and mass spectrometry analysis suggested that ZCC oligomers were exclusively cross-linked (Figures 4.25-4.26). The identification of single BS^3-modified PBD peptides was particularly puzzling but remains unresolved (Tables 4.4-4.5). Finally, we attempted purification of recombinant PBD with the unnatural photoactivatable cross-linker, AbK, incorporated at R557 for amine-independent cross-linker analysis and identification of novel transient phosphorylation-dependent PBD interactors (Figure 4.29). The selection of this site was based on a pre-existing structure of PBD bound to its optimal phosphopeptide; however, mass

spectrometry-based validation of our recombinant artificial protein was unsuccessful (*Elia et al., 2003*). Ultimately, these analyses did not provide any conclusive evidence of a phosphorylation-dependent interaction between PBD and ZCC-pT690 at the protein level. Furthermore, we could not provide any evidence that phosphorylation of T690 impacts the oligomeric or conformational state of ZCC.

4.4.1 Future experiments

The impact of phosphorylation of T690 in the zinc-hook domain has previously focused on changes in Zn^{2+}-binding affinity, as well as on local and global conformational changes affecting ZCC homodimer stability (*Tran, Padjasek, and Krężel, 2022*). These analyses have shown that phosphorylation decreases Zn^{2+} binding by 4-fold, which correlates with minor conformational changes localised to the conserved CPVC Zn^{2+}-binding motif (*Tran, Padjasek, and Krężel, 2022*). Importantly, the Zn^{2+} binding affinity of the zinc-hook motif is in the femtomolar range, therefore a phosphorylation-dependent 4-fold decrease in Zn^{2+}affinity is unlikely to significantly impact Zn^{2+}-mediated dimerization in a cellular context (*Tran, Padjasek, and Krężel, 2022*). Additionally, the effects of T690 phosphorylation were diminished in the full-length ZCC domain. In our study we have explored a potential novel interaction between ZCC-pT690 and PBD at the phosphopeptide and phosphoprotein level. Despite extensive using methods such as MST, ITC, x-ray crystallography, SEC, and cross-linkers, we have not been able to successfully confirm a direct interaction. Consequently, the downstream effects of ZCC-T690 phosphorylation remain unclear.

While it is conceivable that T690 phosphorylation is functionally redundant, this seems unlikely due to the conservation of this site throughout mammalian species. The importance of T690

phosphorylation is further supported by proteomic studies that have shown that phosphorylation of T690 is triggered by DNA damage following ionising-radiation, and upon mitotic entry (*Bennetzen et al., 2010; Olsen et al., 2010*). It seems crucial, therefore, that we gain an understanding of whether ablation or mimicry of this phosphorylation event has any downstream cellular consequences, particularly within the context of MRN assembly and functionality, the DDR, and mitosis.

To address these points in a cellular context, we propose that future work is focused on generation of stable cell lines that overexpress exogenous Rad50 wild-type, T690A phosphonull, and T690E phosphomimetic protein, while endogenous Rad50 is targeted by siRNA-mediated knock-down. To that end, we have successfully cloned the full-length *Rad50* gene and phosphonull/mimetic mutants in-frame with an N-terminal GFP tag into the mammalian pcDNA5-FRT-TO Flp-In genome integration vector. We have also identified an siRNA that specifically targets endogenous Rad50 protein, but at present our exogenous expression vector is not siRNA resistant. Consequently, further experiments should focus on generation of silent mutations in the siRNA-targeted region, so that the exogenous GFP-Rad50 mRNA is not knocked down. Following this, it is important to establish the relative MRN incorporation efficiency of phosphomutant T690 proteins, in order to understand whether T690 phosphorylation has any effect on MRN complex formation. Although previous structural analyses suggest that T690 phosphorylation affects local conformational changes, it may be necessary to understand whether T690 phosphorylation has an allosteric effect on Rad50 ATPase activity and overall Rad50 conformation. This could be addressed by purification of exogenous proteins using GFP-trap technology coupled with ADP-Glo ATPase assays in the absence or presence of DNA. Once these control experiments have been addressed, it is necessary to elucidate whether T690 phosphorylation impacts DNA damage resolution following ionising

radiation through monitoring of γH2AX clearance rates. Similarly, mitotic duration could be monitored by live-cell fluorescence microscopy or by Western blot analysis using markers of mitotic entry, progression, and exit, such as APC3 hyperphosphorylation status, cyclin B1 degradation, and PLK1 degradation, respectively. Both sets of experiments should also be accompanied with CK2-substrate Western blot analysis and mass spectrometry analysis to determine the phosphorylation status of T690.

From a biochemical perspective, it seems clear that using the recombinant ZCC with a pT inserted at the T690 position could yield valuable insights. We believe that our recombinant ZCC-pT690 offers a more physiologically relevant approximation of endogenously phosphorylated Rad50-T690 compared to the synthetic phosphopeptide counterpart. As such, it could serve as a powerful tool for the identification of pT690-dependent interactors, provided that the preliminary proteomic analyses are repeated. Ultimately, identification of any novel pT690-dependent interactor would likely inform subsequent cellular experiments, while cellular experiments are likely to provide context for understanding the role of these interactions.

An avenue to biochemically explore that would provide an explanation for the original Rad50 phospho-interactomics data is that the interaction between Rad50 and PLK1 occurs indirectly and involves unidentified mediator protein(s). One such candidate as a mediator is FOXK1, which was identified as the 5th highest-ranking protein in the screen (*Dullovi et al., 2023*). FOXK1 is one of two members of the forkhead box (FOX) transcription factor subfamily K (FOXK1 and FOXK2), which both uniquely contain FHA domains (*Katoh and Katoh, 2004*). FHA domains crucially select for pThr-containing epitopes, and while the Rad50-pT690 fits this prerequisite there is also selectivity at the +3 position. The Rad50-pT690 +3 is a E residue, which has not been conclusively shown to interact

with FHA domains, however the FOXK1 FHA domain interaction sequence is currently unexplored. Moreover, previous evidence has demonstrated that PLK1-dependent phosphorylation of JNK-associated leucine zipper protein (JLP)-T351 regulates FOXK1-FHA-dependent binding to JLP-T351 (*Ramkumar et al., 2015*). Moreover, circumstantial immunoprecipitation evidence suggests that PLK1 interacts directly with a FOXK1 fragment corresponding to amino acids 195-300. Interestingly, this region encompasses the FOXK1 DNA-binding domain (DBD) and contains 4 potential Ser and Thr that fit the minimal PBD S-S/T interaction consensus sequence: T195 (S**T**A), S264 (S**S**S), S265 (S**S**Y), and S320 (S**S**A). However, there is currently no phosphoproteomics evidence to support phosphorylation of any these sites. Initial experiments could focus on exploring the potential phosphorylation-dependent interaction between Rad50-pT690 and the FOXK1-FHA, as well as the potential phosphorylation-dependent interaction between PBD and FOXK1-T195/S264/S265/S320.

Chapter 5

Biochemical Characterisation of a SUMO E1 Phosphorylation Site Induced in Mitosis and the DNA Damage Response

5.1 Introduction

Small ubiquitin-like modifiers (SUMOs) are an evolutionarily conserved family of ~11kDa (~100 amino acids) post-translational modifier proteins that dynamically and reversibly modify >6000 human proteins in response to changes within the intra- and extracellular environment. In humans, at least 4 *SUMO* genes are ubiquitously expressed as C-terminally extended type I UbL preproteins (SUMO1-4) (*Mannen et al., 1996; Matunas, Coutavas, and Blobel, 1996; Lapenta et al., 1997; Mahajan et al., 1997; Kamitani et al., 1998; Guo et al., 2004*). A fifth gene (*SUMO5/ SUMO1P1*) is expressed tissue-specifically but is generally considered a non-functional pseudogene (*Bouchard et al., 2021*).

The attachment of sentrins/SUMOs is central to the regulation of virtually all nuclear processes, including the DDR and DNA repair pathways *(Zhu et al., 2008; Morris et al., 2009; Gali et al., 2012; Luo et al., 2012; Garvin and Morris, 2017)*. Like ubiquitination, SUMOylation involves formation of an isopeptide linkage between a SUMO C-terminal glycine carboxyl group and a substrate lysyl residue ε-NH_2 group *(Matunas, Coutavas, and Blobel, 1996; Kamitani, Nguyen, and Yeh, 1997; Kamitani et al., 1998))*. SUMOylation is primarily a non-proteolytic PTM that regulates substrate localisation, stability, and functionality by forming non-covalent binding surfaces for SUMO-binding motif (SBM)-containing interactors *(Song et al., 2004)*. Although SUMOylation is functionally distinct from ubiquitination, in some cases both pathways act in concert, as SUMO-modified substrates can prime SUMO-targeted ubiquitin ligase (STUbL)-mediated substrate degradation *(Uzunova et al., 2007; Tatham et al., 2008)*. The hierarchical SUMOylation cascades are similar to those of ubiquitination, consisting of sequential SUMO-specific E1, E2, and E3 enzymes (Figure 5.1). First, a pool of conjugatable SUMO molecules containing C-terminally exposed GG (diGly) motifs, termed

matured SUMOs, are generated through the catalytic activities of SUMO maturation and deconjugation enzymes *(Gong et al., 2000; Hang and Dasso, 2002; Mikolajczyk et al., 2007)*. Specifically, canonical SUMO1-3 members are the only paralogues that are enzymatically matured and deconjugated, therefore, are the only conjugatable SUMO members *(Gong et al., 2000)*. In contrast, the maturation and conjugation capacity of SUMO4 is controversial, due to sequence variation within the C-terminal tail, which renders SUMO4 incompatible with all SUMO processing enzymes *(Owerbach et al., 2005)*. Mature SUMOs are activated through ATP-dependent adenylation and thioesterification reactions by the SUMO-activating enzyme (Sae) E1 heterodimer, composed of Sae1/AOS1 and Sae2/Uba2. SUMOs are then transferred from a Sae1:Sae2-linked high-energy thioester bond to the only known SUMO E2, Ubc9/UbE2I, catalytic cysteine active site through a transfer reaction, termed transthioesterification *(Yasugi and Howley, 1996; Johnson and Blobel, 1997; Gong et al., 1999)*. Ubc9 can then facilitate direct SUMO ligation E3-independently to substrate lysine acceptor sites, which reside in at least one of five SUMOylation consensus motifs (SCMs), including the canonical SCM: ϕ-K-x-D/E, where ϕ represents a hydrophobic residue (C/F/I/L/M/P/V) and x denotes any amino acid *(Rodiguez, Dargemont, and Hay, 2001; Tatham et al., 2001)*. Other SCM variants include D/E-x-K-ϕ inverted SCMs (ISCMs), ϕ-ϕ-ϕ-K-x-D/E hydrophobic cluster SUMOylation motifs (HCSMs), ϕ-K-x-D/E-x-x-D/E-D/E-D/E-D/E negatively charged amino acid-dependent SUMOylation motifs (NDSMs), and ϕ-K-x-E-x-x-pS-P phosphorylation-dependent SUMOylation motifs (PDSMs) *(Hietakangas et al., 2006; Yang et al., 2006; Matic et al., 2010)*. There are approximately 14 SUMO E3 enzymes that bind to both thioester-linked Ubc9~SUMO and the target substrate *(Shi et al., 2022)*. SUMO E3s confer substrate and SUMO isoform specificity, as well as enhance the catalytic efficiency of Ubc9 through promotion of Ubc9~SUMO closed-state conformers in a manner analogous to RING E3 ligases *(Pruneda et al., 2012)*.

Like Ub, SUMO can form a diverse array of attachment topologies on modified substrates. These include single (monoSUMOylation) or multiple (multi-monoSUMOylation) monomeric attachments. Notably, monoSUMOylation is the predominant topology observed in cultured cells. Additionally, SUMOylation E1-E2-(E3) cascades can repeat through the sequential attachment of SUMO to internal lysine residue acceptor sites on pre-existing SUMO moieties to form homomeric or mixed polymeric SUMO chains (polySUMOylation). In SUMO1, 10 lysine residues (K7, K16, K17, K23, K25, K37, K39, K45, K48, and K78) serve as SUMO acceptor sites, however the SUMO1-K7 D/E-x-K-ϕ inverted SCM is the primary SUMO2/3 acceptor site under unstressed conditions (*Hendriks et al., 2017*). Under normal conditions, over 90% of polySUMO chains are linked by SUMO2/3. SUMO2/3 chains form through 7 lysine acceptor sites (K5, K7, K11, K21/20, K33/32, K35/34, and K42/41), however the canonical SCM-containing SUMO2/3-K11 is the predominant acceptor site in polySUMO2/3 chains (*Hendriks et al., 2017*). Furthermore, polymeric SUMO chains can also incorporate Ub and NEDD8 molecules, resulting in complex branched heterotypic chains that finetune SUMOylation cascades through conferring a further layer of divergent substrate specificity (*Hendriks et al., 2017*).

Taken together, SUMO conjugation topologies create complex and versatile barcodes that determines a diverse array of cellular outcomes. For example, p53 monoSUMO1ylation at K386 inhibits p53 DNA-binding and subsequent transcriptional activation, while monoSUMO1ylation of cytoplasmic Ran-GTPase (RanGAP1)-K524 triggers its relocalisation to nuclear pore complexes (NPCs) and subsequent interaction with the Nup358/RanBP2 SUMO E3 ligase *(Bernier-Villamor et al., 2002; Wu and Chiang, 2009)*. On the other hand, early mitotic topoisomerase IIα (Top2a) multi-monoSUMO2/3ylation triggers centromeric recruitment of Top2a and the chromosomal passenger

complex (CPC), while inhibiting Top2a DNA decatenation activity until onset of anaphase *(Mao, Desai, and Liu, 2000; Agostinho et al., 2008)*. Additionally, DSB signalling-dependent MDC1 polySUMOylation regulates MDC1 turnover through triggering recruitment of the SUMO-targeted Ub ligase, RNF4, which polyubiquitinates MDC1, thus targeting it for 26S proteasomal degradation (*Wu and Chiang, 2009; Luo et al., 2012*).

Here, we provide a comprehensive mechanistic and structural overview of how the heterodimeric SUMO E1 complex orchestrates SUMO activation and Ubc9-mediated transthioesterification at the molecular level. Despite extensive research into the SUMOylation cascade, whether other post-translational modifications, such as phosphorylation, influence SUMOylation cascade proficiency remains poorly understood. Consequently, we also explore the possibility of a single Sae1 phosphosite directly affecting SUMO E1 catalysis through the use of phosphorylated recombinant Sae1:Sae2 heterodimers in *in vitro* reconstituted RanGAP1 SUMOylation assays.

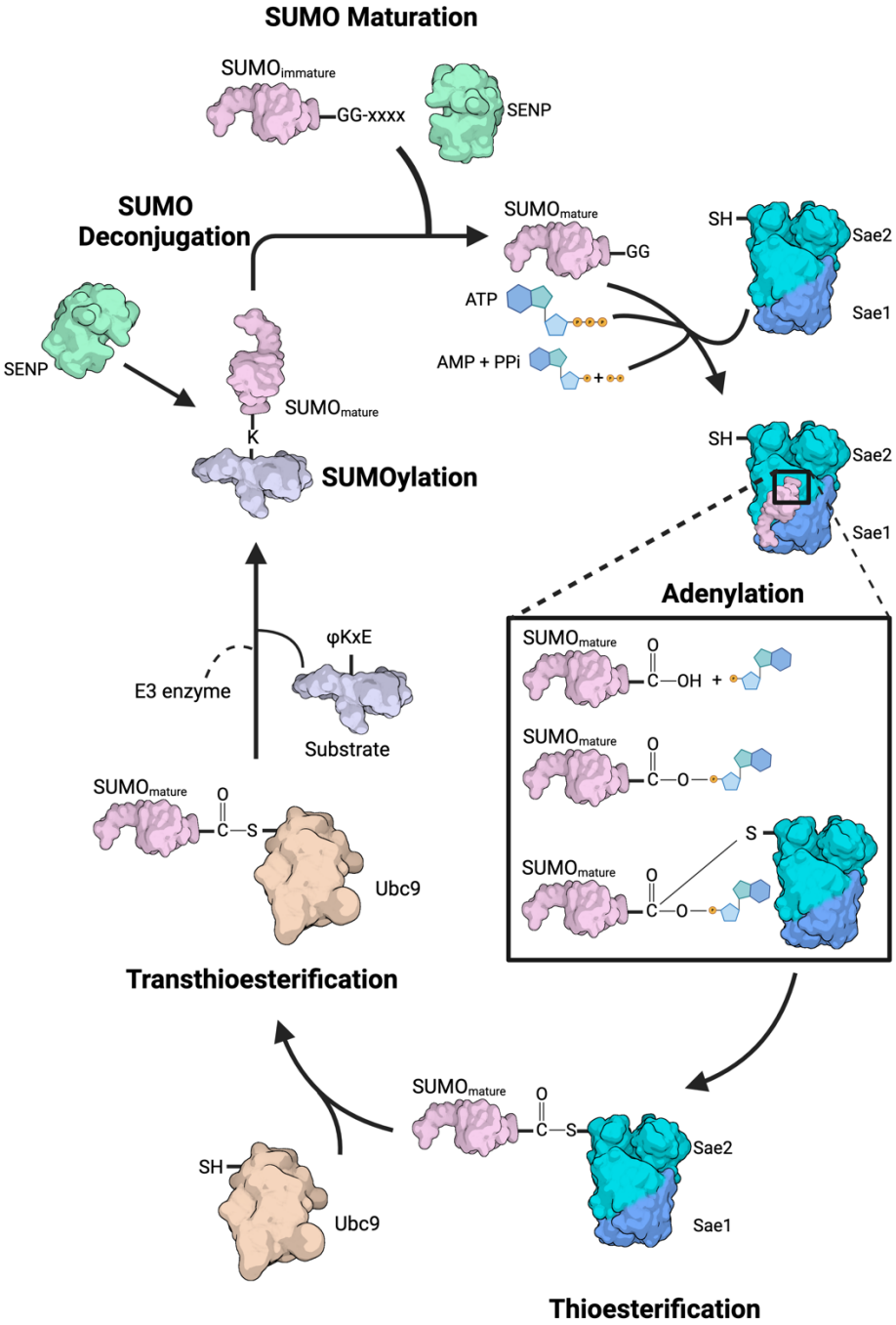


Figure 5.1 - The SUMOylation cascade. The solvent-accessible surface of SUMO 1 (pink; accession code: 1A5R), Sae1:Sae2 (light and dark blue; accession code: 1Y8Q), SENP1 (green; accession code: 2G4D), Ubc9 (orange; accession code: 1U9A), and RanGAP1 (purple; accession code: 2IO2) are represented in the cartoon format. The SUMO adenylate and SUMO:adenylate:SUMO E1 tetrahedral intermediate are shown in the inlet box. Figure created with Biorender.com.

5.1.1 SUMO Isoforms

SUMO proteins are conserved throughout eukaryotes, with *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans* expressing a single SUMO1 orthologue, Smt3, while *Xenopus laevis* express 5*, Arabidopsis thaeliana* 8*,* and mammals express at least 4 isoforms, respectively (*Meluh and Koshland, 1995; Mannen et al., 1996; Matunas, Coutavas, and Blobel, 1996; Choudhury and Li, 1997; Lapenta et al., 1997; Mahajan et al., 1997; Huang et al., 1998; Kamitani et al., 1998; Saitoh et al., 1998; Kurepa et al., 1999; Guo et al., 2004*). As previously mentioned, humans encode for 5 SUMO paralogues, SUMO1-5 (*Mannen et al., 1996; Lapenta et al., 1997; Guo et al., 2004; Liang et al., 2016*). SUMO1-5 paralogues share a moderate level of sequence conservation (Figure 5.2), with pre-SUMO1 sharing 47%, 47%, and 41%, and 88% sequence identity with pre-SUMO2/3/4/5, respectively. SUMO1-3 are considered the canonical SUMO isoforms due to high expression across all human tissues, and are also the only paralogues that are subject to sentrin/SUMO-specific protease (SENP) family-mediated maturation, deconjugation, required for inherent SUMO conjugation activities (*Gong et al., 2000; Owerbach et al., 2005; Mikolajczyk et al., 2007*; *Bouchard et al., 2021*). Pre-SUMO2 and pre-SUMO3 are the most closely related family members, sharing 97% identity, which cause both isoforms to be indistinguishable by antibody probes and exhibit similar substrate preferences. As such, both isoforms are collectively referred to as SUMO2/3. The majority of SUMO2/3 sequence divergence is derived from a VY/VP dipeptide variation immediately following the diGly motif, and 8-amino acid C-terminal extension in SUMO3, providing SUMO2 with more proficient Sentrin/SUMO-specific proteases (SENP)-mediated maturation capacity (*Xu and Au, 2005; Mikolajczyk et al., 2007*). Non-canonical pre-SUMO4 shares 85% and 84% sequence identity with pre-SUMO2 and pre-SUMO3, respectively, however a unique

SUMO4-P90 residue preceding the diGly motif, instead of a highly conserved Q residue (SUMO1-Q94, SUMO2-Q90, SUMO3-Q89, and SUMO5-Q94), limits SENP-mediated SUMO4 maturation and conjugation capacity (*Owerbach et al., 2005*). In contrast, recent evidence suggests that immature SUMO4 forms non-covalent interactions with SENP1, to promote SENP1 deconjugation activity, and prevent excessive global SUMOylation dynamics and BRCA1-A complex recruitment during DSB signalling (*Preprint; Garvin et al., 2024*). Primate-specific pre-SUMO5 shares a high level of sequence identity with pre-SUMO1, however *SUMO5* is considered a non-functional *SUMO1* pseudogene, because of its low expression is predominantly limited to the peripheral blood leukocytes and testes (*Liang et al., 2016; Bouchard et al., 2021*). In spite of this, promyelocytic leukemia protein (PML)-K160 SUMO5ylation was identified as a crucial regulator of PML nuclear body nucleation and disassembly dynamics (*Liang et al., 2016*).

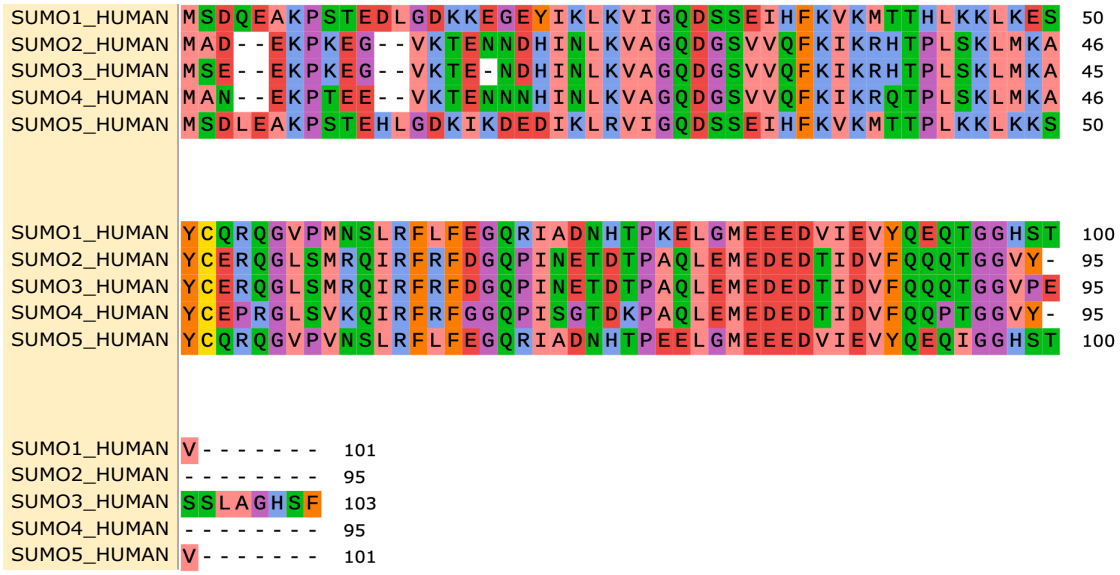


Figure 5.2 - Sequence conservation of SUMO family members. The primary sequence of human SUMO 1-5 proteins was aligned using the Clustal Omega algorithm (Sievers et al., 2011). Amino acids are coloured according to physicochemical properties (red- acidic, blue- basic, green- polar, salmon- hydrophobic, orange- aromatic, purple- conformationally special, and yellow- cysteine). Absolutely conserved amino acids are denoted with an asterisk (*). Amino acids with similar properties are indicated with a colon (:), and weakly similar amino acids are indicated with a full stop (.). Figure created with SnapGene software v8.0 (www.snapgene.com).

5.1.2 Selectivity Amongst Canonical SUMO1-3 Paralogues

At a glance, canonical SUMO1-3 isoforms exhibit overlapping localisation patterns and target substrates, as well as similar SENP1-mediated maturation and deconjugation efficiency, suggesting that canonical isoforms can compensate for one another to maintain global SUMOylation dynamics (*Ayaydin and Dasso, 2004; Vertegaal et al., 2006*). For example, all 3 members localise throughout the nucleoplasm and to PML bodies, and can SUMOylate both PML and MDC1, while *SUMO1* knockout mice are viable due to functional compensation of canonical RanGAP1 SUMO1ylation by SUMO2/3ylation (*Ayaydin and Dass, 2004; Evdokimov et al., 2008; Luo et al., 2012*).

However, SUMO1-3 do indeed possess distinct tissue-dependent expression levels, cellular localisation patterns and dynamics, as well as divergent phenotypes upon CRISP-Cas9-mediated knockout (*Saitoh and Hinchey, 2000; Ayaydin and Dasso, 2004; Bouchard et al., 2021*). For example, *SUMO1* expression is higher than *SUMO2* in the liver, adrenal gland, and muscle, whereas *SUMO2* expression levels are typically higher in reproductive organs, while *SUMO3* is expressed at a higher level than *SUMO1* and *SUMO2* in the majority of human tissues. In addition, SUMO1 localises to cytoplasmic foci, as well as the nuclear envelope and nucleolus, whereas SUMO2/3 localises throughout the nucleoplasm. Furthermore, SUMO1 localises to the mitotic spindle during metaphase and anaphase, whereas SUMO2/3 does not (*Ayadin and Dasso, 2004*). The *in vitro* SUMO2/3 pool is greater than SUMO1, such that the unconjugated SUMO2/3 pool is ~50-fold greater than that of free SUMO1 under normal conditions, wherein 90% of SUMO1 is conjugated (*Saitoh and Hinchey, 2000*). SUMO1 conjugate proportions are also less dynamic than SUMO2/3 conjugates, with SUMO1 unperturbed after 10 minutes of heat stress, while SUMO2/3 conjugates increase to ~80% of the total SUMO2/3 pool (*Ayaydin and Dasso, 2004*). This dynamic SUMO2/3 stress

response was further solidified by nucleoplasmic FRAP experiments, whereby YFP-pre-SUMO1 chimeras recovered 10-fold slower than YFP-pre-SUMO2/3 counterparts (*Ayadin and Dasso, 2004*). Moreover, CRISPR-Cas9 *SUMO* knockout cell lines demonstrated that *SUMO2* knockout cells display a unique altered cellular morphology and enrichment of cells with >2N DNA content (*Bouchard et al., 2021*). These phenotypes were associated with differentially expressed patterns between *SUMO* knockout cell lines, and a unique role of SUMO2 in promoting cell survival upon hydroxyurea-induced DNA replication stress (*Bouchard et al., 2021*). Collectively, these data indicate that canonical SUMO paralogues exhibit non-redundant biological activities, despite conflicting evidence from mouse knockout models.

Additionally, a significant proportion of SENPs, SUMO E3s SUMO substrates and downstream SUMO-dependent interactors are paralog-specific. For example, SENP2/3/5/6/7 and USPL1 exhibit preferential SUMO2/3 deconjugation/isopeptidase activity, with SENP7 only acting on SUMO2/3. In particular, SENP2/SENP5 and USPL1 also exhibit preferential SUMO2/3 maturation/endopeptidase efficiency (*Mikolajczyk et al., 2007; Kolli et al., 2010; Schulz et al., 2012*). SUMO E3 paralogue specificity includes the preferential SUMO1-binding, and enhanced E3 activity and protection of RanGAP1-SUMO1 from SENP1-mediated deconjugation imparted by the RanBP2/Nup358 nuclear pore complex subunit, as well as the SUMO2/3 preferential activity of ZNF451 (*Reverter and Lima, 2005; Gareau, Reverter, and Lima, 2012; Cappadocia, Pichler, and Lima, 2015*). RanBP2 achieves paralogue specificity through an internal repeat domain, IR1-M-IR2 that contains two internal SUMO-interaction motifs (SIMs), of which the former generates more interactions and a higher level of structural complementarity with the hydrophobic groove of SUMO1 compared to SUMO2 (*Reverter and Lima, 2005; Gareau, Reverter, and Lima, 2012*). Classic examples of paralogue-

selective substrates include SUMO1-preferential modification of RanGAP1 and proliferating cell nuclear antigen (PCNA), while polySUMO2/3 chains preferentially modify topoisomerase 2A (TOP2A), USP25, and BLM RecQ-helicase (*Matunas, Coutavas, and Blobel, 1996; Agostinho et al., 2008; Meulmeester et al., 2008; Zhu et al., 2008; Gali et al., 2012; Sakin et al., 2015*).

5.1.3 Conserved Structural Features of the SUMO-Fold Drive SUMOylation Cascade Functionality while Unique SUMO Features Drive Paralogue Specificity

All eukaryotic SUMO isoforms adopt a core β-grasp fold, consisting of a mixed 5-standed β-sheet (↓↑↑↓↑) that encircles a central α-helix and 3_{10}-helix, which is flanked by both intrinsically disordered and extended N- and C-termini (*Kraulis, 1991; Overington, 1992; Bayer et al., 1998; Burroughs et al., 2007*). In particular, the β-grasp core is a characteristic feature of the ubiquitin-like (UbL) superfamily across all domains of life. The C-terminal tail houses the diGly motif, which is exposed during sentrin/SUMO-specific protease (SENP)-mediated SUMO maturation and is the substrate lysine conjugation site, thereby providing the SUMO family with type I UbL status (*Kamitani, Nguyen, and Yeh, 1997; Bayer et al., 1998; Gong et al., 2000; Jentsch and Pyrowolakis, 2000*). On the other hand, the flexible 13-23 amino acid N-terminal extension is a unique feature present in only in SUMO and ATG12 families (*Bayer et al., 1998; Otomo et al., 2012*).

The β-grasp structural core contributes at least 2 major binding surfaces that mediate non-covalent interactions with all components of the SUMOylation cascade and modified substrate interactors (Figure 5.3). Firstly, the α-helical-opposed (Figure 5.3, right panel), or exposed surface of the β-sheet forms non-covalently interactions with SENPs, the Sae1:Sae2 SUMO E1 heterodimer, and Ubc9, wherein the overlapping surface is centred around the β_3-β_5 region, involving moderately

conserved SUMO1-R63, E67, G68, and E89 residues (SUMO2/3-R59/58, D63/62, G64/63, and D85/84) (*Lois and Lima, 2005; Reverter and Lima, 2005; Reverter and Lima, 2006; Capili and Lima, 2007; Olsen et al., 2010; Lussier-Price et al., 2020*). In particular, SUMO1-E89 residue acts as a crucial hydrogen bond acceptor with the majority of SUMOylation cascade components, interacting with the Sae2 adenylation domain residues N423, K425, the Ubc9 backside residue G23, and the SENP2 catalytic site-proximal residue K445 (*Lois and Lima, 2005; Reverter and Lima, 2006; Capili and Lima, 2007*). Consequently, the promiscuous nature of the E89 side chain enables flexible positioning of the SUMO C-terminal tail for SUMO maturation and deconjugation, activation events, but is also involved in SUMO-loading during polySUMO chain formation. This is further evidenced by a E89R charge reversal mutation that inhibits all of the aforementioned processes (*Knipscheer et al., 2007*).

The second binding surface is formed by the obscured β-grasp face within the β_2-α-helix region, which forms a conserved hydrophobic groove across SUMO paralogues (*Song et al., 2004*) (Figure 5.3, left panel). The hydrophobic groove facilitates weak affinity interactions with SIM-containing proteins, including Sae2, SUMO E3s, substrates, and SUMO-modified substrate interactors (*Reverter and Lima, 2005; Song et al., 2005; Sekiyama et al., 2008; Olsen et al., 2010; Gareau, Reverter, and Lima, 2012; Xu et al., 2014; Cappadocia et al., 2015; Anamika and Spyracopoulos, 2016; Lussier-Price et al., 2020; Lussier-Price et al., 2022*). In particular, the C-terminal Sae2 SIM, corresponding to residues 632-640, contacts the SUMO1 hydrophobic groove in the SUMO adenylated-mimic state, which may suggest that the SUMO E1 inhibits SIM-binding during SUMO activation to ensure unperturbed SUMO activation (*Olsen et al., 2010*). Additionally, myeloproliferative and mental retardation (MYM)-type tandem zinc fingers, such as those found in

ZNF198 and ZNF261 bind to the SUMO hydrophobic groove independently of zinc-binding status, and are required for localisation of ZNF198 to SUMO-enriched PML bodies (*Guzzo et al., 2014*). Putative SIMs are characterised by a hydrophobic core of 3-4 amino acids, which fit a loose consensus sequence of ϕ-ϕ-ϕ-ϕ, ϕ-ϕ-x-ϕ or ϕ-x-ϕ-ϕ, where ϕ denotes a I/V/L aliphatic amino acid and x is typically a D/E/S/T residue (*Song et al., 2004; Hannich et al., 2005; Hecker et al., 2006; Beauclair et al., 2015*). Upon binding, unstructured SIMs adopt an extended β-strand-like parallel or antiparallel conformations relative to SUMO β_2 (*Reverter and Lima, 2005; Song et al., 2005; Xu et al., 2014; Cappadocia et al., 2015*). The hydrophobic SIM core is often flanked either side by acidic or phosphorylated serine residues, which form electrostatic interactions with basic residues within SUMO hydrophobic groove and significantly enhance the affinity and stability of free SUMO/SUMO conjugate:SIM complexes (*Stehmeier and Muller, 2009; Chang et al., 2011*). Examples of phosphorylated-SIMs (pSIMs) include the protein inhibitor of activated STAT protein 1 (PIAS1) SUMO E3, as well as the DAXX transcription factor and PML substrates, and the SUMO-modified substrate interactor, BRCA1-A complex subunit RAP80 (*Cappadocia et al., 2015; Anamika and Spyracopoulos, 2016; Lussier-Price et al., 2020*). SUMO1 and SUMO2/3 exhibit slight structural variations in the basic residues that contact pSIMs, wherein SUMO1-K39-K46 structurally align to SUMO2-K35-K42, however SUMO1-H43 and SUMO2-H37 are dissimilar (*Anamika and Spyracopoulos, 2016*). This structural variation is proposed to be a contributing factor that drives divergent SUMO paralogue interaction specificity and affinity amongst pSIM-containing interactors.

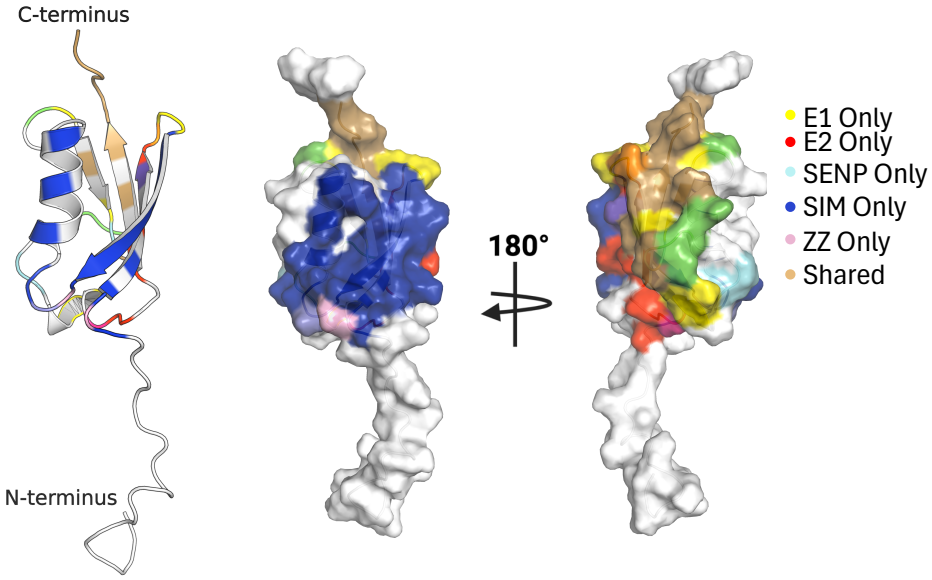


Figure 5.3 - Structural analysis of SUMO interaction surfaces. The secondary structure (left panel) of the full-length SUMO 1 NMR structure (accession code: 1A5R) is provided. The solvent-accessible surface of SUMO 1 is provided and rotated about the y-axis by 180°. Individual residues are coloured accorded to E1 (yellow), E2 (red), SENP (cyan), SIM (dark blue), ZZ (pink) unique interaction surfaces, and overlapping interaction surfaces (brown). Analysis was performed based on the following accession codes: 1Y8R, 3KYC, 3KYD, 1Z5S, 6V7P, and 2N1A.

The N-terminal extension is the most divergent feature of canonical SUMO paralogues and is the predominant initiation site for polymeric SUMO chain formation (*Hendriks et al., 2017; Hendriks et al., 2018*). In particular, over 90% of PolySUMO chains are initiated by SUMO2/3 under normal conditions, compared to the mere 6.8% formed by SUMO1 (*Hendriks et al., 2017*). Specifically, 35% of endogenous polySUMO2/3 chains are initiated by a SUMO2/3-specific K11, which fits a canonical ϕ-K-x-D/E SUMO-conjugation motif (SCM), where ϕ represents a hydrophobic residue (C/F/I/L/M/P/V) and x denotes any amino acid (*Rodriguez, Dargemont, and Hay, 2001; Tatham et al., 2001; Matic et al., 2010; Hendriks et al., 2018*). Moreover, the proportion of SUMO2/3-K11 chains rise to ~50% under heat and proteotoxic stress (*Hendriks et al., 2018*). In addition, SUMO2/3-K11 is a sirtuin 1 (SIRT1)-mediated acetylation target site, which inhibits canonical polySUMO2/3 chain formation and instead promotes formation of non-canonical SUMO2/3-K5 and K35-linked polymeric

chains (*Gärtner et al., 2018*). While SUMO1 does not contain the SUMO2/3-K11 SCM, 70% of SUMO1-primed polySUMO2/3 chains are attached through an inverted SCM (ISCM), D/E-x-K-ϕ, located at SUMO1-K7. In contrast to endogenous stress-induced polySUMO2/3-K11 dynamics, SUMO1-K7 primed polySUMO2/3 chains decrease upon heat and proteotoxic stress, and instead coincide with an increase in K17/23/37/45/78-linked SUMO1 chains (*Hendriks et al., 2018*). Interestingly, recent evidence has elucidated that the SUMO1 N-terminus acts as transient intramolecular competitive inhibitor of SIMs, whereas the SUMO2/3 N-terminal region does not (*Lussier-Price et al., 2022; preprint Richter et al., 2024*). Consequently, this finding provides a molecular insight into how different SUMO paralogues achieve SIM-specific interactions with substrates or effector proteins, such as RanBP2, USP25, Daxx, and PML. These findings, together with the earlier mentioned SUMO paralogue expression and localisation data, provide an initial framework for understanding the preferential paralogue-specific modification of substrates proteins and the subsequent recruitment of paralogue-specific interaction partners.

5.1.4 The Sae1:Sae2 SUMO Heterodimer is a Catalytic Hub for SUMO Activation and Transfer

As mentioned, mature SUMO1-3 canonical isoforms are activated by the SUMO E1 heterodimeric complex, composed of the ~39kDa SUMO activating enzyme 1 (Sae1/AOS1) and the ~71kDa Sae2/Uba2 (*Johnson et al., 1997; Desterro et al., 1999; Gong et al., 1999*). Specifically, Sae1:Sae2 catalyse three distinct and sequential chemical reactions: SUMO adenylation, thioesterification, and Ubc9-mediated transthioesterification, which are catalysed at near identical rates across canonical SUMO isoforms (*Tatham et al., 2001; Lois and Lima, 2005*). Of note, SUMO1-3 adenylation and thioesterification reactions are collectively termed SUMO activation.

SUMO activation begins through the pseudo-ordered binding of ATP, Mg^{2+}, and SUMO to the Sae1:Sae2 inactive and active adenylation domains (*Lois and Lima, 2005; Wang and Chen, 2010*). These first Sae1:Sae2-mediated ligand binding events are associated with Sae1:Sae2 adopting an open structural conformation, hereafter termed "open 1" and "open 2" structural conformations (*Lois and Lima, 2005*) (Figure 5.4). The SUMO E1 adenylation active site catalyses ATP hydrolysis, yielding AMP and pyrophosphate (PP_i), which enables formation of a high-energy acyl SUMO adenylate intermediate ($SUMO_{GG}$-adenosine 5'-monophosphate; AMP) and release of PP_i (*Haas et al., 1982; Johnson et al., 1997*). The $SUMO_{GG}$-AMP intermediate is derived from a phosphoester bond, or acyl-phosphate linkage, between the phosphate group of AMP with the SUMO C-terminal glycine carboxyl group. Sae1:Sae2 bound to the $SUMO_{GG}$-AMP intermediate is also associated with Sae1:Sae2 adopting an open structural conformation, hereafter termed "open 3" (*Olsen et al., 2010*) (Figure 5.4). Following adenylation, the Sae1:Sae2 adenylation active site undergoes structural remodelling to adopt a structural conformation, hereafter termed "closed", to facilitate nucleophilic attack of the $SUMO_{GG}$-AMP intermediate by the Sae2 catalytic cysteine, C173 (*Lu et al., 2010; Olsen et al., 2010*) (Figure 5.4). The half-reaction of this process, termed thioesterification, produces a tetrahedral intermediate (Sae1:Sae2-C173~$SUMO_{GG}$-AMP) and promotes release of PP_i (*Olsen et al., 2010*). The second half thioesterification reaction results in the formation of a high-energy covalent thioester bond between the $SUMO_{GG}$ C-terminal glycyl carboxylate and the Sae1:Sae2-C173 catalytic cysteine thiol moieties (Sae1:$Sae2_{C173}$~$SUMO_{GG}$), and also coincides with release of AMP from the Sae1:Sae2 active site. (*Haas et al., 1982; Johnson et al., 1999; Lu et al., 2010; Olsen et al., 2010*). Successful thioesterification stimulates return of Sae1:Sae2 to an "open-like" structural conformation that can stimulate recruitment of the SUMO E2 enzyme, Ubc9 (*Wang et al., 2010; Boucher et al., 2015*). Some of the structural details of this process, such as the

Sae1:Sae2:Ubc9 recruitment complex interface, have been elucidated, but not all (*Wang et al., 2010; Boucher et al., 2015*). In this state, Sae1:Sae2 can also bind Mg^{2+} and ATP again, and a second SUMO molecule forms non-covalent interactions with the Sae1:Sae2 adenylation pocket, to yield a doubly loaded SUMO E1 ternary complex ($SUMO_{GG}$-Sae1:$Sae2_{C173}$ ~$SUMO_{GG}$) (*Haas et al., 1982; Pickart et al., 1994; Huang et al., 2007; Hann et al., 2019*). The doubly loaded SUMO E1 ternary complex is the most efficient conformational state at transferring the $SUMO_{GG}$ acyl group from the Sae2-C173 thiol to the Ubc9 C93 catalytic cysteine thiol, in a process interchangeably termed transthioesterification and transthiolation (*Haas et al., 1982; Gong et al., 1997; Johnson and Blobel, 1997*). In this process, the doubly loaded ternary complex Sae1:$Sae2_{C173}$~$SUMO_{GG}$ thioester bond undergoes nucleophilic attack from a proximal Ubc9-C93 catalytic cysteine, forming an initial tetrahedral intermediate ($SUMO_{GG}$-Sae1:$Sae2_{C173}$ -$SUMO_{GG}$-$Ubc9_{C73}$) and then full transfer of $SUMO_{GG}$ to the Ubc9-C73 thiol group ($SUMO_{GG}$-Sae1:$Sae2_{C173}$ + $SUMO_{GG}$~$Ubc9_{C73}$) (*Gong et al., 1997; Johnson and Blobel, 1997 ;Kochańczyk et al., 2024*). Upon transthioesterification, Ubc9 dissociates from the SUMO E1, serving to return Sae1:Sae2 to the “open 2” conformation, which enables another round of SUMO activation and thioesterification. The molecular basis of improved transthioesterification efficiency of doubly SUMO-loaded Sae1:Sae2 relative to the singly SUMO-loaded counterpart is currently unknown. However, a recent structure of the *S. pombe* Ub tetrahedral transthioesterification intermediate has revealed that the second adenylation and PP_i release events are coupled to optimal positioning of thioester-linked Ub for transfer through rotation of the catalytic cysteine domain (*Kochańczyk et al., 2024*).

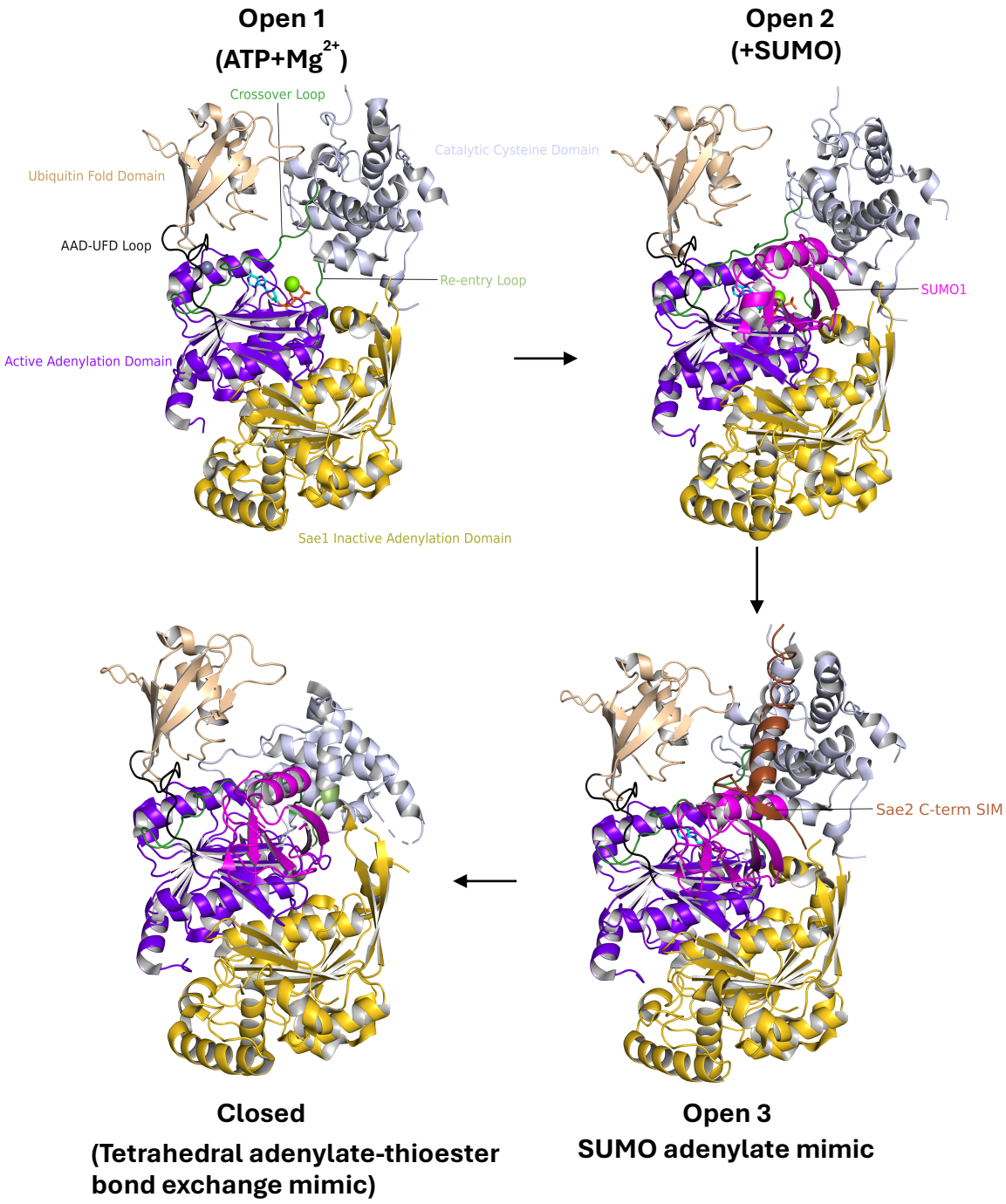


Figure 5.4 - Sae1:Sae2 structural conformational changes during ATP-binding (open 1), SUMO-binding (open 2), and SUMO adenylation (open 3), and thioesterification (closed). The secondary structure of the SUMO E1 and SUMO1 is shown (accession codes: open 1- 1Y8Q, open 2- 1Y8R, open 3- 3KYC and closed- 3KYD). Sae1 (gold) and coloured according to the following scheme: Sae1 (gold), SUMO1 (magenta), Sae2-AAD (purple), Sae2-CC (grey), Sae2-UFD (wheat), Sae2-crossover loops (dark green), Sae2-UFD loop (black), and Sae2-C-term SIM (brown). Atoms for ATP (coloured by element, except the cyan adenine ring) and ATP analogues are represented as sticks, while Mg^{2+} (green) ions are represented as spheres.

From a technical perspective, understanding the molecular basis for Sae1:Sae2 driving the aforementioned reactions was dependent on the generation of crystal structures of open 1-3 and closed 1 conformations (Figure 5.4). The first breakthrough occurred ~20 years when open 1 and open 2 structures were solved by x-ray crystallography (accession codes: 1Y8Q and 1Y8R, respectively) (*Lois and Lima, 2005*). The structure of Sae1:Sae2 bound to the adenylate intermediate, or open 3 structure, was initially elusive, because PP_i release is the rate-limiting step in adenylation, and the reverse reaction can occur. This latter feature has resulted in in the majority of available Sae1:Sae2 crystal structures in the open 2 structural conformation. To circumnavigate this issue, a nonhydrolyzable analogue of the $SUMO_{GG}$-AMP acyl intermediate was generated to trap the first-half reaction for structural analysis. This was achieved through intein-mediated expressed protein ligation (intein EPL) technology, involving a C-terminally truncated form of SUMO1 (residues 1-94) containing a terminal reactive thioester and a synthetic CGG tripeptide containing a C-terminal 5'-(aminodeoxy)adenosine-5'-*N*-monosulfamide (AMSN) group (Figure 5.5). The monosulfamide group is non-hydrolysable, but also mimics the monophosphate in AMP, which is crucial for efficient binding to the Sae1:Sae2 adenylation pocket and inhibition of SUMO activation (*Lu et al., 2010*). The open 3 conformation structure (accession code: 3KYC) confirmed almost all of the relative orientations of the open 1-2 conformation structures, however there was an understandable lack of PP_i coordination (*Olsen et al., 2010*). Similarly, structures of the tetrahedral adenylate-thioester bond exchange tetrahedral intermediate, or closed conformation, and thioester-linked SUMO structures were also elusive, due to the reversibility of the thioesterification half-reaction. Like the SUMO adenylate-mimic structure, the closed Sae1:Sae2 conformation crystal structure (accession code: 3KYD) was solved through intein EPL technology-based trapping of the thioesterification half reaction. In this case, the tripeptide C-terminal group was instead 5'-

(aminodeoxy)adenosine-5'-N-vinyl sulfonamide, whereby the vinyl sulfonamide electrophilic moiety forms an irreversible covalent bond with the nucleophilic C173 catalytic cysteine (Figure 5.5; *Lu et al., 2010; Olsen et al., 2010*).

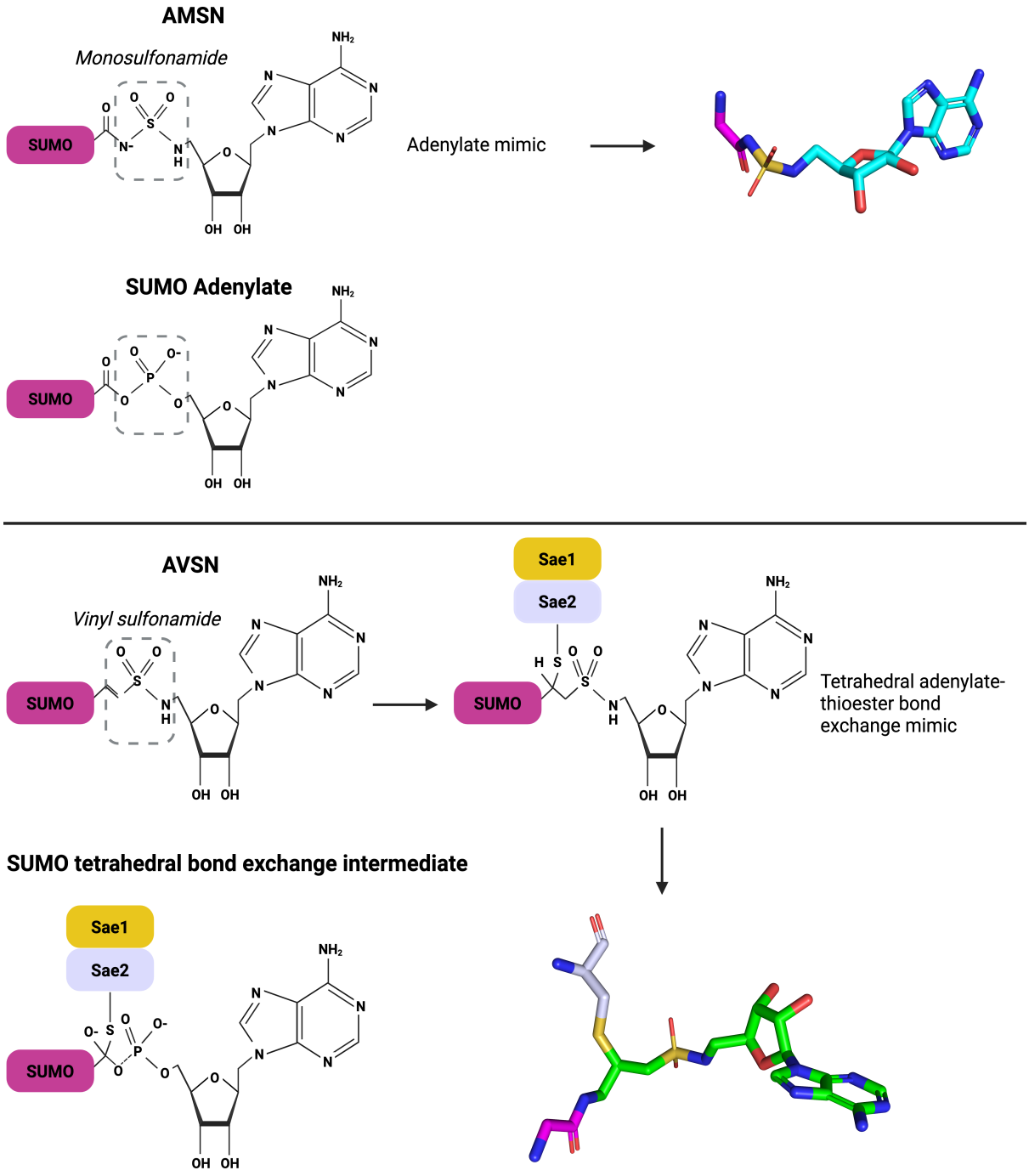


Figure 5.5 – The SUMO adenylate mimic (AMSN) compared to native SUMO adenylates (top panel) and the SUMO tetrahedral adenylate-thioester bond exchange mimic compared to native SUMO tetrahedral bond exchange intermediates (bottom panel). Chemical structures are presented using skeletal formula with the crystal structure of each mimic provided for each mimic (accession codes: 3KYC and 3KYD; Olsen et al., 2010). Figures made using Biorender.com and the PyMOL Molecular Graphics System v3.1.1

5.1.5 The Structural Basis of Sae1:Sae2-mediated SUMO Processing

The Sae1:Sae2 heterodimer forms a canyon-like structure, which is characterised as a canonical UbL E1 enzyme, due to the presence of two pseudosymmetric adenylation domains: the Sae1 inactive adenylation domain (IAD) and the Sae2 active adenylation domain (AAD) (*Capadoccia and Lima, 2018*). These two domains are crucial for SUMO E1 ATP and SUMO-binding, and subsequent adenylation activity. Moreover, the IAD and AAD form a structural base for Sae2 catalytic cysteine (CC) and ubiquitin-fold domains (UFD) to sit atop, respectively, and form opposing walls of the heterodimer (*Lois and Lima, 2005; Capadoccia and Lima, 2018*). The latter subdomains are fundamental for thioesterification, Ubc9 recruitment, and transthioesterification-mediated SUMO transfer (*Lois and Lima, 2005; Wang et al., 2010*).

5.1.5.1 The Sae1 Inactive Adenylation Domain

The 346-amino acid Sae1 IAD predominantly acts as a rigid structural platform for the Sae2 AAD (*Gong, Millas, and Yeh, 1999; Lois and Lima, 2005*). It does so by maintaining a network of salt-bridges, hydrogen-bonding, and hydrophobic contacts with the Sae2 AAD throughout all major SUMO E1 structural transitions, including ATP and SUMO-binding, SUMO activation, and transthioesterification-mediated SUMO transfer (*Lois and Lima, 2005; Olsen et al., 2010*). The preserved salt bridges involve two-fold symmetry related reciprocal Sae1-α_3-E50-K53 and Sae2-α_2-E31-K34 E-x-x-K motifs- a particular feature that is conserved throughout the E1 superfamily (*Lois and Lima, 2005; Burroughs, Iyer, and Aravind, 2009*). An additional salt bridge contact is formed between Sae1-IAD-β_{10}-E341 and Sae2-AAD-K426, located in the loop preceding β_{10} *(Lois and Lima, 2005*. The SUMO E1 IAD:AAD hydrogen-bonding contacts that are maintained throughout all

structural transitions include both Sae1-R235, located in the loop following α_7, and Sae1-β_{10} residues G338 and V340 with Sae2-AAD-L428-L429 and Sae2-AAD-K426, respectively (*Lois and Lima, 2005; Olsen et al., 2010*). An additional hydrogen-bonding contact between Sae1-α_2-Q22 with Sae2-N58, located in the g1-helix between β_2-α_3, stabilises the heterodimer in open 1-3 confirmations (*Lois and Lima, 2005; Olsen et al., 2010*). Reciprocal hydrophobic contacts between residues from Sae1-IAD α_{10} and Sae2-AAD α_{14}, which immediately follows the re-entry loop, also facilitate stabilisation of the SUMO E1 heterodimer.

Sae1 also contributes to CC positioning and the adenylation pocket in open 1-3 conformational states through a two-fold symmetry-related N-terminal helical dyad (*Lois and Lima, 2005; Burroughs, Iyer, and Aravind, 2009; Olsen et al., 2010*). The second of these helices, corresponding to residues 21-36, is positioned proximal to the adenylation active site through stabilising hydrophobic interactions between an R24-L25-W26 motif with Sae2-P385-I387-Y144 sidechain groups (*Olsen et al., 2010*). R24 also forms a sidechain-mediated hydrogen-bonding contact with the main chain carbonyl oxygen atom of Sae2-CC-F374. Importantly, this helix also contains an arginine finger, Sae1-R21, which is conserved throughout the E1 superfamily (*Burroughs, Iyer, and Aravind, 2009*). The role of R21 is to anchor ATP, primed for adenylation, through a sidechain-mediated hydrogen bond contact with the γ-phosphate group of ATP (*Lois and Lima, 2005*). Mutation of Sae1-R21 or the RLW motif to alanine residues inhibits SUMO adenylation, thereby demonstrating the importance of these sites (*Olsen et al., 2010*). Furthermore, the first 25-amino acids of Sae1, encompassing α_1 of the helical dyad, becomes disordered in the closed conformational state, wherein Sae1-α_2 is displaced away from the adenylation active site by the melted Sae2 g_1-helix, located in the AAD between β_2-α_3 (*Olsen et al., 2010*).

One curious feature of the Sae1-IAD is a unique intrinsically disordered insertion, located between β_7 and β_8 that encompasses residues 183-204 (*Lois and Lima, 2005; Olsen et al., 2010*). This region corresponds to the first catalytic half (FCCH) and CC domain in Ub and NEDD8 E1 enzymes, respectively, however the corresponding Sae1 region is unnamed (*Cappadocia and Lima, 2017*). Interestingly, residues 183-198 contain two patches of basic residues that adhere to both monopartite K-[K/R]-x-[K/R] and bipartite, [R/K]-x_{10-12}-K-R-x-K nuclear localisation signal (NLS) motifs, with residues 191-200 sharing a high level of sequence identity with a known c-MYC NLS (*Moutty, Sakin, and Melchior, 2011*). Moreover, alanine mutants of residues 195-196 have previously been shown to abrogate importin α/β-mediated Sae1 nuclear import (*Moutty, Sakin, and Melchior, 2011*). In addition, this region of the IAD appears to be a PTM hotspot, with phosphorylation, acetylation, SUMOylation, and ubiquitylation sites identified by high-throughput proteomic screens (*Hornbeck et al., 2015*).

5.1.5.2 Sae2

Sae2 consists of 640-amino acids that form three structured domains, including the AAD, the CC and a UFD (*Gong, Millas, and Yeh, 1999; Louis and Lima, 2005*). Sae2 also contains a C-terminal 90 amino acid region, corresponding to residues 550-640, that is unique across Ub/UbL E1s and moderately conserved across Sae2 homologues. Interestingly, this region is a major PTM hotspot with phosphorylation, ubiquitylation, SUMOylation, and acetylation sites also identified by high-throughput proteomic screens (*Hornbeck et al., 2015*). Residues 550-606 are dispensable for SUMO E1 adenylation, thioesterification, and transthioesterification reactions, and are unstructured in all available Sae1:Sae2 crystal structures (*Lois and Lima, 2005; Olsen et al., 2010*). However, NMR studies have shown that residues 585-591 in isolation form a canonical SIM-based

contact with SUMO1 (*Song et al., 2004*). Residues 607-640 adopt a helix-loop-strand secondary structure in the SUMO1 adenylate-mimic crystal structure (*Olsen et al., 2010*). The C-terminal 10 amino acids contain a stretch of negatively charged amino acids followed by a hydrophobic core, which adopts an anti-parallel β-strand that forms canonical SIM-based contacts with SUMO. In addition, Residues 610-626 contain two basic patches that also adhere to monopartite and bipartite NLS motifs (*Dohmen et al., 1995; Moutty, Sakin, and Melchior, 2011; Truong et al., 2012*). Deletion of the orthogonal 610-640 residues in *S. cerevisiae* Uba2 disrupt its nuclear localisation, and also minorly affect the corresponding Smt3-mediated substrate conjugation pattern (*Lois and Lima, 2005*). In human Sae2, there are 5 SUMOylation acceptor lysine residues (K611, K613, K617, K623, and K630) located within, or proximal to, the NLS. These 5 SUMOylation sites are important for nuclear retention of the SUMO E1, since mutation of all sites to an R residue results in cytoplasmic localisation of Sae1:Sae2, which is inhibited upon nuclear export inhibition (*Truong et al., 2012*).

5.1.5.2.1 The Sae2 Active Adenylation Domain

The Sae2 AAD, corresponding to residues 1-151 and 387-433 adopts a 3-layered α/β sandwich containing a central 8-stranded β-sheet ($\alpha_1\beta_1\alpha_2\beta_2g_1g_2\alpha_3\beta_3\alpha_4\beta_4\alpha_5\beta_5\beta_6\alpha_{14}\beta_9\beta_{10}$), with β_6 and β_{10} antiparallel to all other strands, which is a characteristic feature of the E1 superfamily (*Lois and Lima, 2005; Burroughs, Iyer, and Aravind, 2009*). The core of the AAD is a 4-stranded Rossmannoid fold (Figure 5.6 left panel) composed of alternating βα units, ($\beta_1\alpha_2\beta_2g_1g_2\alpha_3\beta_3\alpha_4\beta_4\alpha_5$), which is essential for ATP-binding during SUMO adenylation (*Lois and Lima, 2005; Burroughs, Iyer, and Aravind, 2009*). As mentioned, Sae1 is pseudosymmetric with Sae2, and therefore also adopts a similar core 4-stranded Rossmannoid fold (Figure 5.6 right panel; $\beta_1\alpha_3\beta_2g_1\alpha_4\beta_3g_2\alpha_5\beta_4\alpha_6$), however Sae1 lacks necessary catalytic machinery, except required for ATP-binding and adenylation activity.

Sae2 AAD residues involved in the ATP-binding/adenylation pocket are located in highly conserved regions throughout the Ub/UbL E1 superfamily. These include the N-terminus of α_2, the β_2-g1-g2 extended coil region and the loops separating β_3-α_4 and β_4-α_5 (*Lois and Lima, 2005*).

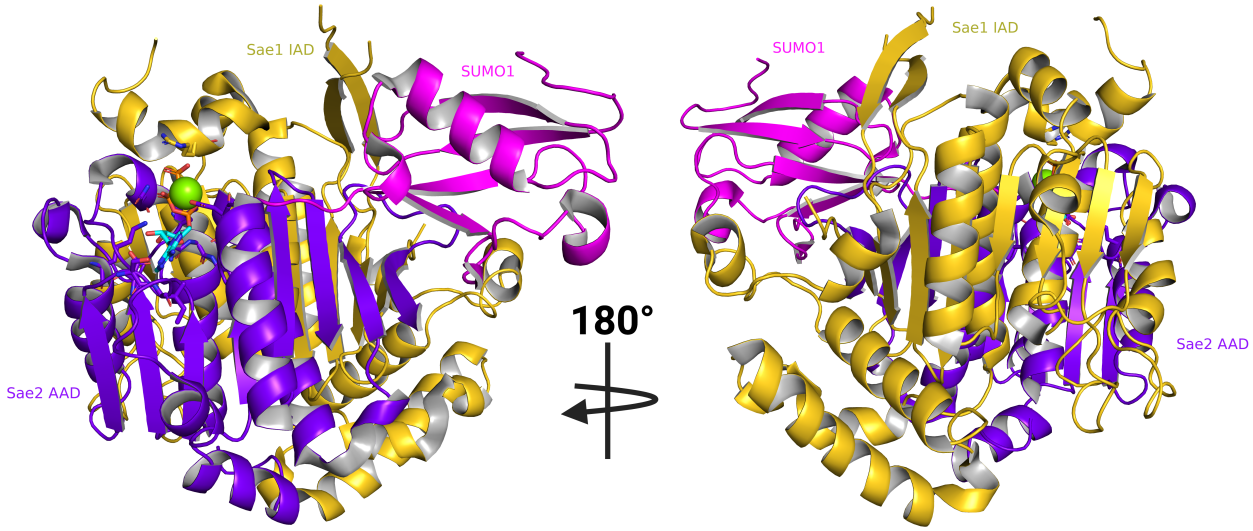


Figure 5.6 – The Sae1 inactive adenylation domain and Sae2 active adenylation domain. The pseudosymmetric Sae1 inactive adenylation (IAD; gold) and Sae2 active adenylation domain (AAD; purple) Rossmanoid folds that comprise the SUMO (magenta) adenylation active site from the open 2 state crystal structure (accession code: 1Y8R; *Lois and Lima, 2005*) are indicated. Sae1, Sae2 and SUMO1 secondary structures are represented in the cartoon format, while the ATP molecule (cyan) is represented in the stick format and Mg^{2+} ion cofactor (green) as a sphere. The structure was rotated about the y-axis by 180° to highlight the AAD (left panel) and IAD (right panel) Rossmanoid folds. Figures made using the PyMOL Molecular Graphics System v3.1.1

5.1.5.2.2 Sae2 Active Adenylation Domain Elements Involved in SUMO Adenylation

An E1-conserved glycine-rich region, located at the N-terminus of Sae2 AAD α_2, engages in hydrogen bonding contacts through G26 and G27 main chain atoms with the ribose 3'-hydroxyl group and the α-phosphate, respectively (Figure 5.7; *Lois and Lima, 2005*). Furthermore, the complementary

positive electrostatic potential of the α_2 helix dipole, facilitates between the main chain amide atoms of G27 and I28 residues with the negatively charged SUMO C-terminal glycyl oxygen atom (Figure 5.7; *Lois and Lima, 2005; Olsen et al., 2010*). This feature of α_2, termed the oxyanion hole, stabilises the SUMO-AMP and adenylated-thioester bond exchange tetrahedral intermediates formed during adenylation and thioesterification reactions, thus serving as the catalytic centre of the adenylation pocket (*Olsen et al., 2010*). A highly E1-conserved 'DNRK' motif, Sae2-D48-N56-R59-K72, located in the β_2-g_1/g_2 extended coil region between β_2-α_3 forms hydrogen-bonding contacts with the ribose 2'- and 3'-hydroxyl groups (D48 and K72), α-phosphate (K72), β-phosphate (N56 and K72), and γ-phosphate (R59) moieties (Figure 5.7; *Lois and Lima, 2005; Burroughs, Iyer, and Aravind, 2009*). The g_1 helix melts to a loop in the closed conformational state, and is displaced from the active site through R59 hydrogen bonding and van der Waals interaction-mediated stabilisation with N56 and L57, respectively (Figure 5.7; *Olsen et al., 2010*). An additional residue within this region, D50, is also highly conserved site throughout Ub/UbL E1s (*Burroughs, Iyer, and Aravind, 2009*). D50 forms sidechain-mediated hydrogen bonds with the main chain amide atoms of N177 and T178 to stabilise the C173 thioester active site loop within the adenylation active site during thioesterification, as observed in the closed structural conformation (*Olsen et al., 2010*). Conversely, N177 and T178 are solvent exposed in open 1-3 structural conformations. As a result, alanine mutation of D50 inhibited SUMO thioesterification activity and cross-linking-mediated stabilisation of the Sae1:Sae2 closed conformation, but has no effect on SUMO adenylation (*Olsen et al., 2010*). Residues S95 and I96, located in the loop between β_3-α_4, are moderately conserved in Ub/UbL E1s. S95 and I96 form key hydrogen bonds with the adenine exocyclic amino group and nitrogen N1 atom, respectively, to anchor the ATP to the binding pocket (Figure 5.7; *Olsen et al., 2005*). Moreover, S95 and M97 sites are key binding sites for the SUMO E1 activation inhibitor,

[(1R,2S,4R)-4-[[5-[[1-[(3-bromophenyl)methyl]-1H-pyrazol-3-yl]carbonyl]-4-pyrimidinyl]amino]-2-hydroxycyclopentyl]methyl ester (ML-792), since exogenous overexpression of mutant Sae2-S95N-M97T can restore cellular SUMO activation (*He et al., 2017*). The highly conserved D117 residue, located in the loop between β_4-α_5 coordinates a Mg^{2+} ion by near-octahedral geometry, in conjunction with the ATP α- and β-phosphates (*Lois and Lima, 2005; Burroughs, Iyer, and Aravind, 2009*). Additionally, D117 forms a hydrogen bond with Sae2-CC-R176 in the closed structural state, such that alanine mutation of D117 or R176 impaired closed state cross-linking and Sae2~SUMO thioester bond formation, although Sae2-R176A mutants were affected to a lesser extent (*Olsen et al., 2010*). Ultimately, the importance of these sites is underscored by inhibited SUMO adenylation upon single alanine mutation of N56, L57, R59, K72, and D117 (*Olsen et al., 2010*). Moreover, the Sae1:Sae2 allosteric inhibitor, COH000, effectively inhibits adenylation, upon covalent bond formation with Sae2-C30, through induction of local conformational changes in the IAD and AAD, leading to disassembly of the adenylation active site (*Li et al., 2019*). Specifically, COH000 causes the AAD-g1-g2 helices to become disordered, leading to displacement of the R21-containing IAD from the ATP-binding pocket. Meanwhile, COH000-induces distortion of the AAD-α_2 oxyanion hole, such that the hydrogen-bond and electrostatic interaction networks involving the ATP α-phosphate and C-terminal SUMO G residue carbonyl oxygen atom are disrupted (*Li et al., 2019*).

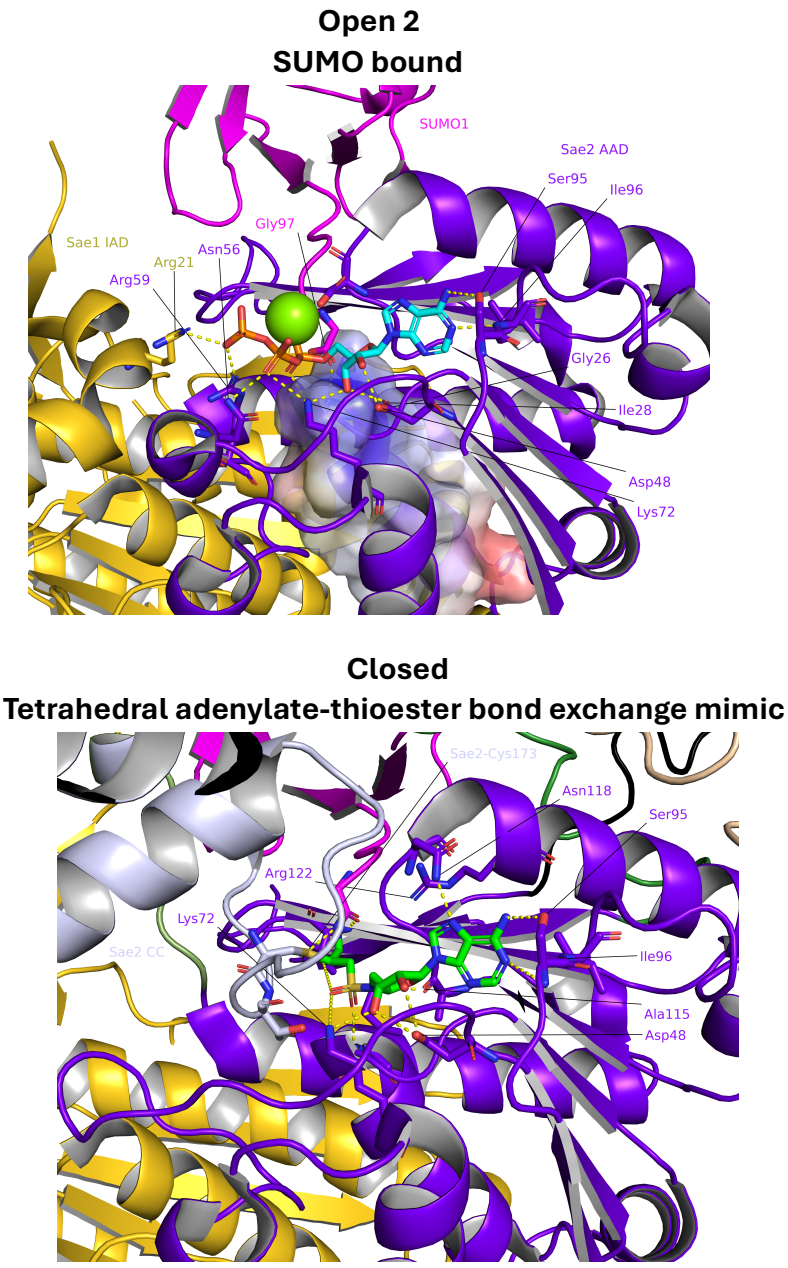


Figure 5.7 – Remodelling of the SUMO adenylation active site drives adenylate-thioester bond exchange. (Top panel) the SUMO adenylation active site from the open state 2 crystal structure (accession code: 1Y8R; *Lois and Lima, 2005)* is provided with key residues involved in ATP-binding and SUMO adenylation indicated and presented in the stick format. Key residues are coloured according to subdomain. Dashed lines depict hydrogen-bonding contacts. The Sae2 α_2 oxyanion hole surface area is represented and coloured by electrostatic potential (electropositive- blue, electronegative- red, and electroneutral- white), as calculated by the Adaptive Poisson-Boltzmann Solver Plugin using the PyMOL Molecular Graphics System v3.1.1 (*Schrödinger, LLC; Jurrus et al., 2018*). (Bottom panel) the SUMO adenylation active site during tetrahedral adenylate-thioester bond exchange is depicted (accession code: 3KYD; *Olsen et al., 2010)*. Key residues are coloured according to subdomain (Sae1 IAD- gold, Sae2-AAD- purple, Sae2-CC-blue white, and SUMO1-magenta). Figures made using the PyMOL Molecular Graphics System v3.1.1

5.1.5.3 Sae2 Crossover and Re-entry Loops: Structural Hinges Required for SUMO Thioesterification

The individual Sae2 AAD inserts connect to the CC through the so-called crossover loop, corresponding to residues 152-170, and the re-entry loop, corresponding to residues 381-386, respectively (Figure 5.8; *Lois and Lima, 2005*). The crossover and re-entry loops are separated by at least 10Å in open 1-3 Sae1:Sae2 structural conformations, however both form a 2-stranded parallel β-sheet ($\beta_7\beta_8$) in the closed conformation (Figure 5.8; *Lois and Lima, 2005; Olsen et al., 2010*). The loop-β-sheet structural transition facilitates the 130° rotation 3Å translation of the Sae2-CC domain, which enables a 34Å movement of the Sae2-C173 thioester active site towards the adenylation active site during SUMO thioesterification (Figures 5.7 and 5.8; *Olsen et al., 2010*). CC domain translocation and subsequent remodelling of the adenylation active site is underpinned by a 125° ψ-angle rotation of crossover loop residues 164-168 and a 147° ϕ-angle rotation between G381 and N382 re-entry loop residues, respectively (*Olsen et al., 2010*). Previous mutation of individual crossover loop residues had no effect on SUMO~Sae2 thioester bond formation, nor cross-linking-mediated stabilisation of the SUMO E1 closed-state conformation (*Olsen et al., 2010*). Consequently, the 164-168 patch cumulatively contributes to the hinge properties of the crossover loop. In contrast, Sae2-G381P-N382P and Sae2-N382P re-entry loop mutants inhibit both of SUMO thioester bond formation and closed conformational state cross-linking, thereby highlighting the importance of these particular sites for CC movement and subsequent remodelling of the Sae1:Sae2 active site during thioesterification (*Olsen et al., 2010*).

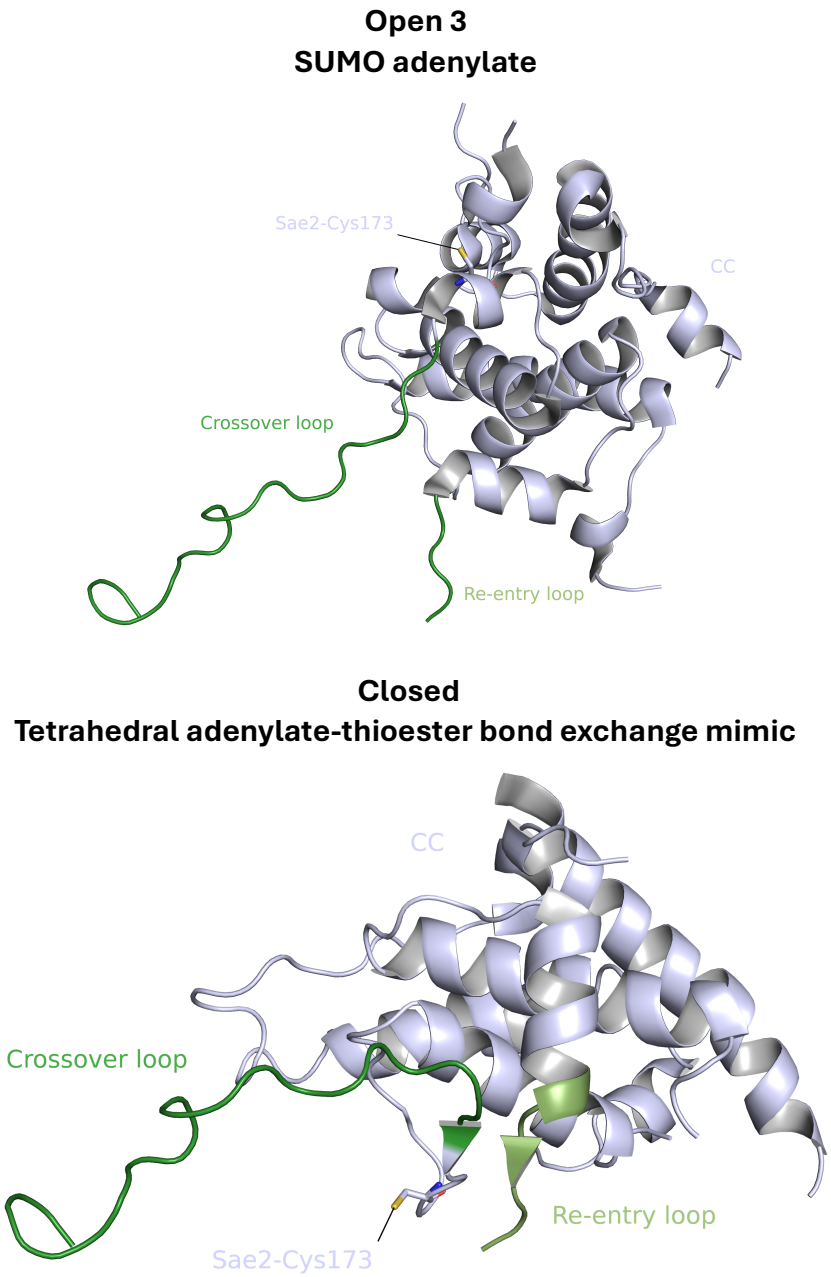


Figure 5.8 – The Sae2 crossover loop and re-entry loops act as structural hinges during SUMO thioesterification. The SUMO catalytic domain (blue white; CC), Sae2 catalytic Cys173, crossover loop (forest), and re-entry loop (smudge) from the open state 3 SUMO adenylate mimic (top panel; accession code: 3KYC; *Olsen et al., 2010)* and closed state tetrahedral adenylate-thioester bond exchange mimic (bottom accession code: 3KYD; *Olsen et al., 2010)* crystal structures is provided. Protein secondary structures are represented in the cartoon format and Sae2 catalytic Cys173 is highlighted in the stick format. Figures made using the PyMOL Molecular Graphics System v3.1.1

5.1.5.4 The Sae2 Crossover Loop CxxC Motif

Interestingly, Sae2-cross-over loop residues, C158 and C160, along with Sae2-UFD-C441-C444, both possess CxxC motifs responsible for tetrahedral chelation of a Zn^{2+} ion, as observed in all available Sae1:Sae2 crystal structures (*Lois and Lima, 2005; Olsen et al., 2010*). The role of the cross-over:UFD bipartite Zn^{2+}-binding motif is to maintain the structural integrity of the SUMO E1. It is also proposed that the Zn^{2+}-binding motif is involved in a pseudo-ordered substrate-binding mechanism that proceeds by initial binding of ATP, prior to SUMO binding (*Wang and Chen, 2010*).

5.1.5.5 The Sae2 Catalytic Cysteine Domain

The CC domain, corresponding to residues 171-381, is predominantly helical, consisting of $\alpha_6\alpha_7g_3g_4\alpha_8\alpha_9g_6\alpha_{10}g_7\alpha_{11}\alpha_{12}\alpha_{13}$ in open conformation structures (Figure 5.8; *Lois and Lima, 2005; Olsen et al., 2010*). The CC domain is characterised by the catalytic cysteine, C173, residing in α_6 immediately following the crossover loop (Figure 5.8), which is a conserved feature throughout evolution of all thioester-competent canonical E1s (*Lois and Lima, 2005; Burroughs, Iyer, and Aravind, 2009*). C173 is positioned below the g4 helix, or so-called Cys cap (*Olsen et al., 2010; Cappadocia and Lima, 2017*). The Cys cap is proposed to maintain low C173 nucleophilicity, or the thiol form of the sidechain group, in open 1-3 conformations through raising the pKa by the acidic sidechain groups of D212 and E214 (*Olsen et al., 2010*). In the closed conformation state, the g3-loop-g4 region becomes disordered, thereby reducing the pK_a and relieving the nucleophile reactive nature of the C173 thiolate (Figure 5.8). As mentioned, α_6 melts to a loop in the thioester tetrahedral intermediate state, enabling positioning of C173 proximal to the C-terminal $SUMO_{GG}$~AMP adenylate within the adenylation active site (Figure 5.7; *Olsen et al., 2010*). In turn, this primes C173 to attack the $SUMO_{GG}$~AMP nucleophile and catalyse SUMO thioesterification. Further CC structural

components undergo melting or disordering to enable the melted α_6 to occupy the Sae1 adenylation pocket in the closed conformation (Figure 5.8; *Olsen et al., 2010*). The α_8 helix partially melts to form the g5 helix, which maintains the majority of stabilising hydrophobic contacts with the base of α_9. Meanwhile, part of the loop separating α_{10}-α_{11} and the succeeding g7 helix become disordered, which results in a loss of a stabilising hydrogen bond contact between Sae2-K308 and Sae1-G9, associated with the disordering of the N-terminus of Sae1 (*Lois and Lima, 2005; Olsen et al., 2010*). Additionally, residues 337-344 from the loop separating α_{11} and α_{12} become disordered, resulting in a loss of intradomain contacts with α_7, α_{12}, and α_{13} (*Lois and Lima, 2005; Olsen et al., 2010*).

5.1.5.6 The Sae2 Ubiquitin Fold Domain Recruits Ubc9

The UFD, corresponding to residues 448-547, forms a modified 6-stranded β-grasp fold ($\beta_{11}\alpha_{15}\beta_{12}\beta_{13}g_8\beta_{14}\beta_{15}\beta_{16}$), which exhibits a high level of structural similarity with SUMO and Ub (*Lois and Lima, 2005*). The importance of this region was established through deletion of the corresponding UFD region in *Saccharomyces cerevisiae* and human Uba2/Sae2, which abrogated SUMO transfer to Ubc9 without impacting E1~SUMO thioester formation, and was unable to rescue colony lethality in a Δ*uba2* null strain (*Lois and Lima, 2005*). The Sae2-UFD functions as a Ubc9-specific recruitment domain to facilitate transthioesterification-mediated transfer of SUMO from Sae2-C173 to Ubc9-C93 (Figure 5.9; *Wang et al., 2010; Boucher et* al., 2015). The Sae2-UFD and Ubc9 form an interlocked complex provided by a UFD W-shaped groove that is derived from $\beta_{12}\beta_{13}g_8\beta_{14}$ (Figure 5.9; *Wang et al., 2010; Boucher et* al., 2015). This region contacts both termini of the Ubc9 N-terminal α_1 helix, as well as the β_1-loop-β_2 region, which are positioned distally from the catalytic cysteine (*Wang et al., 2010; Boucher et* al., 2015). There are two major regions within Ubc9 that determine the specificity of Ubc9 recruitment to the UFD (*Liu, Lois, and Reverter, 2017*). Firstly,

a basic patch within the C-terminal side of the Ubc9-α_1 helix, derived from R13-K14-R17-K18 residues, forms a network of hydrogen bonding and salt bridge contacts with the complementary acidic side chain groups of UFD-D479-E494-E497-D520, and the hydroxyl group of UFD-S492 (Figure 5.9; *Boucher et al., 2015; Liu, Lois, and Reverter, 2017*). Moreover, the UFD-F522 sidechain fits into a hydrophobic pocket formed by the aliphatic portion of Ubc9 K14-R17-K18 sidechains (*Boucher et al., 2015; Liu, Lois, and Reverter, 2017*). Specifically, the Ubc9 basic patch-contacting residues are located in UFD β_{12}, the loop separating β_{13}-g_8, and β_{14} (*Boucher et al., 2015*). The complementary electrostatic interface between the Ubc9 basic patch and the Sae2-UFD acidic patch is unique to the SUMO cascade (*Liu, Lois, and Reverter, 2017*). The second Ubc9 binding surface is a hydrophobic patch formed by residues L6, M36, and L38 from the N-terminal side of both α_1 and β_2 (Figure 5.9; *Boucher et al., 2015; Liu, Lois, and Reverter, 2017*). Hydrophobic Ubc9 residues contact Sae2-UFD residues Q481,G487, and I489 from the β_{12}-loop-β_{13} region. L6 forms weak hydrophobic contacts with Sae2-UFD residues Q481 and G487, while M36 and L38 form a total of three main chain-mediated hydrogen bonds with G487 and I489 (*Boucher et al., 2015*). The remaining Sae2-UFD:Ubc9 interface is formed by Ubc9 residues D33, G34, and T35 located in the loop separating β_1-β_2 with UFD residues N501 and F509 from the loop-g8 helix region (*Boucher et al., 2015*). UFD-N501 forms an indirect sidechain-mediated hydrogen bonding contact with the D33 main chain oxygen atom and a direct hydrogen bond with the G34 main chain oxygen atom G34, while the F509 benzyl sidechain provides supporting hydrophobic contacts with G34 and T35 (*Boucher et al., 2015*).

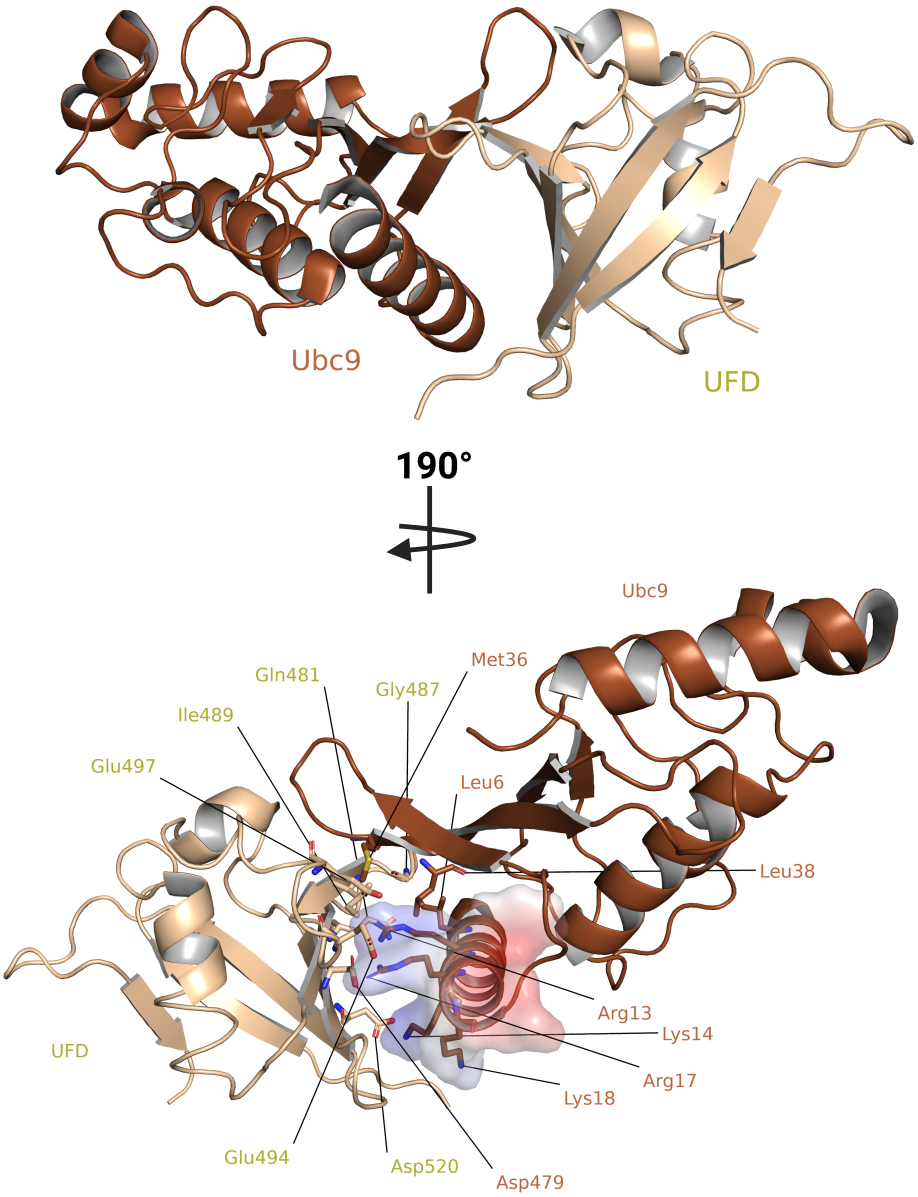


Figure 5.9 – The Sae2 ubiquitin fold domain (UFD) recruits Ubc9. (Top panel) the secondary structure of Sae2-UFD (wheat) bound to Ubc9 (brown) from the crystal structure of the complex (accession code: 4W5V; *Boucher et al., 2015)* is provided in the cartoon format. (Bottom panel) the interface between Sae2-UFD is rotated around the y-axis clockwise by 190°. Residues involved in the interface are highlighted and presented in the stick format. The Ubc9 basic patch surface area is represented and coloured by electrostatic potential (electropositive- blue, electronegative- red, and electroneutral- white), as calculated by the Adaptive Poisson-Boltzmann Solver Plugin using the PyMOL Molecular Graphics System v3.1.1 (*Schrödinger, LLC; Jurrus et al., 2018*). Figures made using the PyMOL Molecular Graphics System v3.1.1

5.2 Preliminary Work Suggests that Sae1-S185 Phosphorylation in the Unstructured loop is Induced by Multiple Stress-Inducing Stimuli

Unpublished research conducted by Dr Alexander Lanz (Professor Jo Morris' Lab) has investigated a potential novel mitotic DDR pathway, involving phospho-regulation of the SUMO E1 heterodimer subunit, Sae1. Specifically, the phosphorylation status of the 24-amino acid unstructured loop S185 residue was monitored using a phospho-specific antibody (Sae1-pS185) in response to mitotic and DNA damage-specific treatments. These experiments were performed following depletion of endogenous Sae1:Sae2 levels and complementation with exogenously overexpressed wild-type heterodimers.

5.2.1 Sae1-S185 Phosphorylation is Induced in Response to Genotoxic Stress and Non-Phosphorylatable Sae1-S185A-Containing SUMO E1 Heterodimers Sensitises Cells to Genotoxic Stress

Following irradiation (10Gy), exogenous S185 was phosphorylated in an ATM-dependent manner, and Sae1-pS185 levels peaked 2 hours post-IR recovery, and persisted until 8 hours. Minor Sae1-pS185 levels were induced prior to ATM autophosphorylation at S1981, ATM-mediated phosphorylation of the transcriptional co-repressor KRAB-associated protein 1 (KAP1)/transcription intermediary factor 1β (TIF1β) at S824, and ATR-mediated phosphorylation of Chk1 at S317. Interestingly, similar IR-dependent pS185 levels and persistence were observed in a catalytically deficient Sae1:Sae2$_{C173G}$ background mutant, suggesting that S185 phosphorylation occurs independently of SUMO E1 activation status. Importantly, complementation with exogenous Sae1:Sae2$_{C173G}$ slightly increased levels of global endogenous SUMO1 and RanGAP-SUMO conjugates relative to Sae1:Sae2 depletion alone, thereby suggesting that endogenous SUMO E1

levels were not consistently depleted. IR-dependent Sae1-S185 phosphorylation was blocked by p38 kinase inhibition induced by adezmapimod/SB203589 treatment and was also reduced by olaparib-induced polyadenosine 5'-diphosphoribose polymerase 1/2 (PARP1/2) inhibition. Conversely, Sae1-pS185 levels 2-hour post-IR recovery were enhanced upon inhibition of CDK1/2/5/7/9, Chk2, Src, mitogen-activated protein kinases 1/2 (MEK1/2), JNK1-3 kinases, as well as upon microtubule destabilisation, and inhibition of Rad51 and sirtuin 1/2 (SIRT1/2) deacetylation activity. The strongest enhancement in Sae1-pS185 levels were observed upon SIRT1/2 inhibition, microtubule destabilisation, and Rad51 inhibition. Enhanced pS185 2-hours post-IR recovery was induced by roscovitine, Chk2 inhibitor hydrate, SU6656, PD0325901, SP600125, colcemid, B02, and Salermide treatments, respectively. Of note, IR-dependent γH2AX levels and kinetics were not associated with Sae1:Sae2 levels. Additionally, exogenous Sae1-S185 phosphorylation was induced by camptothecin (CPT)-mediated DNA damage, peaking after 1-hour post-recovery, and was near inhibited upon ATM inhibition. Sae1:Sae2 levels and Sae1-S185 phosphorylation status was associated with IR- and CPT-dependent sensitivity, as determined by colony survival assays. Sae1:Sae2 depletion significantly increased cellular IR and CPT-sensitivity, which was rescued upon overexpression of the exogenous wild-type heterodimer and $Sae1_{S185D}$:Sae2 phosphomimetic, however exogenous overexpression of catalytically deficient Sae1:$Sae2_{C173G}$ and phosphonull $Sae1_{S185A}$:Sae2 proteins failed to rescue the IR-sensitivity phenotype.

5.2.2 Sae1-S185 Phosphorylation is Induced by Mitotic Arrest, Enhanced by Genotoxic Stress in an Apparent DNA-PK, p38, ATM, and ATR-Dependent Manner, and is Associated with Mitotic SUMO 2/3 Chain Formation

Interestingly, Sae1-pS185 was induced upon prometaphase arrest induced by nocodazole-mediated microtubule destabilisation, and Sae1-pS185 levels peaked in intensity after cells were released into mitosis after 15-30 minutes. Moreover, Sae1-pS185 levels remained after 8 hours of nocadazole release, despite H3-pS10 levels were ablated after 2 hours. Consequently, these results suggested that Sae1-pS185 levels persist throughout all mitotic stages and upon G_1 entry. Mitotic Sae1-pS185 levels were abrogated upon DNA-protein kinase (DNA-PK) and p38 inhibition and were drastically reduced upon ATM and ATR inhibition. Strikingly, mitotic Sae1-pS185 levels were further enhanced in response to IR-dependent DNA damage, and this was associated with a reduced level of H3-pS10 following nocodazole release. IR-damaged mitotic Sae1-pS185 levels were abolished upon ATM inhibition, and were near absent upon ATR, DNA-PK, and p38 inhibition. Moreover, Sae1:Sae2 thioester-competency and $Sae1_{S185}$:Sae2 phosphorylation status were intimately linked with mitotic sensitivity to IR treatment and 2-hour γH2AX foci clearance in mitotic cells following IR damage. Again, exogenous expression of wild-type Sae1:Sae2 and $Sae1_{S185D}$:Sae2 restored mitotic IR sensitivity, while complementation with catalytically/thioester-deficient Sae1:$Sae2_{C173}$ or phosphonull $Sae1_{S185A}$:Sae2 exhibited similar levels of survival to that of SUMO E1-depleted cells. Strikingly, $Sae1_{S185D}$:Sae2 complementation enhanced 2-hour γH2AX foci clearance in pH3-postive cells post IR-recovery. Conversely, catalytically deficient Sae1:$Sae2_{C173G}$ and phosphonull $Sae1_{S185A}$:Sae2 exhibited comparable 2-hour γH2AX foci recovery levels post mitotic IR-recovery to that of depleted cells, while ATM, p38, and ATR-inhibited cells possessed similar 2-hour γH2AX foci recovery levels post mitotic IR-recovery to that of untreated cells. Interestingly, a similar defect in

mitotic IR-recovery was observed upon inhibition of Sae1:Sae2 with ML-792, however combined Rad52 inhibition restored normal γH2AX foci clearance. Recent experiments have highlighted a dependency of mitotic SUMO chain formation on the phosphorylation status of Sae1-S185. In these experiments endogenous SUMO E1 activity was inhibited upon treatment with ML-792, while ML-792-resistant Sae1:Sae2 SUMO E1s (Sae1:Sae2-S95N-M97T) containing wild-type, phosphonull (S185A), and phosphomimetic (S185D) mutant Sae1 were overexpressed, and mitotic SUMO chain formation was monitored. Mitotic SUMO1 chains were low across all ML792-treated cells, in agreement with the observation that mitotic SUMOylation predominantly involve SUMO2/3 chains, however S185D phosphomimetic-expressing cells did possess an enhanced level of mitotic SUMO1 chains. Strikingly, mitotic SUMO2/3 chains were abolished in exogenously expressed S185A cells, while wild type were unperturbed and phosphomimetic were enhanced. This preliminary result, therefore, suggests that S185 phosphorylation is involved in mitotic SUMO chain formation by an unknown mechanism.

5.2.3 *In vitro* SUMOylation Assays with Phosphomimetic Sae1-S185D-Containing SUMO E1 Heterodimers May Not Affect RanGAP1 SUMO Conjugation

The potential role of S185 phosphorylation status in E1 activation was then explored through *in vitro* SUMOylation assays. *In vitro* SUMOylation assays were reconstituted using recombinant wild-type Sae1:Sae2 proteins, Ubc9, combinations of SUMO1 or SUMO2 or both SUMO isoforms, and a RanGAP1 C-terminal fragment ($RanGAP1_{C\text{-}term}$) was used as a SUMO1/2 modifiable substrate. Specifically, wild-type Sae1:Sae2, catalytically deficient $Sae1{:}Sae2_{C173G}$, double phosphomimetic $Sae1_{S185D/S289D}$, and double phosphonull $Sae1_{S185A/S289A}$ mutant SUMO E1 heterodimers were employed, and the RanGAP1 C-terminal fragment was derived from residues 398-597. In all cases,

RanGAP1 SUMO1/2ylation levels were equivalent after 10 minutes across wild-type and phosphomutant E1 heterodimers. Despite the strong Sae1:Sae2 phosphomutant DNA damage and mitotic DNA damage rescue phenotypes, the *in vitro* SUMOylation assays suggested that S185 phosphorylation status was not involved in overall RanGAP1 SUMOylation efficiency, nor in SUMO isoform specificity. It was speculated that the -1 negative charge imparted by the S185D side chain was insufficient to recapitulate S185 phosphorylation-dependent effects in E1 activity and SUMO isoform preference. However, this model did not consistently fit with the observed cellular phenotypes observed in S185D phosphomimetic-expressing cells.

5.2.4 Summary of Preliminary Data

In summary, these data present a complex relationship between the phosphorylation status of Sae1-S185, the DDR, and mitosis. From a DDR perspective, these results clearly suggest that Sae1-S185 phosphorylation is an early IR and CPT-dependent DNA damage response event that is positively associated with intact p38 MAPK and ATM signalling. In particular, DNA damage-dependent activation of p38 MAPK signalling is well-established, however the role of p38 in downstream G_2/M DNA damage checkpoint activation is possibly dispensable (*Phong et al., 2010*). In contrast, IR-dependent S185 phosphorylation was inversely associated with intact microtubules, as well as SIRT1/2 deacetylation and Rad51 recombinase activity. In particular, the negative relationship between Sae1-S185 phosphorylation and intact microtubules was mirrored by enhanced Sae1-pS185 levels upon nocodazole-induced prometaphase arrest experiments. This particular result is likely similar, because nocodazole-induces microtubule destabilisation and subsequent activation of SAC-mediated mitotic arrest. From a mitotic perspective, peak Sae1-pS185 levels after 15-30 minutes nocodazole release and reduced levels throughout latter

timepoints suggest that maximal Sae1-S185 phosphorylation likely occurs between prometaphase and the metaphase-to-anaphase transition, which is corroborated by presence of H3-pS10 levels during this timeframe. Moreover, Sae1-pS185 levels were ablated upon DNA-PK inhibition, which is a known regulator of microtubule dynamics, kinetochore attachments, and subsequent chromosomal segregation (*Huang et al., 2014*). It is important, therefore, that further markers of mitotic progression, such as securin, cyclin B1, and PLK1, are used to further hone on the specific mitotic stage. It is equally curious that mitotic Sae1-S185 levels were further increased after 15-30 minutes IR-recovery, since H3-pS10 levels decreased more rapidly at these timepoints compared to nocodazole release-induced mitotic progression alone. These data would suggest that DNA damage in mitosis increases Sae1-S185 phosphorylation and the subsequent propensity for cells to exit mitosis, perhaps through mitotic catastrophe-based cell death, or simply through accelerated progression through latter mitotic phases.

5.2.4 Limitations of the Preliminary Data suggest a Lack of Importance of Sae1-S185 Phosphorylation in SUMO Activation

One of the pieces of this Sae1-pS185 mitotic DDR puzzle that did not fit with the observed cellular phenotypes was the *in vitro* SUMOylation assay data conducted using recombinant phosphonull and phosphomimetic S185 mutant-containing SUMO E1s, which showed no significant difference. These data would suggest that Sae1-S185 phosphorylation does not directly affect SUMO E1 activation or isoform bias, as supported by the distal positioning of the Sae1 unstructured loop away from SUMO recognition, adenylation, thioesterification, and transthioesterification sites. While it was conceivable that the phosphomimetic S185D residue -1 charge carboxyl side chain did not accurately reflect the -2-charge imparted by a pS residue γ-phosphate side chain, the same logical

conclusion could be applied to the mitotic DDR sensitivity experiments, which were rescued upon overexpression of phosphomimetic Sae1-S185D-containing SUMO E1 heterodimers. Additionally, *in vitro* SUMOylation assays were performed in the context of double phosphomimetic $Sae1_{S185D/S289D}$:Sae2 mutants, and therefore the role of S185 phosphorylation was not formally investigated in isolation. It was equally possible that single 10-minute timepoints were an insufficient quantifiable metric to quantify SUMO activation and transfer rates. In addition, the conducted *in vitro* SUMOylation assays combine the sum of adenylation, thioesterification, transthioesterification, and Ubc9-mediated RanGAP1 SUMOylation reactions, which may have overall masked the contribution of SUMO E1-specific reactions. This point was later addressed by SUMO E1-specific adenylation reactions using BODIPY FL ATP, involving wild-type, phosphonull, and phosphomimetic Sae1-S185-containing SUMO E1s, but no significant difference in adenylation activity was observed (personal communication). Taking these considerations into account, comparing wild-type and $Sae1_{pS185}$:Sae2 RanGAP1 SUMO1/2ylation efficiencies by *in vitro* SUMOylation assays with full time courses were the most likely productive avenue of research to explore the functional importance of Sae1-S185 phosphorylation, before consideration of alternative hypotheses.

5.3 Aims and Objectives

The potential role of Sae1-S185 phosphorylation in SUMO E1 activation was particularly interesting to us, because S185 is located in the unstructured loop between β_7-β_8 in the upper region of the Sae1:Sae2 heterodimer (Figure 5.10), proximal to the Sae2-UFD, and the unstructured loop has no known functional role in E1 activity. Structural comparisons of the open 3 and closed structural conformations suggested that the unstructured loop undergoes a minor conformational change

between adenylation and thioesterification states. In the open 3 structural conformation, the Sae2 C-terminus structured region is extended by density from residues 607-640, of which residues 632-640 form a SIM that contacts the SUMO1 hydrophobic groove. In the closed state the Sae2 residues 607-640 are disordered, suggesting that the Sae2 SIM contact with SUMO1 is disrupted, whereas the Sae1 β_7 unstructured loop N-terminal exit site is extended by 3 amino acids, which all form weak hydrophobic contacts with SUMO1 centred around β_2 residues 33-35. Moreover, the local sequence surrounding the Sae1-pS185 site, VAKV**pS**, loosely resembles a Φ-Φ-x-Φ-pS phospho-SIM consensus sequences. Taken these initial analyses together, we speculated that S185 phosphorylation creates a weak pseudo phospho-SIM. The pseudo phospho-SIM could conceivably compete with the adenylate-state Sae2 SIM for SUMO hydrophobic patch binding, thus promoting processive SUMO adenylate-thioester bond-exchange. Consequently, we anticipated that pS185 Sae1:Sae2 heterodimers could exhibit increased initial RanGAP1 SUMOylation rates. As a result, mitotic or DNA damage stimuli could enable accelerated substrate SUMOylation, which would provide a molecular basis for the observed DNA damage sensitivity in phosphonull S185-expressing cells.

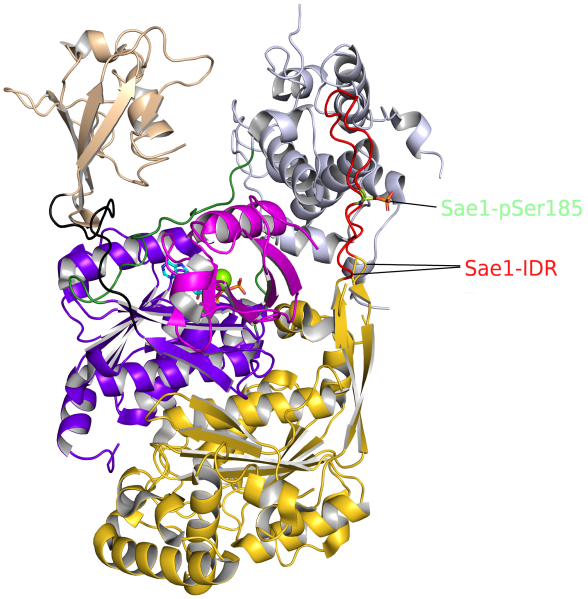


Figure 5.10 – The Sae1-IDR pSer185 site is predicted to be located distally to the SUMO processing active site. The secondary structure of the SUMO E1 and SUMO1 is shown in the open 2 conformational state (accession code: 1Y8R). The Sae1-IDR (red) secondary structure from an AlphaFold 3 structural prediction of the phosphorylated SUMO E1 is highlighted, following structural alignment to the open 2 structure *(Abramson et al., 2024)*. The Sae1-pSer185 (lime) is represented in the stick format and coloured by atom. All amino acids within the Sae1-IDR have a very low per residue local confidence (pLDDT) score of >50. The SUMO E1 is coloured according to the following scheme: Sae1 (gold), SUMO1 (magenta), Sae2-AAD (purple), Sae2-CC (grey), Sae2-UFD (wheat), Sae2-crossover loops (dark green), and Sae2-UFD loop (black). Atoms for ATP (coloured by element, except the cyan adenine ring) are represented as sticks, while Mg^{2+} (green) ions are represented as spheres.

5.4 Results

5.4.1 Recombinant Expression and Purification of Site-Specfically Phosphorylated Sae1-pS185 Protein

Based on these goals, we, therefore, sought to purify recombinant $Sae1_{pS185}$:Sae2 heterodimers using the well-established pSer OTS. The pSer OTS diverges slightly from the pThr OTS, whereby IPTG-inducible pKW2-SepRS(2)-pSertRNA(B4)-EF-Sep and pSerOTS-C1* are used for expression of

the orthogonal pSer-$tRNA_{CUA}$:pSerRS translational pair in Bl21ΔserB and rEcoli XpS *E.coli* expression strains. The pKW2-SepRS(2)-pSertRNA(B4)-EF-Sep vector and Bl21ΔserB was gifted by the Chin lab, while pSerOTS-C1* and rEcoli XpS were gifted by the Rhinehart lab. In the BL21ΔserB strain, the serine phosphatase gene was deleted to prevent dephosphorylation of site-incorporated pSer. rEcoli XpS is an *E.coli* K12 MG1655 substrain derivative with a few key differences. Firstly, all 321 genomic amber stop codons were recoded to TAA ochre stop codons to reduce global nonsense amber suppression and OTS cytoxicity. Additionally, the polypeptide release factor 1 (*prfA*) gene was deleted to prevent premature translation termination at internal amber stop codons, and serB was deleted. We then proceeded by cloning the full-length *Sae1* gene in-frame of N-terminal MBP affinity tag and 3C protease cleavage site, and a C-terminal His_6 affinity tag, into the aforementioned RSF-origin-containing AC6 expression vector. Following this, the codon encoding S185 was mutated to an amber TAG stop codon by site-directed mutagenesis, in order to specify the phosphoserine incorporation site. We then conducted 2L LB, TB, and 2TY expression trials of MBP-3C-$Sae1_{pS185}$-His_6 using the AC6 expression vector with the Chin BL21ΔserB:pKW2-SepRS(2)-pSertRNA(B4)-EF-Sep pSer OTS, and later using the Rhinehart rEcoli XpS:pSerOTS-C1* pSer OTS (Figure 5.10). Recombinant protein expression and OTS expression was conducted alike the MBP-3C-ZCC_{pT690}-His_6 trials, except the media was further supplemented 1mM O-phospho-L-serine to provide the bacteria with an extracellular source of pSer. Initial dual-tagged $Sae1_{pS185}$ recombinant proteins were purified by initial amylose affinity chromatography, followed by cobalt affinity chromatography and extensive dialysis to remove imidazole contaminants (Figure 5.10). Again, this strategy was employed to initially isolate N-terminal MBP-tagged truncated MBP-3C-$Sae1_{185*}$ and dual-tagged full-length MBP-3C-$Sae1_{pS185}$-His_6 protein variants, and subsequent cobalt affinity tag would isolate full-length protein only. SDS-PAGE analysis of the purified recombinant dual-tagged protein

indicated that we obtained a single ~100kDa band across all trialled medias, which roughly corresponded to the predicted molecular weight of 80.6kDa (Figure 5.11). Of note, the best performing expression medias were TB and 2TY, while LB yields were moderately reduced. This process of expression trial optimisation was repeated again using the Rhinehart pSer OTS. The typical MBP-3C-Sae1$_{pS185}$-His$_6$ yields using the Chin and Rhinehart pSer OTSs expression and purification trials were ~50μg/L culture and 420μg/L culture, respectively (Figure 5.11).

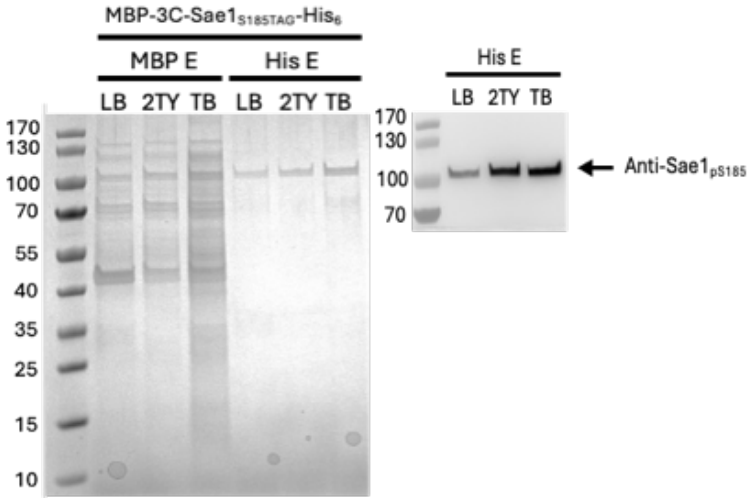
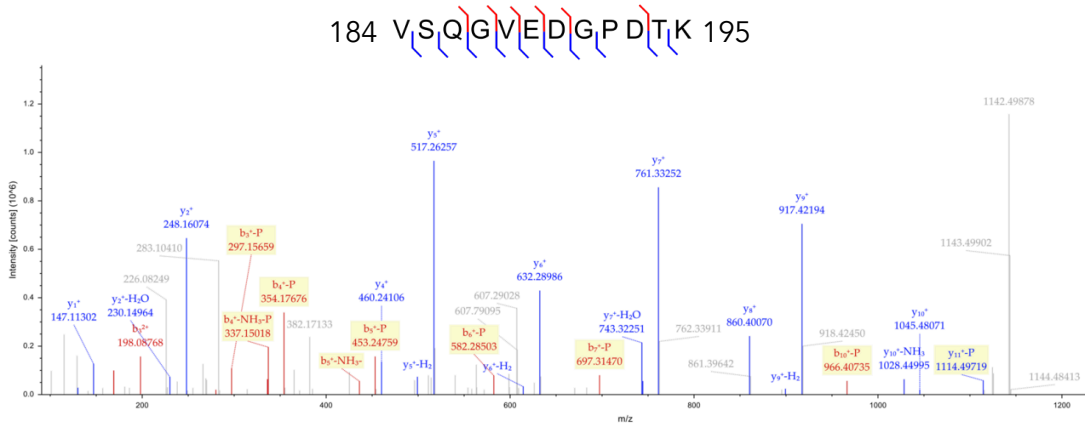


Figure 5.11 - Purification and Western blot validation of site-specifically phosphorylated recombinant MBP-3C-Sae1-pS185-His6 protein. (left panel) SDS-PAGE analysis of two-step affinity purification of recombinant Sae1-pS185 from media expression trials. Elutions (E) from first-step amylose (MBP) and second-step cobalt (His) affinity chromatography purifications of MBP-3C-Sae1-pS185-His$_6$ after expression trials conducted in LB, 2x Yeast Extract Trypton, and TB media is indicated. (right panel) Western blot validation of site-specific phosphoserine incorporation at Sae1-S185 in recombinant MBP-3C-Sae1-pS185-His6 protein. His elutions from (left panel) were probed with a phospho-specific anti-Sae1-pS185 antibody.

Given that our observed recombinant protein was ~20kDa larger than the predicted molecular weight, we validated successful Sae1$_{pS185}$ incorporation two-fold. Firstly, we conducted Western blot analysis of the purified proteins using the anti-Sae1$_{pS185}$ phospho-specific antibody (Figure 5.11), which revealed that the S185 residue was phosphorylated across all medias tested. Following this, all purified proteins were pooled together and concentrated to ~5μM and the >90% pure ~100kDa

band was excised and analysed by mass spectrometry-based protein/peptide identification and post-translational modification analyses (Figure 5.12). These analyses were performed by the University of Birmingham Advanced Mass Spectrometry Facility, and were performed in an unbiased manner, due to the absence of a phosphoenrichment strategy. In turn, the MS1 spectra demonstrated that we had 85.38% sequence coverage of MBP-3C-Sae1$_{pS185}$-His$_6$, and the MS2 spectra of a fragmented V**pS**QGVEDGPDTK phosphopeptide indicated that only the S185 site was phosphorylated (Figure 5.12). Specifically, the y_{11}^+ ion underwent collision-induced dissociation of a phosphate group, while the b_3^{2+} ion possessed an observed mass of 198.08768, which was 40Da larger than the theoretical predicted mass of 158.0832 (Figure 5.12). Given the 2^+ charge, this would in turn, indicate that the mass increased by 80Da, thereby corresponding to the molecular mass of a γ-phosphate group (Figure 5.12). Additionally, we observed a b_3^+ ion that underwent collision-induced phosphate dissociation. Taken together, these results demonstrate with 100% certainty that we purified full-length dual-tagged Sae1$_{pS185}$ that successfully incorporated pS185 in a site-specific manner. Due to time constraints, the relative level of pSer incorporation in the recombinant MBP-3C-Sae1$_{pS185}$-His$_6$ was not analysed, however it is well-documented that both the Chin and Rhinehart pSer OTSs achieved >95% incorporation efficiencies.



Ion	Theoretical mass (Da)	Observed mass (Da)
b_3-H_3PO_4	297.1557	297.1566
b_4-H_3PO_4-NH_3	337.1507	337.1502
b_4-H_3PO_4	354.1772	354.1768
b_5-H_3PO_4	453.2456	453.2476
b_6-H_3PO_4	582.2882	582.2850
b_7-H_3PO_4	697.3151	697.3147
b_{10}-H_3PO_4	966.4163	966.4075
y_{11}-H_3PO_4	1114.5011	1114.4972

Figure 5.12 - Tandem mass spectrometry (MS/MS) validation of site-specific phosphoserine incorporation at Sae1-S185 in recombinant MBP-3C-Sae1-pS185-His6 protein. Mass-to-charge ratio (m/z) relative to ion abundance (intensity) of ions identified from MS2 analysis of a Sae1 VpSQGVEDGPDTK phosphopeptide is indicated (top panel). Neutral loss (-) of H_2, H_2O, H_3PO_4 (P), NH_3, and combined neutral loss of NH_3 and H_3PO_4 (NH_3-P) during collision-induced dissociation is indicated. b^+ (blue) and y^+(red) ions are mapped to the phosphopeptide sequence (top). Experimentally-observed phospho-specific ion masses are compared to theoretical masses (table) calculated using the UCSF Protein Prospector MS-Digest webserver (https://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest). **Sample preparation, MS/MS and data analysis was performed by the University of Birmingham Advanced Mass Spectrometry Facility.**

5.4.2 Purification of Sae1$_{pS185}$:Sae2 SUMO E1 Heterodimers

We, therefore, proceeded with Sae1:Sae2 heterodimer generation. In the first instance, we decided that the best course of action would be to determine the conditions necessary for wild-type heterodimer generation, and then apply the same methodology for the phospho-Sae1:Sae2 heterodimer. Wild-type Sae1 was expressed using the AC6 expression vector and BL21(DE3) parental strain in LB media, while the purification strategy was modified to remove the N-terminal MBP affinity tag and 3C cleavage site (Figure 5.13A). Consequently, untagged Sae1 proteins were purified by initial amylose affinity chromatography, followed by overnight 3C digestion at 4°C, cobalt affinity chromatography, and extensive dialysis to remove imidazole contaminants. The observed molecular weight of purified Sae1-His$_6$ bands were ~45-50kDa, roughly corresponding to the predicted molecular weight of 39.3kDa (Figure 5.13A). An N-terminally His$_6$-tagged Sae2 recombinant protein was expressed using a pET28b-Sae2 construct and BL21(DE3) strain in 2TY, and was purified by cobalt affinity chromatography and extensive dialysis to remove imidazole contaminants. Of note, the pET28b-Sae2 vector was gifted by Dr Alex Lanz. Purification of His$_6$-Sae2 yielded a predominant band with an observed mass of ~100kDa, which roughly corresponded to the predicted molecular weight of ~72.0kDa, and some faint lower bands, which were presumed to be degradation products (Figure 5.13C). In general, the yields of His$_6$-Sae2 were ~1mg/L culture, which is lower than expected, and was rather unstable, since we observed time-dependent dynamic light scattering at near UV wavelengths of 330-350nm. Clearly, these protein instability metrics suggest that either the pET28b-Sae2 expression system is marginally compromised, or the purification strategy requires further optimisation, but this was not further pursued due to time constraints. However, we would propose that the pET28b-Sae2 plasmid is re-sequenced, or Sae proteins are co-

expressed, while purification buffer salt concentration and pH should be explored, as well as addition of cryoprotectants, such as glycerol. We then conducted size-exclusion chromatography of individual Sae proteins, so we could determine the baseline elution volumes, and compared these to the heterodimer elution volume when both proteins were mixed in an equimolar ratio. The peak elution volumes of Sae1-His_6 and His_6-Sae2 alone were 20.7mL and 19.3mL, respectively, and the 1.4mL elution volume difference was expected based on the observed ~50-55kDa larger molecular weight His_6-Sae2 compared to Sae1-His_6 (data not shown). Of note, we observed a large His_6-Sae2 peak elution volume at 14.2mL, which corresponds to our predicted void volume of 14.4mL, thereby demonstrating the aforementioned His_6-Sae2 instability. The Sae1-His_6:His_6-Sae2 SEC profile contained two elution peaks of 17.8mL and 19.7mL. The first peak corresponded to a 1.5mL earlier retention time than His_6-Sae2 alone, while the second peak corresponded to the SEC profile of Sae1-His_6. Additionally, we observed a third peak elution volume of 13.1mL, roughly corresponding to the void volume elution profile observed in the His_6-Sae2 SEC analysis. Consequently, our SEC analyses suggested that we had generated a heterogenous protein population containing both purified Sae1-His_6:His_6-Sae2 heterodimers, Sae1-His_6 alone, and His_6-Sae2 aggregates. We then performed SDS-PAGE analysis of peak elution volume fractions from all aforementioned SEC analyses to validate our data interpretation (Figure 5.13D). Indeed, we observed bands corresponding to His_6-Sae2 in the void volume peak, co-elution of Sae1-His_6 and His_6Sae2 from the first heterodimer peak, and Sae1-His_6 only bands in the second peak, thereby confirming our interpretation of the SEC analyses. Given the overlap in elution volumes between His_6-Sae2 and the Sae1-His_6:His_6-Sae2 heterodimer, we cautiously decided that we would select only heterodimer fractions 17-18 for *in vitro* SUMOylation assays, because Sae1-His_6 alone did not elute at these fractions. Although Sae1-His_6 did not elute at fraction 19, we decided that a 1mL spacer fraction would be sufficient to only obtain

heterodimers. The SEC analysis was then repeated for $Sae1_{pS185}$ alone and $Sae1_{pS185}$:His_6-Sae2 heterodimers (Figure 5.13B,E). The $Sae1_{pS185}$-His_6 was peak elution volume was much broader and eluted earlier than wild-type, spanning 15.2-20.4mL, which suggested that our protein was a heterogenous population of non-phosphorylated and phosphorylated S185, or S185 phosphorylation affects the conformational dynamics, oligomeric status and/or size of Sae1-His_6 protein (data not show). In any case, SDS-PAGE analysis indicated that fractions 15-20 contained $Sae1_{pS185}$-His_6 protein. To generate the $Sae1_{pS185}$-His_6:His_6-Sae2 heterodimer we mixed the proteins in a respective 1:3.3 molar ratio, and we observed a single broad peak, spanning 11.9-18.4mL, which peaked at 16.0mL. We observed co-elution of $Sae1_{pS185}$-His_6 and His_6-Sae2 proteins in fractions 15-17, and to a lesser extent fraction 18, from SDS-PAGE analysis (Figure 5.13E). These analyses indicated that we had successfully purified the $Sae1_{pS185}$-His_6: His_6-Sae2 heterodimer, and again suggested that phosphorylation of S185 may affect the conformational dynamics, size, and oligomeric status of the SUMO E1 heterodimer. Finally, we proposed that separate purification and gel filtration of $Sae1_{(pS185)}$-His_6 and His_6-Sae2 proteins was a direct cause of our low yields and Sae2 aggregation patterns, because both proteins typically exist as obligate heterodimers in the cell. We, therefore, purified all subsequent heterodimers by mixing bacterial pellets, prior to cellular lysis, so that the Sae proteins would heterodimerise early in the purification workflow, thereby stabilising His_6-Sae2. Based on this optimised approach, we managed to increase our phospho-heterodimer yields 25-fold (Figure 5.13F).

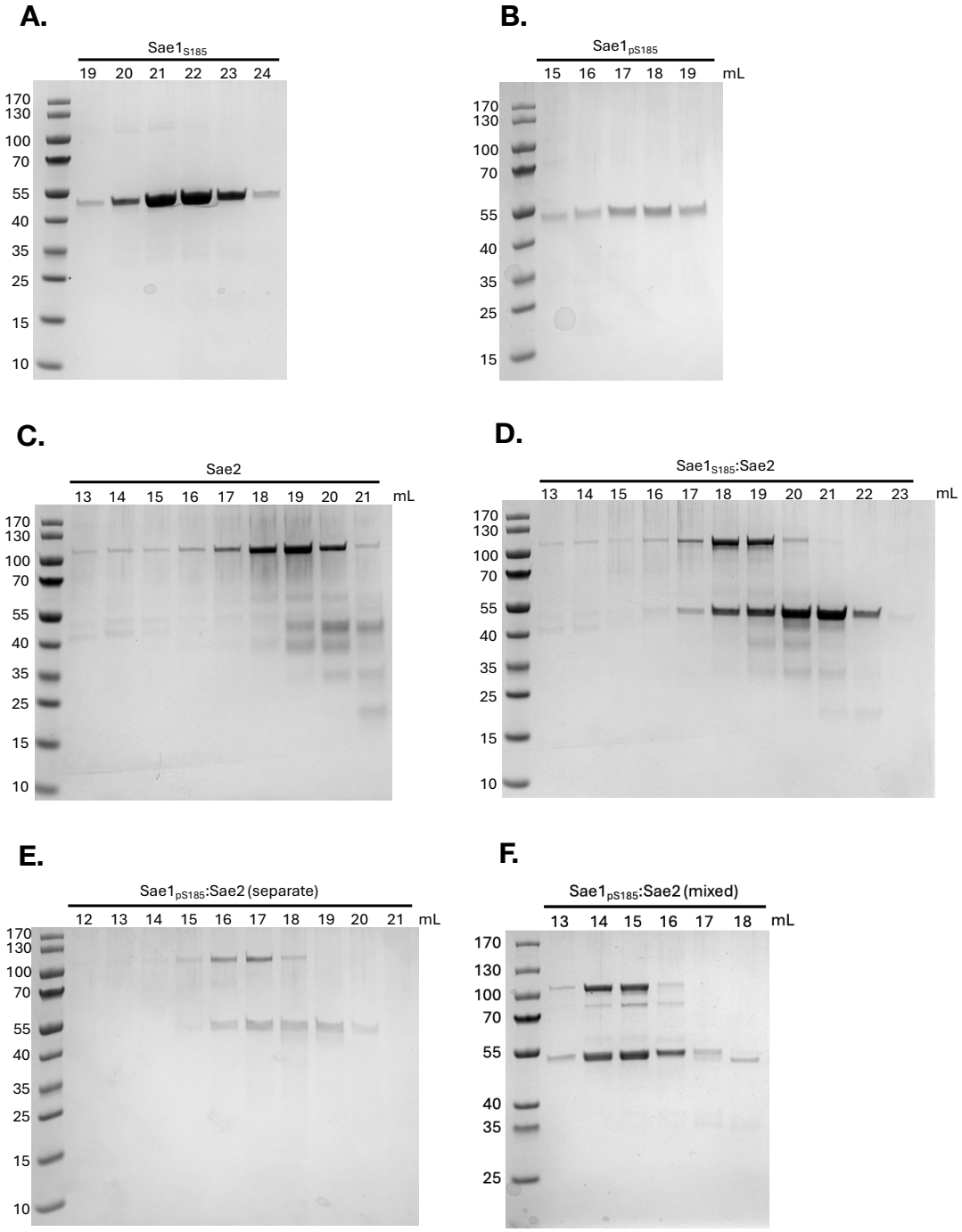


Figure 5.13 - SDS-PAGE analysis of recombinantly purified Sae1-S185 (A,C-D) and Sae1-pS185 (B,C,E-F)-containing SUMO E1 heterodimers. Sae1-pS185 and Sae2 pellets E.coli lysate were mixed together (F) to improve phospho-SUMO E1 heterodimer yields. The elution volume (mL) from SEC analysis is provided.

5.4.3 Recombinant Expression and Purification of Fluorescently-Labelled SUMO 1 and 2 Proteins

Buoyed by the success of wild-type and pS185 Sae1:Sae2 heterodimers, we then set about purification of fluorescently labelled SUMO1 and SUMO2 (Figure 5.14). To achieve this, we sought to purify mature SUMO1 and SUMO2, corresponding to residues 1-97 and 1-93 ($SUMO1_{1\text{-}97}$ and $SUMO2_{1\text{-}93}$) . To conjugate thiol-reactive Alexa Fluor 488 C_5-maleimide and Alexa Fluor 647 C_2-maleimide to the N-terminal region of SUMO1 and SUMO2, respectively, an endogenous C residue in the core α-helix was mutated to an A, while a residue in the extended flexible N-terminal region was mutated to an A. Specifically, these mutations were SUMO1-S9C-C52A ($SUMO1_{1\text{-}97\ S9C\ C52A}$) and SUMO2-A2C-C48A ($SUMO2_{1\text{-}93\ A2C\ C48A}$), as utilised previously to quantify ZNF451-dependent changes in SUMO2 single-turnover conjugation rates of the p53 tetramerisation domain (*Cappadocia, Pichler, and Lima, 2015*). The plasmids encoding SUMO1 and SUMO2 mutants in-frame of an N-terminal GST tag and a TEV cleavage site (ENLYFQ↓G/S), pGEX4T-1- $SUMO1_{1\text{-}97\ S9C\ C52A}$ and pGEX4T-1- $SUMO2_{1\text{-}93\ A2C\ C48A}$, were gifted by the Morris lab. Recombinant SUMO1 and SUMO2 mutant proteins were expressed and then purified by glutathione sepharose chromatography, followed by overnight TEV cleavage at 4°C, and then further purified by size-exclusion chromatography to homogeneity at >90% purity (Figure 5.14). 100μM of purified SUMO proteins were labelled overnight at 4°C with a 10-molar excess of fluorophore-containing thiol-reactive probe, and excess probe was removed the following day using desalting columns. Specifically, $SUMO1_{1\text{-}97\ S9C\ C52A}$ and $SUMO2_{1\text{-}93\ A2C\ C48A}$ were labelled with Alexa Fluor 488 C_5-maleimide and Alexa 647 C_5-maleimide, respectively. The degree of labelling (DOL) for each conjugated protein was calculated by:

$$DOL = \frac{A_{max}}{\varepsilon' \times [SUMO]}, where\ [SUMO] = \frac{A_{280} - (A_{max} \times CF)}{\varepsilon}$$

where A_{max} is the absorbance value at the maximum absorption wavelength for Alexa Fluor 488 (493nm)/647 (651nm) C_5-maleimide, ε' is the extinction coefficient for each dye (Alexa Fluor 488 C_5-maleimide: 72,000 $cm^{-1}M^{-1}$ and Alexa Fluor 647 C_5-maleimide: 265,000 $cm^{-1}M^{-1}$), *CF* is the correction factor for each dye at A_{280} (Alexa Fluor 488 C_5-maleimide: 0.11 and Alexa Fluor 647 C_5-maleimide: 0.03), and ε is the predicted extinction coefficient for each SUMO protein (SUMO1: 4470 $cm^{-1}M^{-1}$ and SUMO2: 1490 $cm^{-1}M^{-1}$). DOL values between 2 and 10 were considered optimal, but this was further validated by the ability of each fluorescent SUMO to be conjugated to RanGAP1.

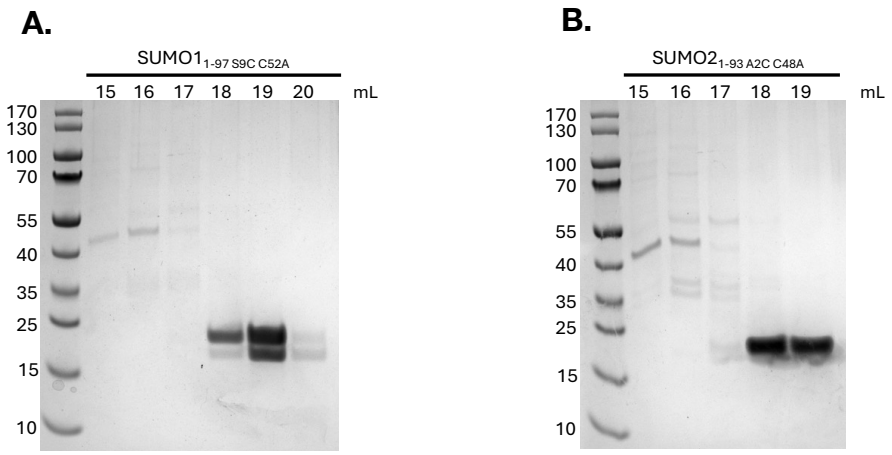


Figure 5.14 - SDS-PAGE analysis of recombinantly purified SUMO 11-97 S9C C52A and SUMO21-93 A2C C48A proteins used for fluorescently labelling with thiol-reactive AlexaFluor 488 and 647 maleimides. The elution volume (mL) from SEC analysis is provided.

5.4.4 Fluorescent Single-Turnover *In Vitro* RanGAP1 SUMOylation Assays Suggest that Sae1-S185 Phosphorylation Does Not Impact SUMO E1 Activity

Following the successful purification of recombinant wild-type and pS185 Sae1:Sae2 heterodimers- and fluorescent AlexaFluor488-SUMO1 and AlexaFluor647-SUMO2 proteins, we proceeded with *in vitro* RanGAP1 SUMOylation assays (Figures 5.15-5.16). Specifically, we conducted minimal *in vitro*

RanGAP1 SUMOylation assays using either mature AlexaFluor488-SUMO1 and AlexaFluor647-SUMO2 as conjugatable modifiers, Ubc9, and the RanGAP1 398-597 C-terminal region ($RanGAP1_{398-597}$) as a substrate. SUMOylation reaction master mixes (n+1 20μL reactions: 100nM $Sae1_{(p)185}$:Sae2, 500nM Ubc9, 10μM $RanGAP1_{398-597}$, 20μM AlexaFluor488-SUMO1/AlexaFluor647-SUMO2) were made up in SUMO Assay Buffer (SAB) (20mM Hepes pH 7.5, 50mM NaCl, 5mM $MgCl_2$) and kept on ice. $RANGAP_1$ SUMOylation reactions were stimulated upon addition of 1mM ATP and warming at 30°C, 400rpm for each given timepoint, whereupon a reaction was quenched by addition of non-reducing sample buffer (62.5mM Tris pH 6.8, 2% (w/v) SDS, 7.5% (w/v) glycerol, 0.005% (w/v) bromophenol blue) and boiling at 95°C for 5 minutes. The time points used were 0mins, 1min, 2.5mins, 5mins, and 10mins. Our 0 mins SUMOylation negative control timepoint absented ATP supplementation and was incubated at 30°C, 400rpm for 10 minutes, in order to measure basal $RanGAP1_{398-597}$ 488/647 fluorescence. Samples were separated by SDS-PAGE and 488/647 fluorescence was quantified using the Bio-Rad ChemiDoc Imaging System, before gels were stained with NEBiotech Quick Coomassie stain to visualise sample loading equivalency, and to determine the observe molecular weights of fluorescently SUMO-conjugated protein bands (Figures 5.15-5.16). Fluorescent images were uniformally thresholded across repeated conditions by applying the same intensity threshold values to each image, to ensure background fluorescence intensity was equally removed, while equally maximising signal intensity. A fluorescence scaling factor was included to account for pipetting error, and was calculated as the fold-change difference between cumulative RFU for each sample and the maximum cumulative RFU across all samples. RanGAP1 SUMOylation Conjugates for each timepoint was calculated as:

$$RanGAP1\ SUMOylation\ Conjugates$$

$$= \frac{AlexaFluor488/647 : SUMO1/2 - RanGAP1_{398-597}\ Scaled\ Fluorescence}{AlexaFluor488/647 : SUMO1/2\ \ Scaled\ Cumulative\ Fluorescence}$$

and these values were normalised against the highest RanGAP1$_{398\text{-}597}$ SUMOylation conjugate value, yielding Normalised RanGAP1 SUMOylation Conjugates. Similarly, Normalised Total SUMOylation Activity was calculated, but instead SUMOylation activity was calculated as:

$$Total\ SUMOylation\ activty = 1 - AlexaFluor488/647 : SUMO1/2\ Scaled\ Fluorescence.$$

Timecourses for both normalised SUMOylation metrics were then fitted using a Michaelis-Menten kinetic model. From these analyses, we found that both recombinant wild-type and pS185 Sae1:Sae2 heterodimers were catalytically active, since RanGAP1$_{398\text{-}597}$ was conjugated with both fluorescent SUMO proteins in an ATP-dependent manner, and minimal E1-E2 conjugation cascades are dependent on E1 adenylation, thioesterification, and transthioesterification activity. Despite reactions did not reach saturation, there was no difference in Normalised RanGAP1 SUMO1ylation Conjugates (wild-type *k*: 68.4nM SUMO1 min^{-1}, pS185 *k*: 86.6nM SUMO1min^{-1}) initial rate constants between wild-type and pS185 Sae1:Sae2 heterodimers (Figure 5.15). Similarly, there was no difference between Sae1:Sae2 heterodimers in normalised fluorescence after 10-minutes across both RanGAP1 and Total SUMO1ylation metrics. These analyses could be improved with a longer end time point, but the inherent difference in activity between SUMO proteins that the labelled AlexaFluor488-SUMO1-A2C-C48A. On the other hand, there was a moderate, albeit non-significant, difference in Normalised RanGAP1 SUMO2ylation Conjugates (wild-type *k*: 1.1μM SUMO2min^{-1}, pS185 *k*: 1.75μM SUMO2min^{-1}) initial rate constants between wild-type and pS185 Sae1:Sae2 heterodimers (Figure 5.16). However, there was no difference between Sae1:Sae2 heterodimers in 10-minute normalised fluorescence across both SUMO2ylation metrics.

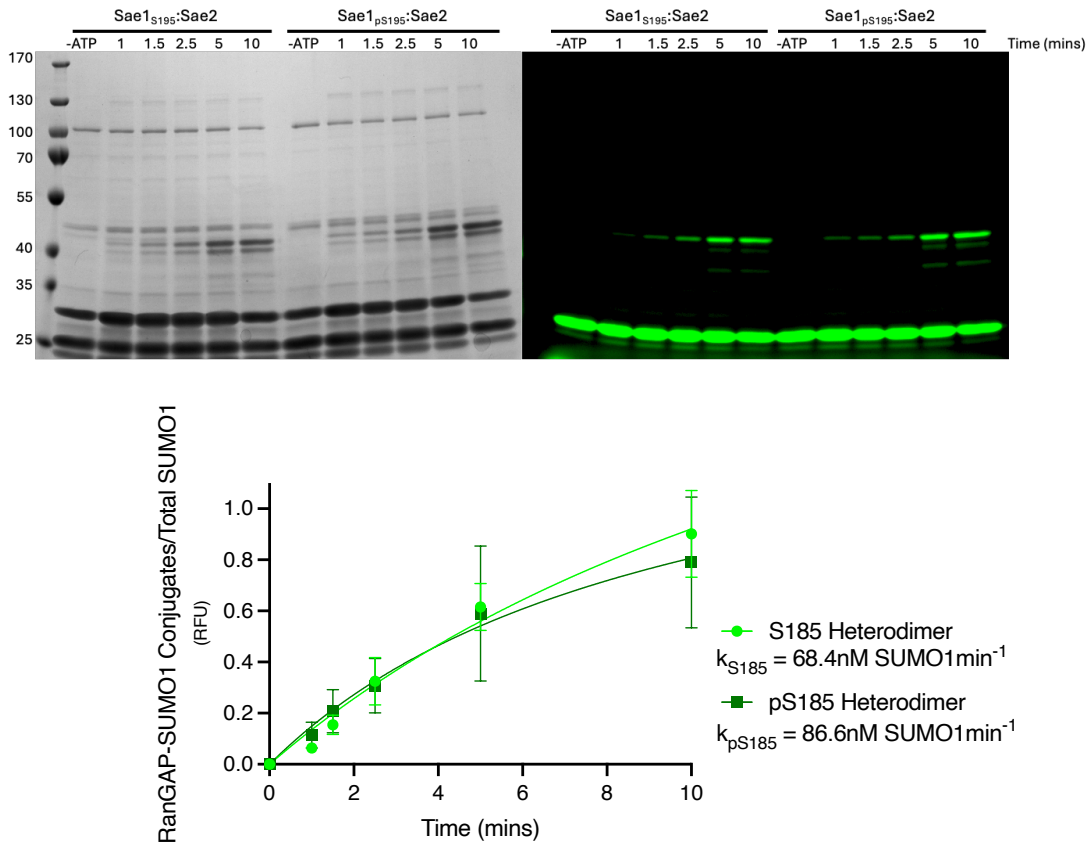


Figure 5.15 - Fluorescent single turnover in vitro RanGAP1 SUMO1ylation assays suggest that Sae1-S185 phosphorylation has no effect on SUMO E1 activity. SDS-PAGE analysis (left panel) and AlexaFluor488 fluorescence imaging (right panel) of a representative in vitro RANGAP1 SUMO1ylation assay timecourse is provided. An ATP control was included for detection of non-specific bands. Experiments were repeated three times. Data points represent the mean of values calculated and the error bars represent the standard deviation. Graphs were fitted using Michaelis-Menten non-linear regression models. Semi-quantitative AlexaFluor488 SUMO1 consumption rates (k) for each SUMO E1 heterodimer were calculated by fitting a linear regression model to the first 2.5 minutes of each assay.

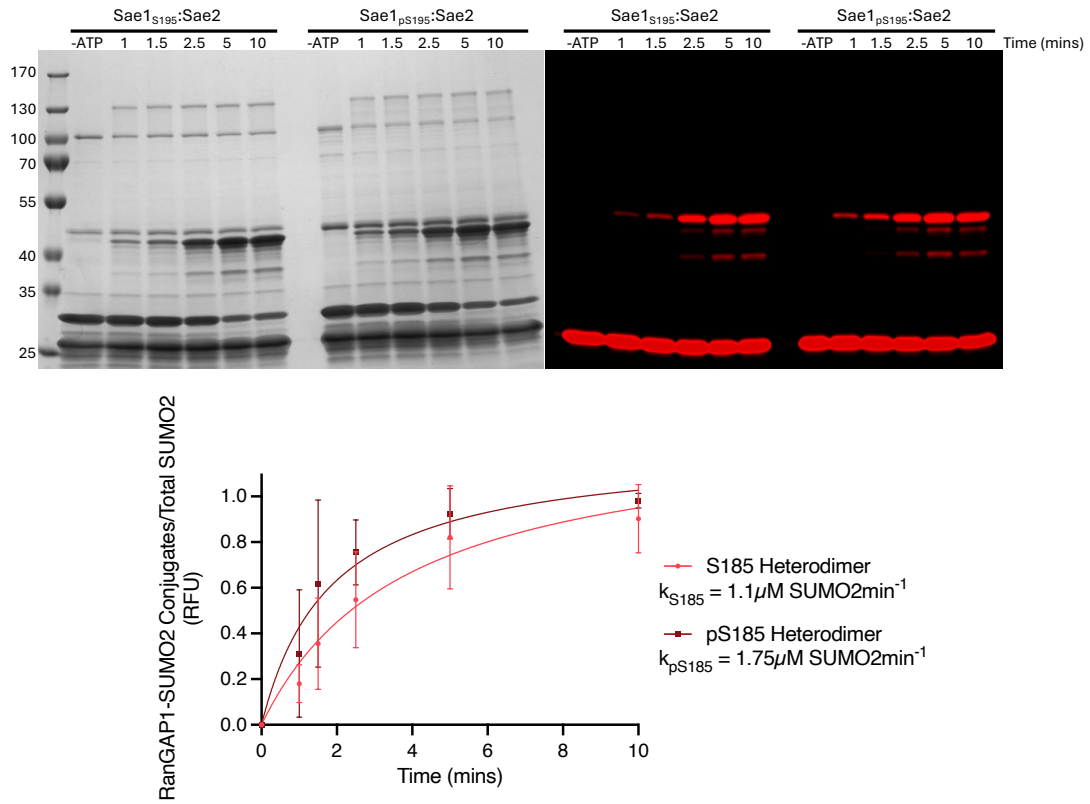


Figure 5.16 - Fluorescent single turnover in vitro RanGAP1 SUMO2ylation assays suggest that Sae1-S185 phosphorylation has no effect on SUMO E1 activity. SDS-PAGE analysis (left panel) and AlexaFluor647 fluorescence imaging (right panel) of a representative in vitro RANGAP1 SUMO1ylation assay timecourse is provided. An ATP control was included for detection of non-specific bands. Experiments were repeated three times. Data points represent the mean of values calculated and the error bars represent the standard deviation. Graphs were fitted using Michaelis-Menten non-linear regression models. Semi-quantitative AlexaFluor647 SUMO2 consumption rates (k) for each SUMO E1 heterodimer were calculated by fitting a linear regression model to the first 2.5 minutes of each assay.

In summary, these integrated analyses of SUMO1/2ylation of RanGAP1 suggests that S185 phosphorylation does not significantly enhance SUMO E1 activity. This aligns with the previous observations showing comparable *in vitro* $RanGAP1_{398-597}$ SUMOylation levels between wild-type and S185 phosphomimetic mutant SUMO E1s after 10 minutes.

Future studies should focus on Michaelis-Menten kinetic characterisation of RanGAP1 SUMOylation by varying RanGAP1 concentrations. This approach would provide detailed insights into how Sae1-

S185 phosphorylation affects SUMO E1 catalytic efficiency. Additionally, targeted SUMO E1-specific assays—such as in vitro Sae1:Sae2 adenylation, thioesterification, and transthioesterification assays—should also be conducted in a similar manner to further validate and expand upon these findings.

In addition, we conducted initial ITC experiments to monitor potential binding between recombinant $Sae1_{pS185}$-His_6 and wild-type SUMO2, thus addressing the potential phospho-SIM mimetic model. We observed no heat exchange upon titration of wild-type SUMO2 up to 500μM, which would suggest that the S185 phosphorylation site does not act as a phospho-SIM mimetic, however recombinant Sae2 alone or Sae1:Sae2 heterodimer positive controls were not included due to time constraints (data not shown). These controls should be included in future experiments to further validate these conclusions.

5.5 Discussion

Our findings provide no evidence that S185 phosphorylation allosterically modulates E1 activity in RanGAP1 SUMOylation assays or affects interactions with SUMO proteins (Figure 5.15-5.16). This conclusion is supported by prior in vitro $RanGAP1_{398–597}$ SUMOy1/2lation assays using wild-type, S185A (phosphonull), and S185D (phosphomimetic) Sae1-containing SUMO E1 heterodimers, which showed no significant differences in $RanGAP1_{398–597}$ SUMO1/2 conjugation levels after 10 minutes. It is important to note that the Sae1-SQ motif is poorly conserved throughout eukaryotes, with conservation only amongst mammals, and conservation of the S185 residue in mammalian, *G. gallus*, and *D. rerio* Sae1 proteins. Consequently, the poor evolutionary conservation of the SQ motif undermines the fundamental importance of ATM and DNA-PK-mediated Sae1-S185

phosphorylation as an essential mitotic SUMO E1 regulatory site involved in DNA damage tolerance. In contrast, the conservation of the SQ motif may suggest that ATM-mediated Sae1-S185 phosphorylation is a unique mammalian-specific event that provides the SUMO E1 with an unknown regulatory advantage that contributes to mitotic SUMO chain formation that is required for DNA damage tolerance. In support of this, Sae1-S185 phosphorylation status was intimately linked to mitotic DNA damage tolerance and mitotic progression in cellular experiments. Additionally, mitotic cells with ML-792-inhibited endogenous SUMO E1, but containing overexpressed ML-792-resistant and Sae1-S185D phosphomimetic-containing SUMO E1 heterodimers formed more SUMO2 chains than wild-type counterparts, while mitotic SUMO2 chain formation was abolished in cells overexpressing ML792-resistant Sae1-S185A phosphonull mutant SUMO E1.

5.5.1 Future Experiments

The role of Sae1-S185 remains enigmatic, and our data is suggestive of an alternative mechanism of phospho-regulation of the SUMO E1 in the proposed mitotic DDR. One possibility is that S185-phosphorylation modulates interactions with an as-yet unidentified protein, potentially influencing SUMOylation dynamics in response to mitotic and DNA damage stimuli. In search of a phosphorylation-dependent interactor, we found that the local sequence surrounding the Sae1-S185 phosphorylation site does not fit the phosphorylation-dependent interaction consensus sequence of several well-characterised phospho-binding module domain families, including the FHA (pThr-dependent), Src Homology 2 (SH2; pY-dependent), PBD (S-pSer-P), and 14-3-3 (R-x-x-pSer/pThr-x-P) domains (*Songyang et al., 1993; Durocher et al,, 2000; Elia, Cantley, and Yaffe, 2003*). It is conceivable that Sae1-S185 phosphorylation could stimulate specific binding with a BRCT or tandem BRCT-containing DDR protein to facilitate SUMO E1 recruitment to DNA damage sites,

reminiscent of the $MDC1_{tBRCT}$:γH2AX, $pTIP_{BRCT5/6}$:γH2AX, $BRCA1_{BRCT}$:BRCTide, and $Crb2_{tBRCT}$:H2A1 complexes (*Williams et al., 2004; Stucki et al., 2005; Kilkenny et al., 2008; Yan et al., 2011*). However, this possibility seems rather unlikely, because of the poor conservation of the pSer +Q residue and specificity of most phospho-binding modules. Nonetheless, the possibility could be specifically addressed by an unbiased mass spectrometry interactomic screens involving pulldowns with Sae1-S185 and Sae1-pS185 proteins from nuclear extracts of DNA damage-treated, mitotic, and DNA damage-treated mitotic cells.

An alternative mode of phosphorylation-mediated regulation of interaction may be responsible for this curious phenomenon. Further sequence analysis of the local sequence surrounding the Sae1-S185 demonstrates that the site lies 10-amino acids upstream of a previously characterised K-[K/R]-x-[K/R] monopartite NLS that is conserved from humans to fruit flies. In addition, the S185 -3 K residue would form a second basic cluster that fits the consensus motif of a [R/K]-x_{10-12}-K-R-x-K bipartite NLS and shares a similar co-evolutionary relationship with S185 (Figure 5.17). Previous reports have demonstrated that the Sae1 monopartite NLS was required for importin α/β-mediated nucleocytoplasmic shuttling of Sae1 alone but was not required for nuclear import of the SUMO E1 heterodimer (*Moutty, Sakin, and Melchior, 2011; Truong et al., 2012*). Nuclear trafficking of the SUMO E1 was instead dependent on the Sae2 C-terminal bipartite NLS, which also contains a stretch of upstream SUMOylated lysines that are involved in nuclear retention of the E1 (*Truong et al., 2012*). Although Sae1 and Sae2 are considered as an obligate heterodimeric complex, both reports consider the possibility that Sae1 and Sae2 could be imported into the nucleus separately and heterodimerise upon release by importins (*Moutty, Sakin, and Melchior, 2011; Truong et al., 2012*). Taken these considerations into account, it raises the exciting possibility that Sae1-S185

phosphorylation may modulate the interaction between the Sae1 monopartite or bipartite NLS and the importin α adaptor component of the importin α/β nuclear receptors. There are at least 7 importin α family members that exhibit similar transport efficiencies of canonical NLS motifs-containing substrates (*Köhler et al., 1999*). With this in mind, initial experiments could biochemically focus on whether Sae1-S185 phosphorylation has any effect on importin binding through direct protein-protein ITC binding experiments, involving importin α1-7 and Sae1-S185 and Sae1-pS185 proteins.

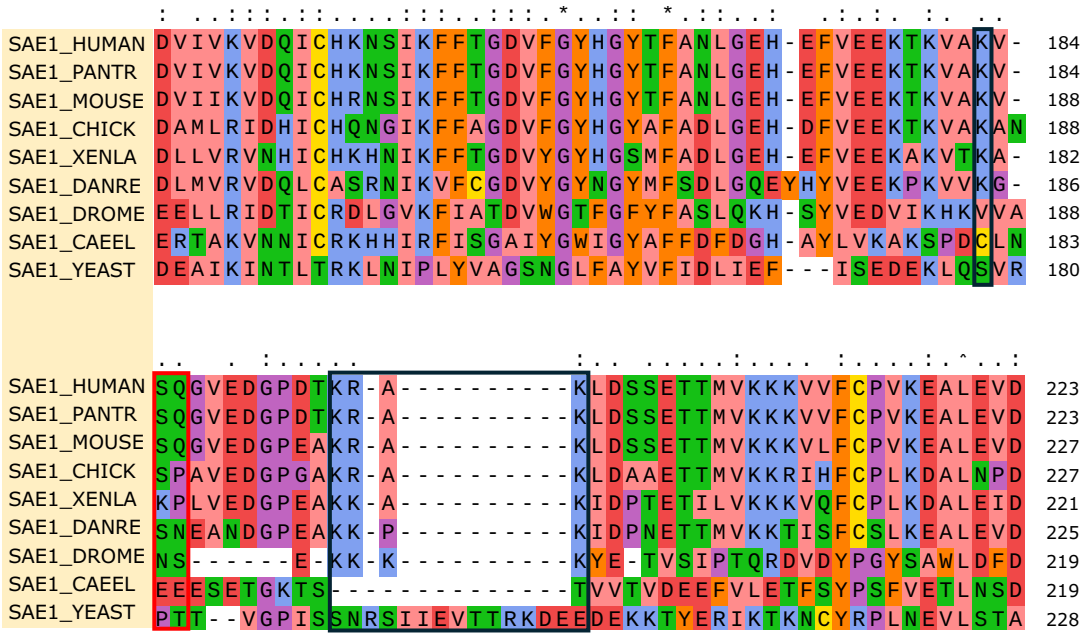


Figure 5.17 - Evolutionary conservation of Sae1-pS185, the mammalian-specific SQ-motif, and the potential animal-specific Sae1 bi-partite NLS. The primary sequence of eukaryotic Sae1 proteins was aligned using the Clustal Omega algorithm (*Sievers et al., 2011*). Amino acids are coloured according to physicochemical properties (red- acidic, blue- basic, green- polar, salmon- hydrophobic, orange- aromatic, purple- conformationally special, and yellow- cysteine). The Sae1-SQ motif is highlighted with a red box and the two basic clusters forming the potential bi-partite NLS are highlighted with a black box. Absolutely conserved amino acids are denoted with an asterisk (*****). Amino acids with similar properties are indicated with a colon (**:**), and weakly similar amino acids are indicated with a full stop (**.**). Figure created with SnapGene software v8.0 (www.snapgene.com).

Chapter 6

Closing Remarks

This thesis has investigated the structural and biochemical basis of phospho-regulation of key DDR-associated proteins that are involved in maintenance of genome stability, with a focus on site-specific Ser/Thr phosphorylation of MDC1-TQxF motifs, Rad50-T690 and Sae1-S185. The core of this work has explored the molecular basis of phospho-epitope recognition by the FHA and PBD reader modules using a combination of biophysical binding experiments, genetic code expansion technologies, structural biology, and AlphaFold structural prediction models. Additionally, this thesis has explored the concept of PTM crosstalk between phosphorylation, ubiquitination, and SUMOylation modular signalling, which is critical for the fine-tuning of DNA damage signalling and DNA repair pathways. Particular attention has focussed on phosphorylation-dependent recruitment of ubiquitin E3 ligases and allosteric phospho-regulation of the SUMO E1 heterodimer.

In the first instance, isothermal titration calorimetry (ITC) binding experiments provided direct evidence that Pellino $1/2_{FHA}$ domains specifically recognise and bind to four distinct MDC1-pTQxF phosphopeptides, and exhibit minor binding preference for MDC1-pT765 (Figure 3.13, Table 3.13). Notably, no detectable binding was observed between Pellino $1/2_{FHA}$ and an experimentally validated γH2AX phosphopeptide (Figure 3.14, Table 3.13). These findings supported cellular data from our collaborators that demonstrated the Pellino 1/2 recruitment to DNA damage sites is directly dependent on binding to ATM-mediated MDC1-TQxF sites (Figures 3.8-11), rather than the ATM-mediated γH2AX pS139 residue, as proposed by a previous Pellino DSB recruitment model (*Ha et al., 2019*). Together, these data outline a simplified model that succinctly establishes a clear link between Pellino recruitment and downstream Pellino RING-mediated K63-linked polyubiquitination of the Nbs1 subunit of the MRN complex, which promotes MRN stabilisation, potentiates ATM signalling, and triggers homologous recombination repair. The most significant achievement of this

thesis was the determination of the crystal structure of the Pellino 1_{FHA}:$MDC1_{pT765}$ phosphopeptide complex (Figure 3.17), which remains the only available structure of a Pellino family member FHA domain bound to a phosphopeptide. This structure not only further demonstrated that Pellino 1 DSB recruitment occurs through MDC1-pTQxF binding in a 1:1 stoichiometry, but it also resolved the long-standing question of how atypical $Pellino_{FHA}$ domains bind to phosphorylated epitopes. Our structural analyses revealed that the basic Pellino FHA core is the only component involved in directly binding to the phosphopeptide (Figure 3.18), with direct pThr γ-phosphate contacts occurring through R104, S135, and R136 (Figure 3.19), which fit within a GR-$[x]_n$-S-[R/K] motif that is highly conserved both in sequence and structure throughout all FHA domain family members (Figure 3.20). Additionally, our structure also revealed that the $Pellino_{FHA}$ 'wing' appendage plays an indirect role in phosphopeptide binding, providing structural support for a unique insertion in the FHA core involved in pThr +3 selectivity (Figure 3.22). The conserved Pellino 1_{FHA} architecture was key for accurate prediction of the experimental structure using AlphaFold models (Figures 3.26-3.27 and Table 3.5), which, at the time this project was initiated, was considered unlikely to reliably predict phosphorylation-dependent complexes. This breakthrough has significant implications, raising the exciting possibility that AlphaFold structural models could be used for future predictions, including Pellino interactions with phosphorylated substrates involved in innate immune signalling. Moreover, AlphaFold could be used to predict the structure of full-length Pellino in complex with a cognate Ub E2 enzyme, Ub, and substrate, as well as the role of allosteric phospho-regulation of E3 activity.

The phosphorylation-dependent interaction between Rad50-pT690 and the PLK1-PBD, which was proposed from published unbiased phosphopeptide interactomics and unpublished SPR data from

our lab, was tested using a combination of biophysical binding experiments (Figures 4.6-4.7,4.11,4.22-4.23), experimental structures (Figure 4.6), AlphaFold predictive structures (Figures 4.9-4.10, 4.12), and chemical cross-linking experiments (Figures 4.27-4.28) (*Dullovi et al., 2023*). Unfortunately, no consistent binding was detected across these approaches. Despite this setback, a notable technical achievement was the determination of the crystal structure of the PBD fragment, solved in four separate instances and two distinct space groups (Figure 4.8). Perhaps this fragment may prove valuable for future studies of PBD:phosphopeptide complexes. A major success in this project was that the successful purification of the recombinant ZCC fragment with site-specific incorporation of a phosphothreonine residue at T690 (Figures 4.16-4.17). This reagent has the potential to serve as a powerful tool for testing of other potential candidates, such as FOXK1 by binding experiments. In addition, the recombinant ZCC-pT690 fragment could be used for the re-identification of a physiological interactor through unbiased phosphoprotein interactomics screens, because it is the most physiologically representative *in vitro* tool available and may also be used for the future characterisation of phosphorylation-dependent conformational in the ZCC. However, it is important to note that further optimisation of the pThr OTS is necessary to achieve higher fidelity and yield of pThr incorporation. One potential optimisation strategy could involve systems-wide analyses. Indeed, a recent study explored the effects of pSer OTS overexpression on global host cell fitness, metabolic resource allocation, and proteome dysregulation, which successfully identified modifiable OTS components (*Mohler et al., 2023*). Upon implementation of these changes, including reducing the pSer-tRNA$_{CUA}$ copy number and pSer-tRNA$_{CUA}$ acceptor stem swap mutations, *Mohler et al.,* observed a significant improvement in host cell fitness, and reduced G and T misincorporation events by 30- and 1000-fold, respectively *(Mohler et al., 2023)*. Ultimately, this

systems-wide analytical approach led to a 6-fold increase in protein yield with a ~50% increase in phosphoprotein purity (*Mohler et al., 2023*).

Finally, the impact of DNA damage and mitosis-dependent phosphorylation of Sae1-S185 by ATM, DNAPK and p38 kinases on the activity of the obligate SUMO E1 heterodimer was investigated. A significant success of this project was the purification of recombinant Sae1 with site-specific incorporation of phosphoserine at S185 (Figures 5.11-5.12), and subsequent purification of the Sae1-pS185:Sae2 heterodimer (Figure 5.13). Importantly, both the monomeric and heterodimeric proteins require further validation by phos-tag SDS-PAGE analyses to determine the relative pSer incorporation efficiency. Unfortunately, no Sae1-S185 phosphorylation-dependent difference in SUMO activity was detected by single turnover fluorescence *in vitro* RanGAP1 SUMO1/2ylation assays (Figures 5.15-5.16). These analyses, along with preliminary data, suggest the possibility of an alternative mode of phospho-regulation of the SUMO E1 holoenzyme. We propose that this regulation may occur through importin α/β nuclear receptors, based on evolutionary conservation of the local S185 sequence and published data that demonstrates a monopartite NLS located 10-amino acids downstream of Sae1-S185 interacts with importin α/β (Figure 5.17). Since importin α is the adapter molecular involved in recognition of the NLS-containing cargo, we propose that this hypothesis could be tested by biophysical binding experiments involving recombinant importin α and Sae1-pS185 proteins.

List of References

Abramson, J., Adler, J., Dunger, J., et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630 (8016): 493–500. doi:10.1038/s41586-024-07487-w.

Abuetabh, Y., Wu, H.H., Chai, C., et al. (2022) DNA damage response revisited: the p53 family and its regulators provide endless cancer therapy opportunities. Experimental & Molecular Medicine, 54 (10): 1658–1669. doi:10.1038/s12276-022-00863-4.

Adams, P.D., Afonine, P. V., Bunkóczi, G., et al. (2010) PHENIX : a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D Biological Crystallography, 66 (2): 213–221. doi:10.1107/S0907444909052925.

Adriaans, I.E., Basant, A., Ponsioen, B., et al. (2019) PLK1 plays dual roles in centralspindlin regulation during cytokinesis. Journal of Cell Biology, 218 (4): 1250–1264. doi:10.1083/jcb.201805036.

Agirre, J., Atanasova, M., Bagdonas, H., et al. (2023) The CCP4 suite: integrative software for macromolecular crystallography. Acta Crystallographica Section D Structural Biology, 79 (6): 449–461. doi:10.1107/S2059798323003595.

Agostinho, M., Santos, V., Ferreira, F., et al. (2008) Conjugation of Human Topoisomerase 2α with Small Ubiquitin-like Modifiers 2/3 in Response to Topoisomerase Inhibitors: Cell Cycle Stage and Chromosome Domain Specificity. Cancer Research, 68 (7): 2409–2418. doi:10.1158/0008-5472.CAN-07-2092.

Ahn, J.-Y., Li, X., Davis, H.L., et al. (2002) Phosphorylation of Threonine 68 Promotes Oligomerization and Autophosphorylation of the Chk2 Protein Kinase via the Forkhead-associated Domain. Journal of Biological Chemistry, 277 (22): 19389–19395. doi:10.1074/jbc.M200822200.

Anamika, A. and Spyracopoulos, L. (2016) Molecular Basis for Phosphorylation-dependent SUMO Recognition by the DNA Repair Protein RAP80. Journal of Biological Chemistry, 291 (9): 4417–4428. doi:10.1074/jbc.M115.705061.

Andrysik, Z., Bernstein, W.Z., Deng, L., et al. (2010) The novel mouse Polo-like kinase 5 responds to DNA damage and localizes in the nucleolus. Nucleic Acids Research, 38 (9): 2931–2943. doi:10.1093/nar/gkq011.

Asea, A., Rehli, M., Kabingu, E., et al. (2002) Novel Signal Transduction Pathway Utilized by Extracellular HSP70. Journal of Biological Chemistry, 277 (17): 15028–15034. doi:10.1074/jbc.M200497200.

Ayaydin, F. and Dasso, M. (2004) Distinct In Vivo Dynamics of Vertebrate SUMO Paralogues. Molecular Biology of the Cell, 15 (12): 5208–5218. doi:10.1091/mbc.e04-07-0589.

Bagdonas, H., Fogarty, C.A., Fadda, E., et al. (2021) The case for post-predictional modifications in the AlphaFold Protein Structure Database. Nature Structural & Molecular Biology, 28 (11): 869–870. doi:10.1038/s41594-021-00680-9.

Bahassi, E.M., Conn, C.W., Myer, D.L., et al. (2002) Mammalian Polo-like kinase 3 (Plk3) is a multifunctional protein involved in stress response pathways. Oncogene, 21 (43): 6633–6640. doi:10.1038/sj.onc.1205850.

Baumann, P., Benson, F.E. and West, S.C. (1996) Human Rad51 Protein Promotes ATP-Dependent Homologous Pairing and Strand Transfer Reactions In Vitro. Cell, 87 (4): 757–766. doi:10.1016/S0092-8674(00)81394-X.

Bayer, P., Arndt, A., Metzger, S., et al. (1998) Structure determination of the small ubiquitin-related modifier SUMO-1. Journal of Molecular Biology, 280 (2): 275–286. doi:10.1006/jmbi.1998.1839.

Beauclair, G., Bridier-Nahmias, A., Zagury, J.-F., et al. (2015) JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs. Bioinformatics, 31 (21): 3483–3491. doi:10.1093/bioinformatics/btv403.

Becker, J.R., Clifford, G., Bonnet, C., et al. (2021) BARD1 reads H2A lysine 15 ubiquitination to direct homologous recombination. Nature, 596 (7872): 433–437. doi:10.1038/s41586-021-03776-w.

Beg, A.A., Ruben, S.M., Scheinman, R.I., et al. (1992) I kappa B interacts with the nuclear localization sequences of the subunits of NF-kappa B: a mechanism for cytoplasmic retention. Genes & Development, 6 (10): 1899–1913. doi:10.1101/gad.6.10.1899.

Benada, J., Burdová, K., Lidak, T., et al. (2015) Polo-like kinase 1 inhibits DNA damage response during mitosis. Cell Cycle, 14 (2): 219–231. doi:10.4161/15384101.2014.977067.

Bennetzen, M. V., Larsen, D.H., Bunkenborg, J., et al. (2010) Site-specific Phosphorylation Dynamics of the Nuclear Proteome during the DNA Damage Response. Molecular & Cellular Proteomics, 9 (6): 1314–1323. doi:10.1074/mcp.M900616-MCP200.

Bernier-Villamor, V., Sampson, D.A., Matunis, M.J., et al. (2002) Structural Basis for E2-Mediated SUMO Conjugation Revealed by a Complex between Ubiquitin-Conjugating Enzyme Ubc9 and RanGAP1. Cell, 108 (3): 345–356. doi:10.1016/S0092-8674(02)00630-X.

Bilbrough, T., Piemontese, E. and Seitz, O. (2022) Dissecting the role of protein phosphorylation: a chemical biology toolbox. Chemical Society Reviews, 51 (13): 5691–5730. doi:10.1039/D1CS00991E.

Botos, I., Segal, D.M. and Davies, D.R. (2011) The Structural Biology of Toll-like Receptors. Structure, 19 (4): 447–459. doi:10.1016/j.str.2011.02.004.

Bouchard, D., Wang, W., Yang, W.-C., et al. (2021) SUMO paralogue–specific functions revealed through systematic analysis of human knockout cell lines and gene expression data. Molecular Biology of the Cell, 32 (19): 1849–1866. doi:10.1091/mbc.E21-01-0031.

Boucher, L.E., Reiter, K.H., Matunis, M.J., et al. (2015) Crystal structure of Human SUMO E2-conjugating enzyme (Ubc9) in complex with E1-activating enzyme (Uba2) ubiquitin fold domain (Ufd). doi:10.2210/pdb4w5v/pdb.

Boyer, P.D., DeLuca, M., Ebner, K.E., et al. (1962) Identification of Phosphohistidine in Digests from a Probable Intermediate of Oxidative Phosphorylation. Journal of Biological Chemistry, 237 (10): PC3306–PC3308. doi:10.1016/S0021-9258(18)50167-8.

Brown, K., Gerstberger, S., Carlson, L., et al. (1995) Control of IκB-α Proteolysis by Site-Specific, Signal-Induced Phosphorylation. Science, 267 (5203): 1485–1488. doi:10.1126/science.7878466.

Brzostek-Racine, S., Gordon, C., Van Scoy, S., et al. (2011) The DNA Damage Response Induces IFN. The Journal of Immunology, 187 (10): 5336–5345. doi:10.4049/jimmunol.1100040.

Burki, F., Roger, A.J., Brown, M.W., et al. (2020) The New Tree of Eukaryotes. Trends in Ecology & Evolution, 35 (1): 43–55. doi:10.1016/j.tree.2019.08.008.

Burma, S., Chen, B.P., Murphy, M., et al. (2001) ATM Phosphorylates Histone H2AX in Response to DNA Double-strand Breaks. Journal of Biological Chemistry, 276 (45): 42462–42467. doi:10.1074/jbc.C100466200.

Burnett, G. and Kennedy, E.P. (1954) THE ENZYMATIC PHOSPHORYLATION OF PROTEINS. Journal of Biological Chemistry, 211 (2): 969–980. doi:10.1016/S0021-9258(18)71184-8.

Burroughs, A.M., Iyer, L.M. and Aravind, L. (2009) Natural history of the E1-like superfamily: Implication for adenylation, sulfur transfer, and ubiquitin conjugation. Proteins: Structure, Function, and Bioinformatics, 75 (4): 895–910. doi:10.1002/prot.22298.

Cai, Z., Chehab, N.H. and Pavletich, N.P. (2009a) Structure and Activation Mechanism of the CHK2 DNA Damage Checkpoint Kinase. Molecular Cell, 35 (6): 818–829. doi:10.1016/j.molcel.2009.09.007.

Cai, Z., Chehab, N.H. and Pavletich, N.P. (2009b) Structure and Activation Mechanism of the CHK2 DNA Damage Checkpoint Kinase. Molecular Cell, 35 (6): 818–829. doi:10.1016/j.molcel.2009.09.007.

Canman, C.E., Lim, D.-S., Cimprich, K.A., et al. (1998) Activation of the ATM Kinase by Ionizing Radiation and Phosphorylation of p53. Science, 281 (5383): 1677–1679. doi:10.1126/science.281.5383.1677.

Capili, A.D. and Lima, C.D. (2007) Structure and Analysis of a Complex between SUMO and Ubc9 Illustrates Features of a Conserved E2-Ubl Interaction. Journal of Molecular Biology, 369 (3): 608–618. doi:10.1016/j.jmb.2007.04.006.

Cappadocia, L. and Lima, C.D. (2018) Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. Chemical Reviews, 118 (3): 889–918. doi:10.1021/acs.chemrev.6b00737.

Cappadocia, L., Pichler, A. and Lima, C.D. (2015a) Structural basis for catalytic activation by the human ZNF451 SUMO E3 ligase. Nature Structural & Molecular Biology, 22 (12): 968–975. doi:10.1038/nsmb.3116.

Cappadocia, L., Pichler, A. and Lima, C.D. (2015b) Structural basis for catalytic activation by the human ZNF451 SUMO E3 ligase. Nature Structural & Molecular Biology, 22 (12): 968–975. doi:10.1038/nsmb.3116.

de Cárcer, G., Escobar, B., Higuero, A.M., et al. (2011a) Plk5, a Polo Box Domain-Only Protein with Specific Roles in Neuron Differentiation and Glioblastoma Suppression. Molecular and Cellular Biology, 31 (6): 1225–1239. doi:10.1128/MCB.00607-10.

de Cárcer, G., Manning, G. and Malumbres, M. (2011b) From Plk1 to Plk5. Cell Cycle, 10 (14): 2255–2262. doi:10.4161/cc.10.14.16494.

Ceccaldi, R., Rondinelli, B. and D'Andrea, A.D. (2016) Repair Pathway Choices and Consequences at the Double-Strand Break. Trends in Cell Biology, 26 (1): 52–64. doi:10.1016/j.tcb.2015.07.009.

Chakravartty, V. and Cronan, J.E. (2015) A series of medium and high copy number arabinose-inducible Escherichia coli expression vectors compatible with pBR322 and pACYC184. Plasmid, 81: 21–26. doi:10.1016/j.plasmid.2015.03.001.

Chang, C.-C., Naik, M.T., Huang, Y.-S., et al. (2011) Structural and Functional Roles of Daxx SIM Phosphorylation in SUMO Paralog-Selective Binding and Apoptosis Modulation. Molecular Cell, 42 (1): 62–74. doi:10.1016/j.molcel.2011.02.022.

Chang, M., Jin, W. and Sun, S.-C. (2009) Peli1 facilitates TRIF-dependent Toll-like receptor signaling and proinflammatory cytokine production. Nature Immunology, 10 (10): 1089–1095. doi:10.1038/ni.1777.

Chapman, J.R. and Jackson, S.P. (2008) Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage. EMBO reports, 9 (8): 795–801. doi:10.1038/embor.2008.103.

Chase, D., Serafinas, C., Ashcroft, N., et al. (2000) The polo-like kinase PLK-1 is required for nuclear envelope breakdown and the completion of meiosis inCaenorhabditis elegans. genesis, 26 (1): 26–41. doi:10.1002/(SICI)1526-968X(200001)26:1<26::AID-GENE6>3.0.CO;2-O.

Chen, B.-R., Wang, Y., Shen, Z.-J., et al. (2021) The RNF8 and RNF168 Ubiquitin Ligases Regulate Pro- and Anti-Resection Activities at Broken DNA Ends During Non-Homologous End Joining. DNA Repair, 108: 103217. doi:10.1016/j.dnarep.2021.103217.

Chen, Y. and Topp, E.M. (2019) Photolytic Labeling and Its Applications in Protein Drug Discovery and Development. Journal of Pharmaceutical Sciences, 108 (2): 791–797. doi:10.1016/j.xphs.2018.10.017.

Cheng, K.-Y., Lowe, E.D., Sinclair, J., et al. (2003) The crystal structure of the human polo-like kinase-1 polo box domain and its phospho-peptide complex. The EMBO Journal, 22 (21): 5757–5768. doi:10.1093/emboj/cdg558.

Choi, K.-C., Lee, Y.S., Lim, S., et al. (2006) Smad6 negatively regulates interleukin 1-receptor–Toll-like receptor signaling through direct interaction with the adaptor Pellino-1. Nature Immunology, 7 (10): 1057–1065. doi:10.1038/ni1383.

Choudhury, B.K. and Li, S.S.-L. (1997) Identification and Characterization of the SMT3 cDNA and Gene from NematodeCaenorhabditis elegans. Biochemical and Biophysical Research Communications, 234 (3): 788–791. doi:10.1006/bbrc.1997.6709.

Chow, J.C., Young, D.W., Golenbock, D.T., et al. (1999) Toll-like Receptor-4 Mediates Lipopolysaccharide-induced Signal Transduction. Journal of Biological Chemistry, 274 (16): 10689–10692. doi:10.1074/jbc.274.16.10689.

Cizmecioglu, O., Krause, A., Bahtz, R., et al. (2012) Plk2 regulates centriole duplication through phosphorylation-mediated degradation of Fbxw7 (human Cdc4). Journal of Cell Science, 125 (4): 981–992. doi:10.1242/jcs.095075.

Clay, F.J., McEwen, S.J., Bertoncello, I., et al. (1993) Identification and cloning of a protein kinase-encoding mouse gene, Plk, related to the polo gene of Drosophila. Proceedings of the National Academy of Sciences, 90 (11): 4882–4886. doi:10.1073/pnas.90.11.4882.

Collins, P.L., Purman, C., Porter, S.I., et al. (2020) DNA double-strand breaks induce H2Ax phosphorylation domains in a contact-dependent manner. Nature Communications, 11 (1): 3158. doi:10.1038/s41467-020-16926-x.

Cori, G.T. and Green, A.A. (1943) CRYSTALLINE MUSCLE PHOSPHORYLASE. Journal of Biological Chemistry, 151 (1): 31–38. doi:10.1016/S0021-9258(18)72111-X.

Corry, P.M. and Cole, A. (1968a) Radiation-Induced Double-Strand Scission of the DNA of Mammalian Metaphase Chromosomes. Radiation Research, 36 (3): 528. doi:10.2307/3572586.

Corry, P.M. and Cole, A. (1968b) Radiation-Induced Double-Strand Scission of the DNA of Mammalian Metaphase Chromosomes. Radiation Research, 36 (3): 528. doi:10.2307/3572586.

Crasta, K., Ganem, N.J., Dagher, R., et al. (2012) DNA breaks and chromosome pulverization from errors in mitosis. Nature, 482 (7383): 53–58. doi:10.1038/nature10802.

D'Amours, D. and Jackson, S.P. (2002) The MRE11 complex: at the crossroads of DNA repair and checkpoint signalling. Nature Reviews Molecular Cell Biology, 3 (5): 317–327. doi:10.1038/nrm805.

Davis, M.W. and Jorgensen, E.M. (2022) ApE, A Plasmid Editor: A Freely Available DNA Manipulation and Visualization Program. Frontiers in Bioinformatics, 2. doi:10.3389/fbinf.2022.818619.

Dephoure, N., Zhou, C., Villén, J., et al. (2008) A quantitative atlas of mitotic phosphorylation. Proceedings of the National Academy of Sciences, 105 (31): 10762–10767. doi:10.1073/pnas.0805139105.

Déry, U., Coulombe, Y., Rodrigue, A., et al. (2008) A Glycine-Arginine Domain in Control of the Human MRE11 DNA Repair Protein. Molecular and Cellular Biology, 28 (9): 3058–3069. doi:10.1128/MCB.02025-07.

Desterro, J.M.P., Rodriguez, M.S., Kemp, G.D., et al. (1999) Identification of the Enzyme Required for Activation of the Small Ubiquitin-like Protein SUMO-1. Journal of Biological Chemistry, 274 (15): 10618–10624. doi:10.1074/jbc.274.15.10618.

Dohmen, R.J., Stappen, R., McGrath, J.P., et al. (1995) An Essential Yeast Gene Encoding a Homolog of Ubiquitin-activating Enzyme. Journal of Biological Chemistry, 270 (30): 18099–18109. doi:10.1074/jbc.270.30.18099.

Dullovi, A., Ozgencil, M., Rajvee, V., et al. (2023) Microtubule-associated proteins MAP7 and MAP7D1 promote DNA double-strand break repair in the G1 cell cycle phase. iScience, 26 (3): 106107. doi:10.1016/j.isci.2023.106107.

Dupré, A., Boyer-Chatenet, L. and Gautier, J. (2006) Two-step activation of ATM by DNA and the Mre11–Rad50–Nbs1 complex. Nature Structural & Molecular Biology, 13 (5): 451–457. doi:10.1038/nsmb1090.

Durocher, D., Taylor, I.A., Sarbassova, D., et al. (2000) The Molecular Basis of FHA Domain:Phosphopeptide Binding Specificity and Implications for Phospho-Dependent Signaling Mechanisms. Molecular Cell, 6 (5): 1169–1182. doi:10.1016/S1097-2765(00)00114-3.

Echols, N., Moriarty, N.W., Klei, H.E., et al. (2014) Automating crystallographic structure solution and refinement of protein–ligand complexes. Acta Crystallographica Section D Biological Crystallography, 70 (1): 144–154. doi:10.1107/S139900471302748X.

Ehlén, Å., Martin, C., Miron, S., et al. (2020) Proper chromosome alignment depends on BRCA2 phosphorylation by PLK1. Nature Communications, 11 (1): 1819. doi:10.1038/s41467-020-15689-9.

Elia, A.E.H., Cantley, L.C. and Yaffe, M.B. (2003a) Proteomic Screen Finds pSer/pThr-Binding Domain Localizing Plk1 to Mitotic Substrates. Science, 299 (5610): 1228–1231. doi:10.1126/science.1079079.

Elia, A.E.H., Rellos, P., Haire, L.F., et al. (2003b) The Molecular Basis for Phosphodependent Substrate Targeting and Regulation of Plks by the Polo-Box Domain. Cell, 115 (1): 83–95. doi:10.1016/S0092-8674(03)00725-6.

Elledge, S.J. (1996) Cell Cycle Checkpoints: Preventing an Identity Crisis. Science, 274 (5293): 1664–1672. doi:10.1126/science.274.5293.1664.

Emsley, P., Lohkamp, B., Scott, W.G., et al. (2010) Features and development of Coot. Acta Crystallographica Section D Biological Crystallography, 66 (4): 486–501. doi:10.1107/S0907444910007493.

Esashi, F., Christ, N., Gannon, J., et al. (2005) CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. Nature, 434 (7033): 598–604. doi:10.1038/nature03404.

Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallographica Section D Biological Crystallography, 62 (1): 72–82. doi:10.1107/S0907444905036693.

Evans, P.R. and Murshudov, G.N. (2013) How good are my data and what is the resolution? Acta Crystallographica Section D Biological Crystallography, 69 (7): 1204–1214. doi:10.1107/S0907444913000061.

Evdokimov, E., Sharma, P., Lockett, S.J., et al. (2008) Loss of SUMO1 in mice affects RanGAP1 localization and formation of PML nuclear bodies, but is not lethal as it can be compensated by SUMO2 or SUMO3. Journal of Cell Science, 121 (24): 4106–4113. doi:10.1242/jcs.038570.

Fields, J.K., Günther, S. and Sundberg, E.J. (2019) Structural Basis of IL-1 Family Cytokine Signaling. Frontiers in Immunology, 10. doi:10.3389/fimmu.2019.01412.

Fischer, E.H., Graves, D.J., Crittenden, E.R.S., et al. (1959) Structure of the Site Phosphorylated in the Phosphorylase b to a Reaction. Journal of Biological Chemistry, 234 (7): 1698–1704. doi:10.1016/S0021-9258(18)69910-7.

Fischer, E.H. and Krebs, E.G. (1955) CONVERSION OF PHOSPHORYLASE b TO PHOSPHORYLASE a IN MUSCLE EXTRACTS. Journal of Biological Chemistry, 216 (1): 121–132. doi:10.1016/S0021-9258(19)52289-X.

Fradet-Turcotte, A., Canny, M.D., Escribano-Díaz, C., et al. (2013) 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. Nature, 499 (7456): 50–54. doi:10.1038/nature12318.

Franchin, C., Cesaro, L., Salvi, M., et al. (2015) Quantitative analysis of a phosphoproteome readily altered by the protein kinase CK2 inhibitor quinalizarin in HEK-293T cells. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1854 (6): 609–623. doi:10.1016/j.bbapap.2014.09.017.

Fuhs, S.R., Meisenhelder, J., Aslanian, A., et al. (2015) Monoclonal 1- and 3-Phosphohistidine Antibodies: New Tools to Study Histidine Phosphorylation. Cell, 162 (1): 198–210. doi:10.1016/j.cell.2015.05.046.

Gali, H., Juhasz, S., Morocz, M., et al. (2012) Role of SUMO modification of human PCNA at stalled replication fork. Nucleic Acids Research, 40 (13): 6049–6059. doi:10.1093/nar/gks256.

Ganguli, S., Shah, A., Hamid, A., et al. (2020) A high energy phosphate jump - From pyrophospho-inositol to pyrophospho-serine. Advances in Biological Regulation, 75: 100662. doi:10.1016/j.jbior.2019.100662.

Gareau, J.R., Reverter, D. and Lima, C.D. (2012) Determinants of Small Ubiquitin-like Modifier 1 (SUMO1) Protein Specificity, E3 Ligase, and SUMO-RanGAP1 Binding Activities of Nucleoporin RanBP2. Journal of Biological Chemistry, 287 (7): 4740–4751. doi:10.1074/jbc.M111.321141.

Gärtner, A., Wagner, K., Hölper, S., et al. (2018) Acetylation of <scp>SUMO</scp> 2 at lysine 11 favors the formation of non-canonical <scp>SUMO</scp> chains. EMBO reports, 19 (11). doi:10.15252/embr.201846117.

Garvin, A.J., Lanz, A.J., Starowicz, K., et al. (2022) SUMO4 promotes SUMO deconjugation required for DNA double-strand break repair. doi:10.1101/2022.03.23.485504.

Garvin, A.J. and Morris, J.R. (2017) SUMO, a small, but powerful, regulator of double-strand break repair. Philosophical Transactions of the Royal Society B: Biological Sciences, 372 (1731): 20160281. doi:10.1098/rstb.2016.0281.

Gary, C.R., Acharige, N.P.N., Oyewumi, T.O., et al. (2024) Kinase-catalyzed biotinylation for discovery and validation of substrates to multispecificity kinases NME1 and NME2. Journal of Biological Chemistry, 300 (8): 107588. doi:10.1016/j.jbc.2024.107588.

Gatti, M., Pinato, S., Maiolica, A., et al. (2015) RNF168 Promotes Noncanonical K27 Ubiquitination to Signal DNA Damage. Cell Reports, 10 (2): 226–238. doi:10.1016/j.celrep.2014.12.021.

Gelot, C., Kovacs, M.T., Miron, S., et al. (2023) Polθ is phosphorylated by PLK1 to repair double-strand breaks in mitosis. Nature, 621 (7978): 415–422. doi:10.1038/s41586-023-06506-6.

Gheghiani, L., Loew, D., Lombard, B., et al. (2017) PLK1 Activation in Late G2 Sets Up Commitment to Mitosis. Cell Reports, 19 (10): 2060–2073. doi:10.1016/j.celrep.2017.05.031.

Golsteyn, R.M., Schultz, S.J., Bartek, J., et al. (1994) Cell cycle analysis and chromosomal localization of human Plk1, a putative homologue of the mitotic kinases Drosophila polo and Saccharomyces cerevisiae Cdc5. Journal of Cell Science, 107 (6): 1509–1517. doi:10.1242/jcs.107.6.1509.

Gong, L., Li, B., Millas, S., et al. (1999) Molecular cloning and characterization of human AOS1 and UBA2, components of the sentrin-activating enzyme complex. FEBS Letters, 448 (1): 185–189. doi:10.1016/S0014-5793(99)00367-1.

Gong, L., Millas, S., Maul, G.G., et al. (2000) Differential Regulation of Sentrinized Proteins by a Novel Sentrin-specific Protease. Journal of Biological Chemistry, 275 (5): 3355–3359. doi:10.1074/jbc.275.5.3355.

Großhans, J., Schnorrer, F. and Nüsslein-Volhard, C. (1999) Oligomerisation of Tube and Pelle leads to nuclear localisation of Dorsal. Mechanisms of Development, 81 (1–2): 127–138. doi:10.1016/S0925-4773(98)00236-6.

Grosstessner-Hain, K., Hegemann, B., Novatchkova, M., et al. (2011) Quantitative Phospho-proteomics to Investigate the Polo-like Kinase 1-Dependent Phospho-proteome. Molecular & Cellular Proteomics, 10 (11): M111.008540. doi:10.1074/mcp.M111.008540.

Guderian, G., Westendorf, J., Uldschmid, A., et al. (2010) Plk4 trans -autophosphorylation regulates centriole number by controlling βTrCP-mediated degradation. Journal of Cell Science, 123 (13): 2163–2169. doi:10.1242/jcs.068502.

Guo, D., Li, M., Zhang, Y., et al. (2004) A functional variant of SUMO4, a new IκBα modifier, is associated with type 1 diabetes. Nature Genetics, 36 (8): 837–841. doi:10.1038/ng1391.

Guzzo, C.M., Ringel, A., Cox, E., et al. (2014) Characterization of the SUMO-Binding Activity of the Myeloproliferative and Mental Retardation (MYM)-Type Zinc Fingers in ZNF261 and ZNF198. PLoS ONE, 9 (8): e105271. doi:10.1371/journal.pone.0105271.

Ha, G.-H., Ji, J.-H., Chae, S., et al. (2019) Pellino1 regulates reversible ATM activation via NBS1 ubiquitination at DNA double-strand breaks. Nature Communications, 10 (1): 1577. doi:10.1038/s41467-019-09641-9.

Haas, A.L., Warms, J. V, Hershko, A., et al. (1982) Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. The Journal of biological chemistry, 257 (5): 2543–8.

Habedanck, R., Stierhof, Y.-D., Wilkinson, C.J., et al. (2005) The Polo kinase Plk4 functions in centriole duplication. Nature Cell Biology, 7 (11): 1140–1146. doi:10.1038/ncb1320.

Häcker, H., Redecke, V., Blagoev, B., et al. (2006) Specificity in Toll-like receptor signalling through distinct effector functions of TRAF3 and TRAF6. Nature, 439 (7073): 204–207. doi:10.1038/nature04369.

Haghayeghi, A., Sarac, A., Czerniecki, S., et al. (2010) Pellino enhances innate immunity in Drosophila. Mechanisms of Development, 127 (5–6): 301–307. doi:10.1016/j.mod.2010.01.004.

Hang, J. and Dasso, M. (2002) Association of the Human SUMO-1 Protease SENP2 with the Nuclear Pore. Journal of Biological Chemistry, 277 (22): 19961–19966. doi:10.1074/jbc.M201799200.

Hanks, S.K. and Hunter, T. (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification 1. The FASEB Journal, 9 (8): 576–596. doi:10.1096/fasebj.9.8.7768349.

Hanks, S.K., Quinn, A.M. and Hunter, T. (1988) The Protein Kinase Family: Conserved Features and Deduced Phylogeny of the Catalytic Domains. Science, 241 (4861): 42–52. doi:10.1126/science.3291115.

Hann, Z.S., Ji, C., Olsen, S.K., et al. (2019) Structural basis for adenylation and thioester bond formation in the ubiquitin E1. Proceedings of the National Academy of Sciences, 116 (31): 15475–15484. doi:10.1073/pnas.1905488116.

Hannich, J.T., Lewis, A., Kroetz, M.B., et al. (2005) Defining the SUMO-modified Proteome by Multiple Approaches in Saccharomyces cerevisiae*. Journal of Biological Chemistry, 280 (6): 4102–4110. doi:10.1074/jbc.M413209200.

Hao, B., Zhao, G., Kang, P.T., et al. (2004) Reactivity and Chemical Synthesis of L-Pyrrolysine— the 22nd Genetically Encoded Amino Acid. Chemistry & Biology, 11 (9): 1317–1324. doi:10.1016/j.chembiol.2004.07.011.

Hardman, G., Perkins, S., Brownridge, P.J., et al. (2019) Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation. The EMBO Journal, 38 (21). doi:10.15252/embj.2018100847.

Harper, J.W. and Elledge, S.J. (2007) The DNA Damage Response: Ten Years After. Molecular Cell, 28 (5): 739–745. doi:10.1016/j.molcel.2007.11.015.

Hartwell, L.H., Mortimer, R.K., Culotti, J., et al. (1973) GENETIC CONTROL OF THE CELL DIVISION CYCLE IN YEAST: V. GENETIC ANALYSIS OF cdc MUTANTS. Genetics, 74 (2): 267–286. doi:10.1093/genetics/74.2.267.

He, X., Riceberg, J., Soucy, T., et al. (2017) Probing the roles of SUMOylation in cancer cell biology by using a selective SAE inhibitor. Nature Chemical Biology, 13 (11): 1164–1171. doi:10.1038/nchembio.2463.

Hecker, C.-M., Rabiller, M., Haglund, K., et al. (2006) Specification of SUMO1- and SUMO2-interacting Motifs. Journal of Biological Chemistry, 281 (23): 16117–16127. doi:10.1074/jbc.M512757200.

Hemmi, H., Takeuchi, O., Sato, S., et al. (2004) The Roles of Two IκB Kinase-related Kinases in Lipopolysaccharide and Double Stranded RNA Signaling and Viral Infection. The Journal of Experimental Medicine, 199 (12): 1641–1650. doi:10.1084/jem.20040520.

Hendriks, I.A., Lyon, D., Su, D., et al. (2018) Site-specific characterization of endogenous SUMOylation across species and organs. Nature Communications, 9 (1): 2456. doi:10.1038/s41467-018-04957-4.

Hendriks, I.A., Lyon, D., Young, C., et al. (2017) Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. Nature Structural & Molecular Biology, 24 (3): 325–336. doi:10.1038/nsmb.3366.

Hietakangas, V., Anckar, J., Blomster, H.A., et al. (2006) PDSM, a motif for phosphorylation-dependent SUMO modification. Proceedings of the National Academy of Sciences, 103 (1): 45–50. doi:10.1073/pnas.0503698102.

Hirao, A., Kong, Y.-Y., Matsuoka, S., et al. (2000) DNA Damage-Induced Activation of p53 by the Checkpoint Kinase Chk2. Science, 287 (5459): 1824–1827. doi:10.1126/science.287.5459.1824.

Hohl, M., Kochańczyk, T., Tous, C., et al. (2015) Interdependence of the Rad50 Hook and Globular Domain Functions. Molecular Cell, 57 (3): 479–491. doi:10.1016/j.molcel.2014.12.018.

Hong, H. (2024) eLife assessment: SUMO's intrinsically disordered N-terminus is an intramolecular inhibitor of SUMO - SIM interactions. doi:10.7554/eLife.95313.1.sa2.

Hopfner, K.-P., Craig, L., Moncalian, G., et al. (2002) The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. Nature, 418 (6897): 562–566. doi:10.1038/nature00922.

Hornbeck, P. V., Zhang, B., Murray, B., et al. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Research, 43 (D1): D512–D520. doi:10.1093/nar/gku1267.

Hu, C.-K., Özlü, N., Coughlin, M., et al. (2012) Plk1 negatively regulates PRC1 to prevent premature midzone formation before cytokinesis. Molecular Biology of the Cell, 23 (14): 2702–2711. doi:10.1091/mbc.e12-01-0058.

Huang, B., Shang, Z., Li, B., et al. (2014) DNA-PKcs Associates With PLK1 and Is Involved in Proper Chromosome Segregation and Cytokinesis. Journal of Cellular Biochemistry, 115 (6): 1077–1088. doi:10.1002/jcb.24703.

Huang, H., Tsoi, S.C. -M., Sun, Y.H., et al. (1998) Identification and characterization of the SMT3 cDNA and gene encoding ubiquitin-like protein from Drosophila melanogaster. IUBMB Life, 46 (4): 775–785. doi:10.1080/15216549800204322.

Hudson, J.W., Kozarova, A., Cheung, P., et al. (2001) Late mitotic failure in mice lacking Sak, a polo-like kinase. Current Biology, 11 (6): 441–446. doi:10.1016/S0960-9822(01)00117-8.

Huen, M.S.Y., Grant, R., Manke, I., et al. (2007) RNF8 Transduces the DNA-Damage Signal via Histone Ubiquitylation and Checkpoint Protein Assembly. Cell, 131 (5): 901–914. doi:10.1016/j.cell.2007.09.041.

Huoh, Y.-S. and Ferguson, K.M. (2014) The Pellino E3 Ubiquitin Ligases Recognize Specific Phosphothreonine Motifs and Have Distinct Substrate Specificities. Biochemistry, 53 (30): 4946–4955. doi:10.1021/bi5005156.

Iliaki, S., Beyaert, R. and Afonina, I.S. (2021) Polo-like kinase 1 (PLK1) signaling in cancer and beyond. Biochemical Pharmacology, 193: 114747. doi:10.1016/j.bcp.2021.114747.

Iwasaki, A. and Medzhitov, R. (2015) Control of adaptive immunity by the innate immune system. Nature Immunology, 16 (4): 343–353. doi:10.1038/ni.3123.

de Jager, M., van Noort, J., van Gent, D.C., et al. (2001) Human Rad50/Mre11 Is a Flexible Complex that Can Tether DNA Ends. Molecular Cell, 8 (5): 1129–1135. doi:10.1016/S1097-2765(01)00381-1.

Janeway, C.A. (1989) Approaching the Asymptote? Evolution and Revolution in Immunology. Cold Spring Harbor Symposia on Quantitative Biology, 54 (0): 1–13. doi:10.1101/SQB.1989.054.01.003.

Jang, C.-Y., Coppinger, J.A., Seki, A., et al. (2009) Plk1 and Aurora A regulate the depolymerase activity and the cellular localization of Kif2a. Journal of Cell Science, 122 (9): 1334–1341. doi:10.1242/jcs.044321.

Janssen, A., van der Burg, M., Szuhai, K., et al. (2011) Chromosome Segregation Errors as a Cause of DNA Damage and Structural Chromosome Aberrations. Science, 333 (6051): 1895–1898. doi:10.1126/science.1210214.

Jensen, L.E. and Whitehead, A.S. (2003) Pellino2 activates the mitogen activated protein kinase pathway. FEBS Letters, 545 (2–3): 199–202. doi:10.1016/S0014-5793(03)00533-7.

Jentsch, S. and Pyrowolakis, G. (2000) Ubiquitin and its kin: how close are the family ties? Trends in Cell Biology, 10 (8): 335–342. doi:10.1016/S0962-8924(00)01785-2.

Jerabek-Willemsen, M., André, T., Wanner, R., et al. (2014) MicroScale Thermophoresis: Interaction analysis and beyond. Journal of Molecular Structure, 1077: 101–113. doi:10.1016/j.molstruc.2014.03.009.

Jiang, Z., Johnson, H.J., Nie, H., et al. (2003) Pellino 1 Is Required for Interleukin-1 (IL-1)-mediated Signaling through Its Interaction with the IL-1 Receptor-associated Kinase 4 (IRAK4)-IRAK-Tumor Necrosis Factor Receptor-associated Factor 6 (TRAF6) Complex. Journal of Biological Chemistry, 278 (13): 10952–10956. doi:10.1074/jbc.M212112200.

Jiang, Z., Ninomiya-Tsuji, J., Qian, Y., et al. (2002) Interleukin-1 (IL-1) Receptor-Associated Kinase-Dependent IL-1-Induced Signaling Complexes Phosphorylate TAK1 and TAB2 at the Plasma Membrane and Activate TAK1 in the Cytosol. Molecular and Cellular Biology, 22 (20): 7158–7167. doi:10.1128/MCB.22.20.7158-7167.2002.

Johnson, E.S. (1997) The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. The EMBO Journal, 16 (18): 5509–5519. doi:10.1093/emboj/16.18.5509.

Johnson, E.S. and Blobel, G. (1997) Ubc9p Is the Conjugating Enzyme for the Ubiquitin-like Protein Smt3p. Journal of Biological Chemistry, 272 (43): 26799–26802. doi:10.1074/jbc.272.43.26799.

Johnson, J.L., Yaron, T.M., Huntsman, E.M., et al. (2023) An atlas of substrate specificities for the human serine/threonine kinome. Nature, 613 (7945): 759–766. doi:10.1038/s41586-022-05575-3.

Jumper, J., Evans, R., Pritzel, A., et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature, 596 (7873): 583–589. doi:10.1038/s41586-021-03819-2.

Jungmichel, S., Clapperton, J.A., Lloyd, J., et al. (2012) The molecular basis of ATM-dependent dimerization of the Mdc1 DNA damage checkpoint mediator. Nucleic Acids Research, 40 (9): 3913–3928. doi:10.1093/nar/gkr1300.

Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. Acta Crystallographica Section D Biological Crystallography, 66 (2): 133–144. doi:10.1107/S0907444909047374.

Kagami, Y., Ono, M. and Yoshida, K. (2017) Plk1 phosphorylation of CAP-H2 triggers chromosome condensation by condensin II at the early phase of mitosis. Scientific Reports, 7 (1): 5583. doi:10.1038/s41598-017-05986-7.

Kaiser, W.J. and Offermann, M.K. (2005) Apoptosis Induced by the Toll-Like Receptor Adaptor TRIF Is Dependent on Its Receptor Interacting Protein Homotypic Interaction Motif. The Journal of Immunology, 174 (8): 4942–4952. doi:10.4049/jimmunol.174.8.4942.

Kalous, J. and Aleshkina, D. (2023) Multiple Roles of PLK1 in Mitosis and Meiosis. Cells, 12 (1): 187. doi:10.3390/cells12010187.

Kamitani, T., Nguyen, H.P., Kito, K., et al. (1998) Covalent Modification of PML by the Sentrin Family of Ubiquitin-like Proteins. Journal of Biological Chemistry, 273 (6): 3117–3120. doi:10.1074/jbc.273.6.3117.

Kanayama, A., Seth, R.B., Sun, L., et al. (2004) TAB2 and TAB3 Activate the NF-κB Pathway through Binding to Polyubiquitin Chains. Molecular Cell, 15 (4): 535–548. doi:10.1016/j.molcel.2004.08.008.

Karn, T., Holtrich, U., Wolf, G., et al. (1997) Human SAK related to the PLK/polo family of cell cycle kinases shows high mRNA expression in testis. Oncology Reports. doi:10.3892/or.4.3.505.

Katoh, M. and Katoh, M. (2004) Identification and characterization of human FOXK1 gene in silico. International Journal of Molecular Medicine. doi:10.3892/ijmm.14.1.127.

Keedy, D.A., Williams, C.J., Headd, J.J., et al. (2009) The other 90% of the protein: Assessment beyond the Cαs for CASP8 template-based and high-accuracy models. Proteins: Structure, Function, and Bioinformatics, 77 (S9): 29–49. doi:10.1002/prot.22551.

van Kempen, M., Kim, S.S., Tumescheit, C., et al. (2024) Fast and accurate protein structure search with Foldseek. Nature Biotechnology, 42 (2): 243–246. doi:10.1038/s41587-023-01773-0.

Khoury, G.A., Baliban, R.C. and Floudas, C.A. (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Scientific Reports, 1 (1): 90. doi:10.1038/srep00090.

Kibbe, W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Research, 35 (Web Server): W43–W46. doi:10.1093/nar/gkm234.

Kilkenny, M.L., Doré, A.S., Roe, S.M., et al. (2008) Structural and functional analysis of the Crb2–BRCT 2 domain reveals distinct roles in checkpoint signaling and DNA damage repair. Genes & Development, 22 (15): 2034–2047. doi:10.1101/gad.472808.

Kim, J.H., Sung, K.S., Jung, S.M., et al. (2011) Pellino-1, an Adaptor Protein of Interleukin-1 Receptor/Toll-like Receptor Signaling, Is Sumoylated by Ubc9. Molecules and Cells, 31 (1): 85–90. doi:10.1007/s10059-011-0006-x.

Kim, T.-S., Park, J.-E., Shukla, A., et al. (2013) Hierarchical recruitment of Plk4 and regulation of centriole biogenesis by two centrosomal scaffolds, Cep192 and Cep152. Proceedings of the National Academy of Sciences, 110 (50). doi:10.1073/pnas.1319656110.

Kinoshita, E., Kinoshita-Kikuta, E., Takiyama, K., et al. (2006) Phosphate-binding Tag, a New Tool to Visualize Phosphorylated Proteins. Molecular & Cellular Proteomics, 5 (4): 749–757. doi:10.1074/mcp.T500024-MCP200.

Kinsella, S., Fichtner, M., Watters, O., et al. (2018) Increased A20-E3 ubiquitin ligase interactions in bid-deficient glia attenuate TLR3- and TLR4-induced inflammation. Journal of Neuroinflammation, 15 (1): 130. doi:10.1186/s12974-018-1143-3.

Kissling, V.M., Reginato, G., Bianco, E., et al. (2022) Mre11-Rad50 oligomerization promotes DNA double-strand break repair. Nature Communications, 13 (1): 2374. doi:10.1038/s41467-022-29841-0.

Knipscheer, P., van Dijk, W.J., Olsen, J. V, et al. (2007) Noncovalent interaction between Ubc9 and SUMO promotes SUMO chain formation. The EMBO Journal, 26 (11): 2797–2807. doi:10.1038/sj.emboj.7601711.

Kobayashi, S., Kasaishi, Y., Nakada, S., et al. (2015) Rad18 and Rnf8 facilitate homologous recombination by two distinct mechanisms, promoting Rad51 focus formation and suppressing the toxic effect of nonhomologous end joining. Oncogene, 34 (33): 4403–4411. doi:10.1038/onc.2014.371.

Kochańczyk, T., Hann, Z.S., Lux, M.C., et al. (2024) Structural basis for transthiolation intermediates in the ubiquitin pathway. Nature, 633 (8028): 216–223. doi:10.1038/s41586-024-07828-9.

Köhler, M., Speck, C., Christiansen, M., et al. (1999) Evidence for Distinct Substrate Specificities of Importin α Family Members in Nuclear Protein Import. Molecular and Cellular Biology, 19 (11): 7782–7791. doi:10.1128/MCB.19.11.7782.

Kolas, N.K., Chapman, J.R., Nakada, S., et al. (2007a) Orchestration of the DNA-Damage Response by the RNF8 Ubiquitin Ligase. Science, 318 (5856): 1637–1640. doi:10.1126/science.1150034.

Kolas, N.K., Chapman, J.R., Nakada, S., et al. (2007b) Orchestration of the DNA-Damage Response by the RNF8 Ubiquitin Ligase. Science, 318 (5856): 1637–1640. doi:10.1126/science.1150034.

Kollewe, C., Mackensen, A.-C., Neumann, D., et al. (2004) Sequential Autophosphorylation Steps in the Interleukin-1 Receptor-associated Kinase-1 Regulate its Availability as an Adapter in Interleukin-1 Signaling. Journal of Biological Chemistry, 279 (7): 5227–5236. doi:10.1074/jbc.M309251200.

Kolli, N., Mikolajczyk, J., Drag, M., et al. (2010) Distribution and paralogue specificity of mammalian deSUMOylating enzymes. Biochemical Journal, 430 (2): 335–344. doi:10.1042/BJ20100504.

Kothe, M., Kohls, D., Low, S., et al. (2007) Structure of the Catalytic Domain of Human Polo-like Kinase 1 ,. Biochemistry, 46 (20): 5960–5971. doi:10.1021/bi602474j.

Kraulis, P.J. (1991) Similarity of Protein G and Ubiquitin. Science, 254 (5031): 581–582. doi:10.1126/science.1658931.

Krebs, E.G., Kent, A.B. and Fischer, E.H. (1958) THE MUSCLE PHOSPHORYLASE b KINASE REACTION. Journal of Biological Chemistry, 231 (1): 73–83. doi:10.1016/S0021-9258(19)77286-X.

Kumagai, A. and Dunphy, W.G. (1996) Purification and Molecular Cloning of Plx1, a Cdc25-Regulatory Kinase from Xenopus Egg Extracts. Science, 273 (5280): 1377–1380. doi:10.1126/science.273.5280.1377.

Kurepa, J., Walker, J.M., Smalle, J., et al. (2003) The Small Ubiquitin-like Modifier (SUMO) Protein Modification System in Arabidopsis. Journal of Biological Chemistry, 278 (9): 6862–6872. doi:10.1074/jbc.M209694200.

Lake, R.J. and Jelinek, W.R. (1993) Cell cycle- and terminal differentiation-associated regulation of the mouse mRNA encoding a conserved mitotic protein kinase. Molecular and Cellular Biology, 13 (12): 7793–7801. doi:10.1128/MCB.13.12.7793.

Lammens, K., Bemeleit, D.J., Möckel, C., et al. (2011) The Mre11:Rad50 Structure Shows an ATP-Dependent Molecular Clamp in DNA Double-Strand Break Repair. Cell, 145 (1): 54–66. doi:10.1016/j.cell.2011.02.038.

Landmann, C., Pierre-Elies, P., Goutte-Gattat, D., et al. (2020) The Mre11-Rad50-Nbs1 complex mediates the robust recruitment of Polo to DNA lesions during mitosis in Drosophila. Journal of Cell Science, 133 (13). doi:10.1242/jcs.244442.

Lapenta, V., Chiurazzi, P., van der Spek, P., et al. (1997) SMT3A,a Human Homologue of theS. cerevisiae SMT3Gene, Maps to Chromosome 21qter and Defines a Novel Gene Family. Genomics, 40 (2): 362–366. doi:10.1006/geno.1996.4556.

Laplantine, E., Fontan, E., Chiaravalli, J., et al. (2009) NEMO specifically recognizes K63-linked poly-ubiquitin chains through a new bipartite ubiquitin-binding domain. The EMBO Journal, 28 (19): 2885–2895. doi:10.1038/emboj.2009.241.

Laskowski, R.A., Jabłońska, J., Pravda, L., et al. (2018) PDBsum: Structural summaries of PDB entries. Protein Science, 27 (1): 129–134. doi:10.1002/pro.3289.

Lee, K.S., Grenfell, T.Z., Yarm, F.R., et al. (1998a) Mutation of the polo-box disrupts localization and mitotic functions of the mammalian polo kinase Plk. Proceedings of the National Academy of Sciences, 95 (16): 9301–9306. doi:10.1073/pnas.95.16.9301.

Lee, K.S., Grenfell, T.Z., Yarm, F.R., et al. (1998b) Mutation of the polo-box disrupts localization and mitotic functions of the mammalian polo kinase Plk. Proceedings of the National Academy of Sciences, 95 (16): 9301–9306. doi:10.1073/pnas.95.16.9301.

Lee, Y.S., Kim, J.H., Kim, S.-T., et al. (2010) Smad7 and Smad6 bind to discrete regions of Pellino-1 via their MH2 domains to mediate TGF-β1-induced negative regulation of IL-1R/TLR signaling. Biochemical and Biophysical Research Communications, 393 (4): 836–843. doi:10.1016/j.bbrc.2010.02.094.

Leimbacher, P.-A., Jones, S.E., Shorrocks, A.-M.K., et al. (2019) MDC1 Interacts with TOPBP1 to Maintain Chromosomal Stability during Mitosis. Molecular Cell, 74 (3): 571-583.e8. doi:10.1016/j.molcel.2019.02.014.

Lera, R.F., Norman, R.X., Dumont, M., et al. (2019) Plk1 protects kinetochore–centromere architecture against microtubule pulling forces. EMBO reports, 20 (10). doi:10.15252/embr.201948711.

Leung, G.C., Hudson, J.W., Kozarova, A., et al. (2002) The Sak polo-box comprises a structural domain sufficient for mitotic subcellular localization. Nature Structural Biology, 9 (10): 719–724. doi:10.1038/nsb848.

Li, B., Ouyang, B., Pan, H., et al. (1996) prk, a Cytokine-inducible Human Protein Serine/Threonine Kinase Whose Expression Appears to be Down-regulated in Lung Carcinomas. Journal of Biological Chemistry, 271 (32): 19402–19408. doi:10.1074/jbc.271.32.19402.

Li, D. and Wu, M. (2021) Pattern recognition receptors in health and diseases. Signal Transduction and Targeted Therapy, 6 (1): 291. doi:10.1038/s41392-021-00687-0.

Li, F., Li, Y., Liang, H., et al. (2018) HECTD3 mediates TRAF3 polyubiquitination and type I interferon induction during bacterial infection. Journal of Clinical Investigation, 128 (9): 4148–4162. doi:10.1172/JCI120406.

Li, J., Williams, B.L., Haire, L.F., et al. (2002) Structural and Functional Versatility of the FHA Domain in DNA-Damage Signaling by the Tumor Suppressor Kinase Chk2. Molecular Cell, 9 (5): 1045–1054. doi:10.1016/S1097-2765(02)00527-0.

Li, P., Allen, H., Banerjee, S., et al. (1995) Mice deficient in IL-1β-converting enzyme are defective in production of mature IL-1β and resistant to endotoxic shock. Cell, 80 (3): 401–411. doi:10.1016/0092-8674(95)90490-5.

Li, W., Wang, H.-Y., Zhao, X., et al. (2019a) A methylation-phosphorylation switch determines Plk1 kinase activity and function in DNA damage repair. Science Advances, 5 (3). doi:10.1126/sciadv.aau7566.

Li, Y.-J., Du, L., Wang, J., et al. (2019b) Allosteric Inhibition of Ubiquitin-like Modifications by a Class of Inhibitor of SUMO-Activating Enzyme. Cell Chemical Biology, 26 (2): 278-288.e6. doi:10.1016/j.chembiol.2018.10.026.

Li, Z., Li, J., Kong, Y., et al. (2017) Plk1 Phosphorylation of Mre11 Antagonizes the DNA Damage Response. Cancer Research, 77 (12): 3169–3180. doi:10.1158/0008-5472.CAN-16-2787.

Liang, Y.-C., Lee, C.-C., Yao, Y.-L., et al. (2016) SUMO5, a Novel Poly-SUMO Isoform, Regulates PML Nuclear Bodies. Scientific Reports, 6 (1): 26509. doi:10.1038/srep26509.

Liby, K., Wu, H., Ouyang, B., et al. (2001) Identification of the Human Homologue of the Early-Growth Response Gene Snk, Encoding a Serum-Inducible Kinase. DNA Sequence, 11 (6): 527–533. doi:10.3109/10425170109041337.

Lienhard, G.E. (2008) Non-functional phosphorylations? Trends in Biochemical Sciences, 33 (8): 351–352. doi:10.1016/j.tibs.2008.05.004.

Lin, C.-C., Huoh, Y.-S., Schmitz, K.R., et al. (2008) Pellino Proteins Contain a Cryptic FHA Domain that Mediates Interaction with Phosphorylated IRAK1. Structure, 16 (12): 1806–1816. doi:10.1016/j.str.2008.09.011.

Linder, M.I., Köhler, M., Boersema, P., et al. (2017) Mitotic Disassembly of Nuclear Pore Complexes Involves CDK1- and PLK1-Mediated Phosphorylation of Key Interconnecting Nucleoporins. Developmental Cell, 43 (2): 141-156.e7. doi:10.1016/j.devcel.2017.08.020.

LITCHFIELD, D.W. (2003) Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. Biochemical Journal, 369 (1): 1–15. doi:10.1042/bj20021469.

Liu, B., Lois, L.M. and Reverter, D. (2017a) Structural analysis and evolution of specificity of the SUMO UFD E1-E2 interactions. Scientific Reports, 7 (1): 41998. doi:10.1038/srep41998.

Liu, J., Luo, S., Zhao, H., et al. (2012) Structural mechanism of the phosphorylation-dependent dimerization of the MDC1 forkhead-associated domain. Nucleic Acids Research, 40 (9): 3898–3912. doi:10.1093/nar/gkr1296.

Liu, J., Qian, C. and Cao, X. (2016) Post-Translational Modification Control of Innate Immunity. Immunity, 45 (1): 15–30. doi:10.1016/j.immuni.2016.06.020.

Liu, Z., Sun, Q. and Wang, X. (2017b) PLK1, A Potential Target for Cancer Therapy. Translational Oncology, 10 (1): 22–32. doi:10.1016/j.tranon.2016.10.003.

Lloyd, J., Chapman, J.R., Clapperton, J.A., et al. (2009) A Supramodular FHA/BRCT-Repeat Architecture Mediates Nbs1 Adaptor Function in Response to DNA Damage. Cell, 139 (1): 100–111. doi:10.1016/j.cell.2009.07.043.

Lois, L.M. and Lima, C.D. (2005) Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. The EMBO Journal, 24 (3): 439–451. doi:10.1038/sj.emboj.7600552.

Lopez-Pelaez, M., Lamont, D.J., Peggie, M., et al. (2014) Protein kinase IKKβ-catalyzed phosphorylation of IRF5 at Ser462 induces its dimerization and nuclear translocation in myeloid cells. Proceedings of the National Academy of Sciences, 111 (49): 17432–17437. doi:10.1073/pnas.1418399111.

Lowery, D.M., Clauser, K.R., Hjerrild, M., et al. (2007) Proteomic screen defines the Polo-box domain interactome and identifies Rock2 as a Plk1 substrate. The EMBO Journal, 26 (9): 2262–2273. doi:10.1038/sj.emboj.7601683.

Lu, X., Olsen, S.K., Capili, A.D., et al. (2010) Designed Semisynthetic Protein Inhibitors of Ub/Ubl E1 Activating Enzymes. Journal of the American Chemical Society, 132 (6): 1748–1749. doi:10.1021/ja9088549.

Luo, K., Zhang, H., Wang, L., et al. (2012) Sumoylation of MDC1 is important for proper DNA damage response. The EMBO Journal, 31 (13): 3008–3019. doi:10.1038/emboj.2012.158.

Lussier-Price, M., Mascle, X.H., Cappadocia, L., et al. (2020) Characterization of a C-Terminal SUMO-Interacting Motif Present in Select PIAS-Family Proteins. Structure, 28 (5): 573-585.e5. doi:10.1016/j.str.2020.04.002.

Lussier-Price, M., Wahba, H.M., Mascle, X.H., et al. (2022) Zinc controls PML nuclear body formation through regulation of a paralog specific auto-inhibition in SUMO1. Nucleic Acids Research, 50 (14): 8331–8348. doi:10.1093/nar/gkac620.

Ma, S., Charron, J. and Erikson, R.L. (2003) Role of Plk2 (Snk) in Mouse Development and Cell Proliferation. Molecular and Cellular Biology, 23 (19): 6936–6943. doi:10.1128/MCB.23.19.6936-6943.2003.

Macůrek, L., Lindqvist, A., Lim, D., et al. (2008) Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery. Nature, 455 (7209): 119–123. doi:10.1038/nature07185.

Mahajan, R., Delphin, C., Guan, T., et al. (1997) A Small Ubiquitin-Related Polypeptide Involved in Targeting RanGAP1 to Nuclear Pore Complex Protein RanBP2. Cell, 88 (1): 97–107. doi:10.1016/S0092-8674(00)81862-0.

Mannen, H., Tseng, H.-M., Cho, C., et al. (1996) Cloning and Expression of Human Homolog HSMT3 to Yeast SMT3 Suppressor of MIF2 Mutations in a Centromere Protein Gene. Biochemical and Biophysical Research Communications, 222 (1): 178–180. doi:10.1006/bbrc.1996.0717.

Manning, G., Whyte, D.B., Martinez, R., et al. (2002) The Protein Kinase Complement of the Human Genome. Science, 298 (5600): 1912–1934. doi:10.1126/science.1075762.

Mao, Y., Desai, S.D. and Liu, L.F. (2000) SUMO-1 Conjugation to Human DNA Topoisomerase II Isozymes. Journal of Biological Chemistry, 275 (34): 26066–26073. doi:10.1074/jbc.M001831200.

Mardin, B.R., Agircan, F.G., Lange, C., et al. (2011) Plk1 Controls the Nek2A-PP1γ Antagonism in Centrosome Disjunction. Current Biology, 21 (13): 1145–1151. doi:10.1016/j.cub.2011.05.047.

Mathur, S., Fletcher, A.J., Branigan, E., et al. (2020) Photocrosslinking Activity-Based Probes for Ubiquitin RING E3 Ligases. Cell Chemical Biology, 27 (1): 74-82.e6. doi:10.1016/j.chembiol.2019.11.013.

Matic, I., Schimmel, J., Hendriks, I.A., et al. (2010) Site-Specific Identification of SUMO-2 Targets in Cells Reveals an Inverted SUMOylation Motif and a Hydrophobic Cluster SUMOylation Motif. Molecular Cell, 39 (4): 641–652. doi:10.1016/j.molcel.2010.07.026.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., et al. (2007) ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage. Science, 316 (5828): 1160–1166. doi:10.1126/science.1140321.

Matsuoka, S., Rotman, G., Ogawa, A., et al. (2000) Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. Proceedings of the National Academy of Sciences, 97 (19): 10389–10394. doi:10.1073/pnas.190030497.

Matunis, M.J., Coutavas, E. and Blobel, G. (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. The Journal of cell biology, 135 (6): 1457–1470. doi:10.1083/jcb.135.6.1457.

Maya, R., Balass, M., Kim, S.-T., et al. (2001) ATM-dependent phosphorylation of Mdm2 on serine 395: role in p53 activation by DNA damage. Genes & Development, 15 (9): 1067–1077. doi:10.1101/gad.886901.

McCoy, A.J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallographica Section D Biological Crystallography, 63 (1): 32–41. doi:10.1107/S0907444906045975.

McDonald, M., Trost, B. and Napper, S. (2018) Conservation of kinase-phosphorylation site pairings: Evidence for an evolutionarily dynamic phosphoproteome. PLOS ONE, 13 (8): e0202036. doi:10.1371/journal.pone.0202036.

MEDZHITOV, R. and JANEWAY, C.A. (1999) Innate Immune Induction of the Adaptive Immune Response. Cold Spring Harbor Symposia on Quantitative Biology, 64 (0): 429–436. doi:10.1101/sqb.1999.64.429.

Meluh, P.B. and Koshland, D. (1995) Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C. Molecular Biology of the Cell, 6 (7): 793–807. doi:10.1091/mbc.6.7.793.

Meng, E.C., Goddard, T.D., Pettersen, E.F., et al. (2023) UCSF ChimeraX: Tools for structure building and analysis. Protein Science, 32 (11). doi:10.1002/pro.4792.

Meulmeester, E., Kunze, M., Hsiao, H.H., et al. (2008) Mechanism and Consequences for Paralog-Specific Sumoylation of Ubiquitin-Specific Protease 25. Molecular Cell, 30 (5): 610–619. doi:10.1016/j.molcel.2008.03.021.

Michel, B. (1997) DNA double-strand breaks caused by replication arrest. The EMBO Journal, 16 (2): 430–438. doi:10.1093/emboj/16.2.430.

Mikolajczyk, J., Drag, M., Békés, M., et al. (2007) Small Ubiquitin-related Modifier (SUMO)-specific Proteases. Journal of Biological Chemistry, 282 (36): 26217–26224. doi:10.1074/jbc.M702444200.

Mirdita, M., Schütze, K., Moriwaki, Y., et al. (2022) ColabFold: making protein folding accessible to all. Nature Methods, 19 (6): 679–682. doi:10.1038/s41592-022-01488-1.

Moen, J.M., Mohler, K., Rogulina, S., et al. (2022) Enhanced access to the human phosphoproteome with genetically encoded phosphothreonine. Nature Communications, 13 (1): 7226. doi:10.1038/s41467-022-34980-5.

Mohler, K., Moen, J.M., Rogulina, S., et al. (2023) System-wide optimization of an orthogonal translation system with enhanced biological tolerance. Molecular Systems Biology, 19 (8). doi:10.15252/msb.202110591.

Morgan, J.A.M., Singh, A., Kurz, L., et al. (2024) Extensive protein pyrophosphorylation revealed in human cell lines. Nature Chemical Biology, 20 (10): 1305–1316. doi:10.1038/s41589-024-01613-5.

Moriguchi, T., Kuroyanagi, N., Yamaguchi, K., et al. (1996) A Novel Kinase Cascade Mediated by Mitogen-activated Protein Kinase Kinase 6 and MKK3. Journal of Biological Chemistry, 271 (23): 13675–13679. doi:10.1074/jbc.271.23.13675.

Morris, J.R., Boutell, C., Keppler, M., et al. (2009) The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. Nature, 462 (7275): 886–890. doi:10.1038/nature08593.

Moutty, M.C., Sakin, V. and Melchior, F. (2011) Importin α/β mediates nuclear import of individual SUMO E1 subunits and of the holo-enzyme. Molecular Biology of the Cell, 22 (5): 652–660. doi:10.1091/mbc.e10-05-0461.

Moynagh, P.N. (2014) The roles of Pellino E3 ubiquitin ligases in immunity. Nature Reviews Immunology, 14 (2): 122–131. doi:10.1038/nri3599.

Myer, D.L., Robbins, S.B., Yin, M., et al. (2011) Absence of polo-like kinase 3 in mice stabilizes Cdc25A after DNA damage but is not sufficient to produce tumors. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 714 (1–2): 1–10. doi:10.1016/j.mrfmmm.2011.02.006.

Nakajima, H., Toyoshima-Morimoto, F., Taniguchi, E., et al. (2003) Identification of a Consensus Motif for Plk (Polo-like Kinase) Phosphorylation Reveals Myt1 as a Plk1 Substrate*. Journal of Biological Chemistry, 278 (28): 25277–25280. doi:10.1074/jbc.C300126200.

Nguyen, T., De Nardo, D., Masendycz, P., et al. (2009) Regulation of IRAK-1 activation by its C-terminal domain. Cellular Signalling, 21 (5): 719–726. doi:10.1016/j.cellsig.2009.01.010.

Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., et al. (1989) A General Method for Site-specific Incorporation of Unnatural Amino Acids into Proteins. Science, 244 (4901): 182–188. doi:10.1126/science.2649980.

Ochoa, D., Jarnuczak, A.F., Viéitez, C., et al. (2020) The functional landscape of the human phosphoproteome. Nature Biotechnology, 38 (3): 365–373. doi:10.1038/s41587-019-0344-3.

Ogawa, T., Yu, X., Shinohara, A., et al. (1993) Similarity of the Yeast RAD51 Filament to the Bacterial RecA Filament. Science, 259 (5103): 1896–1899. doi:10.1126/science.8456314.

Ohkura, H., Hagan, I.M. and Glover, D.M. (1995) The conserved Schizosaccharomyces pombe kinase plo1, required to form a bipolar spindle, the actin ring, and septum, can drive septum formation in G1 and G2 cells. Genes & Development, 9 (9): 1059–1073. doi:10.1101/gad.9.9.1059.

Okamura, E., Sakamoto, T., Sasaki, T., et al. (2017) A Plant Ancestral Polo-Like Kinase Sheds Light on the Mystery of the Evolutionary Disappearance of Polo-Like Kinases in the Plant Kingdom. CYTOLOGIA, 82 (3): 261–266. doi:10.1508/cytologia.82.261.

Okuda, M., Horn, H.F., Tarapore, P., et al. (2000) Nucleophosmin/B23 Is a Target of CDK2/Cyclin E in Centrosome Duplication. Cell, 103 (1): 127–140. doi:10.1016/S0092-8674(00)00093-3.

Olivieri, M., Cho, T., Álvarez-Quilón, A., et al. (2020) A Genetic Map of the Response to DNA Damage in Human Cells. Cell, 182 (2): 481-496.e21. doi:10.1016/j.cell.2020.05.040.

Olsen, J. V., Vermeulen, M., Santamaria, A., et al. (2010a) Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. Science Signaling, 3 (104). doi:10.1126/scisignal.2000475.

Olsen, S.K., Capili, A.D., Lu, X., et al. (2010b) Active site remodelling accompanies thioester bond formation in the SUMO E1. Nature, 463 (7283): 906–912. doi:10.1038/nature08765.

Ordureau, A., Smith, H., Windheim, M., et al. (2008) The IRAK-catalysed activation of the E3 ligase function of Pellino isoforms induces the Lys63-linked polyubiquitination of IRAK1. Biochemical Journal, 409 (1): 43–52. doi:10.1042/BJ20071365.

Orthwein, A., Fradet-Turcotte, A., Noordermeer, S.M., et al. (2014) Mitosis Inhibits DNA Double-Strand Break Repair to Guard Against Telomere Fusions. Science, 344 (6180): 189–193. doi:10.1126/science.1248024.

Otomo, C., Metlagel, Z., Takaesu, G., et al. (2013) Structure of the human ATG12~ATG5 conjugate required for LC3 lipidation in autophagy. Nature Structural & Molecular Biology, 20 (1): 59–66. doi:10.1038/nsmb.2431.

Overington, J.P. (1992) Comparison of three-dimensional structures of homologous proteins. Current Opinion in Structural Biology, 2 (3): 394–401. doi:10.1016/0959-440X(92)90231-U.

Owerbach, D., McKay, E.M., Yeh, E.T.H., et al. (2005) A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. Biochemical and Biophysical Research Communications, 337 (2): 517–520. doi:10.1016/j.bbrc.2005.09.090.

Park, H.-S., Hohn, M.J., Umehara, T., et al. (2011a) Expanding the Genetic Code of Escherichia coli with Phosphoserine. Science, 333 (6046): 1151–1154. doi:10.1126/science.1207203.

Park, Y.B., Chae, J., Kim, Y.C., et al. (2011b) Crystal Structure of Human Mre11: Understanding Tumorigenic Mutations. Structure, 19 (11): 1591–1602. doi:10.1016/j.str.2011.09.010.

Park, Y.B., Hohl, M., Padjasek, M., et al. (2017) Eukaryotic Rad50 functions as a rod-shaped dimer. Nature Structural & Molecular Biology, 24 (3): 248–257. doi:10.1038/nsmb.3369.

Paull, T.T. and Gellert, M. (1998) The 3' to 5' Exonuclease Activity of Mre11 Facilitates Repair of DNA Double-Strand Breaks. Molecular Cell, 1 (7): 969–979. doi:10.1016/S1097-2765(00)80097-0.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., et al. (2023) InterPro in 2022. Nucleic Acids Research, 51 (D1): D418–D427. doi:10.1093/nar/gkac993.

Peng, B., Shi, R., Bian, J., et al. (2021) PARP1 and CHK1 coordinate PLK1 enzymatic activity during the DNA damage response to promote homologous recombination-mediated repair. Nucleic Acids Research, 49 (13): 7554–7570. doi:10.1093/nar/gkab584.

Penkert, M., Yates, L.M., Schümann, M., et al. (2017) Unambiguous Identification of Serine and Threonine Pyrophosphorylation Using Neutral-Loss-Triggered Electron-Transfer/Higher-Energy Collision Dissociation. Analytical Chemistry, 89 (6): 3672–3680. doi:10.1021/acs.analchem.6b05095.

Pennell, S., Westcott, S., Ortiz-Lombardía, M., et al. (2010) Structural and Functional Analysis of Phosphothreonine-Dependent FHA Domain Interactions. Structure, 18 (12): 1587–1595. doi:10.1016/j.str.2010.09.014.

Phong, M.S., Van Horn, R.D., Li, S., et al. (2010) p38 Mitogen-Activated Protein Kinase Promotes Cell Survival in Response to DNA Damage but Is Not Required for the G 2 DNA Damage Checkpoint in Human Cancer Cells. Molecular and Cellular Biology, 30 (15): 3816–3826. doi:10.1128/MCB.00949-09.

Pickart, C.M., Kasperek, E.M., Beal, R., et al. (1994) Substrate properties of site-specific mutant ubiquitin protein (G76A) reveal unexpected mechanistic features of ubiquitin-activating enzyme (E1). The Journal of biological chemistry, 269 (10): 7115–23.

Polycarpo, C., Ambrogelly, A., Bérubé, A., et al. (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. Proceedings of the National Academy of Sciences, 101 (34): 12450–12454. doi:10.1073/pnas.0405362101.

Pruneda, J.N., Littlefield, P.J., Soss, S.E., et al. (2012) Structure of an E3:E2~Ub Complex Reveals an Allosteric Mechanism Shared among RING/U-box Ligases. Molecular Cell, 47 (6): 933–942. doi:10.1016/j.molcel.2012.07.001.

Pruneda, J.N., Stoll, K.E., Bolton, L.J., et al. (2011) Ubiquitin in Motion: Structural Studies of the Ubiquitin-Conjugating Enzyme~Ubiquitin Conjugate. Biochemistry, 50 (10): 1624–1633. doi:10.1021/bi101913m.

Qiao, R., Weissmann, F., Yamaguchi, M., et al. (2016) Mechanism of APC/C CDC20 activation by mitotic phosphorylation. Proceedings of the National Academy of Sciences, 113 (19). doi:10.1073/pnas.1604929113.

Ramazi, S. and Zahiri, J. (2021) Post-translational modifications in proteins: resources, tools and prediction methods. Database, 2021. doi:10.1093/database/baab012.

Ramkumar, P., Lee, C.M., Moradian, A., et al. (2015) JNK-associated Leucine Zipper Protein Functions as a Docking Platform for Polo-like Kinase 1 and Regulation of the Associating Transcription Factor Forkhead Box Protein K1. Journal of Biological Chemistry, 290 (49): 29617–29628. doi:10.1074/jbc.M115.664649.

Reinhardt, R. and Leonard, T.A. (2023) A critical evaluation of protein kinase regulation by activation loop autophosphorylation. eLife, 12. doi:10.7554/eLife.88210.

Resch, K., Jockusch, H. and Schmitt-John, T. (2001) Assignment of homologous genes, Peli1/PELI1 and Peli2/PELI2, for the Pelle adaptor protein Pellino to mouse chromosomes 11 and 14 and human chromosomes 2p13.3 and 14q21, respectively, by physical and radiation hybrid mapping. Cytogenetic and Genome Research, 92 (1–2): 172–174. doi:10.1159/000056895.

Reverter, D. and Lima, C.D. (2005) Insights into E3 ligase activity revealed by a SUMO–RanGAP1–Ubc9–Nup358 complex. Nature, 435 (7042): 687–692. doi:10.1038/nature03588.

Rodriguez, M.S., Dargemont, C. and Hay, R.T. (2001) SUMO-1 Conjugation in Vivo Requires Both a Consensus Modification Motif and Nuclear Targeting. Journal of Biological Chemistry, 276 (16): 12654–12659. doi:10.1074/jbc.M009476200.

Rogerson, D.T., Sachdeva, A., Wang, K., et al. (2015) Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nature Chemical Biology, 11 (7): 496–503. doi:10.1038/nchembio.1823.

Rotheneder, M., Stakyte, K., van de Logt, E., et al. (2023) Cryo-EM structure of the Mre11-Rad50-Nbs1 complex reveals the molecular mechanism of scaffolding functions. Molecular Cell, 83 (2): 167-185.e9. doi:10.1016/j.molcel.2022.12.003.

Roumestand, C., Leiba, J., Galophe, N., et al. (2011) Structural Insight into the Mycobacterium tuberculosis Rv0020c Protein and Its Interaction with the PknB Kinase. Structure, 19 (10): 1525–1534. doi:10.1016/j.str.2011.07.011.

Rubio Gomez, M.A. and Ibba, M. (2020) Aminoacyl-tRNA synthetases. RNA, 26 (8): 910–936. doi:10.1261/rna.071720.119.

Saitoh, H. and Hinchey, J. (2000) Functional Heterogeneity of Small Ubiquitin-related Protein Modifiers SUMO-1 versus SUMO-2/3. Journal of Biological Chemistry, 275 (9): 6252–6258. doi:10.1074/jbc.275.9.6252.

Saitoh, H., Sparrow, D.B., Shiomi, T., et al. (1998) Ubc9p and the conjugation of SUMO-1 to RanGAP1 and RanBP2. Current Biology, 8 (2): 121–124. doi:10.1016/S0960-9822(98)70044-2.

Sakin, V., Richter, S.M., Hsiao, H.-H., et al. (2015) Sumoylation of the GTPase Ran by the RanBP2 SUMO E3 Ligase Complex. Journal of Biological Chemistry, 290 (39): 23589–23602. doi:10.1074/jbc.M115.660118.

Salguero, I., Belotserkovskaya, R., Coates, J., et al. (2019) MDC1 PST-repeat region promotes histone H2AX-independent chromatin association and DNA damage tolerance. Nature Communications, 10 (1): 5191. doi:10.1038/s41467-019-12929-5.

Schauvliege, R., Janssens, S. and Beyaert, R. (2006) Pellino proteins are more than scaffold proteins in TLR/IL-1R signalling: A role as novel RING E3–ubiquitin-ligases. FEBS Letters, 580 (19): 4697–4702. doi:10.1016/j.febslet.2006.07.046.

Schiller, C.B., Lammens, K., Guerini, I., et al. (2012) Structure of Mre11–Nbs1 complex yields insights into ataxia-telangiectasia–like disease mutations and DNA damage signaling. Nature Structural & Molecular Biology, 19 (7): 693–700. doi:10.1038/nsmb.2323.

Schulz, S., Chachami, G., Kozaczkiewicz, L., et al. (2012) Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. EMBO reports, 13 (10): 930–938. doi:10.1038/embor.2012.125.

Schwarz, J.K., Lovly, C.M. and Piwnica-Worms, H. (2003) Regulation of the Chk2 protein kinase by oligomerization-mediated cis- and trans-phosphorylation. Molecular cancer research : MCR, 1 (8): 598–609.

Sekiyama, N., Ikegami, T., Yamane, T., et al. (2008) Structure of the Small Ubiquitin-like Modifier (SUMO)-interacting Motif of MBD1-containing Chromatin-associated Factor 1 Bound to SUMO-3. Journal of Biological Chemistry, 283 (51): 35966–35975. doi:10.1074/jbc.M802528200.

Senior, A.W., Evans, R., Jumper, J., et al. (2020) Improved protein structure prediction using potentials from deep learning. Nature, 577 (7792): 706–710. doi:10.1038/s41586-019-1923-7.

Seong, Y.-S., Kamijo, K., Lee, J.-S., et al. (2002) A Spindle Checkpoint Arrest and a Cytokinesis Failure by the Dominant-negative Polo-box Domain of Plk1 in U-2 OS Cells. Journal of Biological Chemistry, 277 (35): 32282–32293. doi:10.1074/jbc.M202602200.

Sharma, P., Mahen, R., Rossmann, M., et al. (2019) A cryptic hydrophobic pocket in the polo-box domain of the polo-like kinase PLK1 regulates substrate recognition and mitotic chromosome segregation. Scientific Reports, 9 (1): 15930. doi:10.1038/s41598-019-50702-2.

Shen, M., Cai, Y., Yang, Y., et al. (2013a) Centrosomal protein FOR20 is essential for S-phase progression by recruiting Plk1 to centrosomes. Cell Research, 23 (11): 1284–1295. doi:10.1038/cr.2013.127.

Shen, M., Cai, Y., Yang, Y., et al. (2013b) Centrosomal protein FOR20 is essential for S-phase progression by recruiting Plk1 to centrosomes. Cell Research, 23 (11): 1284–1295. doi:10.1038/cr.2013.127.

Shenderov, K., Riteau, N., Yip, R., et al. (2014) Cutting Edge: Endoplasmic Reticulum Stress Licenses Macrophages To Produce Mature IL-1β in Response to TLR4 Stimulation through a Caspase-8– and TRIF-Dependent Pathway. The Journal of Immunology, 192 (5): 2029–2033. doi:10.4049/jimmunol.1302549.

Shi, X., Du, Y., Li, S., et al. (2022) The Role of SUMO E3 Ligases in Signaling Pathway of Cancer Cells. International Journal of Molecular Sciences, 23 (7): 3639. doi:10.3390/ijms23073639.

Shieh, S.-Y., Ikeda, M., Taya, Y., et al. (1997) DNA Damage-Induced Phosphorylation of p53 Alleviates Inhibition by MDM2. Cell, 91 (3): 325–334. doi:10.1016/S0092-8674(00)80416-X.

Shimizu, S. and Dempsey, W.B. (1978) 3-hydroxypyruvate substitutes for pyridoxine in serC mutants of Escherichia coli K-12. Journal of Bacteriology, 134 (3): 944–949. doi:10.1128/jb.134.3.944-949.1978.

Siednienko, J., Jackson, R., Mellett, M., et al. (2012) Pellino3 targets the IRF7 pathway and facilitates autoregulation of TLR3- and viral-induced expression of type I interferons. Nature Immunology, 13 (11): 1055–1062. doi:10.1038/ni.2429.

Singh, P., Pesenti, M.E., Maffini, S., et al. (2021) BUB1 and CENP-U, Primed by CDK1, Are the Main PLK1 Kinetochore Receptors in Mitosis. Molecular Cell, 81 (1): 67-87.e9. doi:10.1016/j.molcel.2020.10.040.

Śledź, P., Stubbs, C.J., Lang, S., et al. (2011) From Crystal Packing to Molecular Recognition: Prediction and Discovery of a Binding Site on the Surface of Polo-Like Kinase 1. Angewandte Chemie International Edition, 50 (17): 4003–4006. doi:10.1002/anie.201008019.

Slevin, L.K., Nye, J., Pinkerton, D.C., et al. (2012) The Structure of the Plk4 Cryptic Polo Box Reveals Two Tandem Polo Boxes Required for Centriole Duplication. Structure, 20 (11): 1905–1917. doi:10.1016/j.str.2012.08.025.

Smith, H., Liu, X.-Y., Dai, L., et al. (2011) The role of TBK1 and IKKε in the expression and activation of Pellino 1. Biochemical Journal, 434 (3): 537–548. doi:10.1042/BJ20101421.

Smith, H., Peggie, M., Campbell, D.G., et al. (2009) Identification of the phosphorylation sites on the E3 ubiquitin ligase Pellino that are critical for activation by IRAK1 and IRAK4. Proceedings of the National Academy of Sciences, 106 (12): 4584–4590. doi:10.1073/pnas.0900774106.

Song, B., Liu, X.S., Davis, K., et al. (2011) Plk1 Phosphorylation of Orc2 Promotes DNA Replication under Conditions of Stress. Molecular and Cellular Biology, 31 (23): 4844–4856. doi:10.1128/MCB.06110-11.

Song, B., Liu, X.S. and Liu, X. (2012) Polo-like kinase 1 (Plk1): an Unexpected Player in DNA Replication. Cell Division, 7 (1): 3. doi:10.1186/1747-1028-7-3.

Song, J., Durrin, L.K., Wilkinson, T.A., et al. (2004) Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. Proceedings of the National Academy of Sciences, 101 (40): 14373–14378. doi:10.1073/pnas.0403498101.

Song, J., Zhang, Z., Hu, W., et al. (2005) Small Ubiquitin-like Modifier (SUMO) Recognition of a SUMO Binding Motif. Journal of Biological Chemistry, 280 (48): 40122–40129. doi:10.1074/jbc.M507059200.

Spycher, C., Miller, E.S., Townsend, K., et al. (2008) Constitutive phosphorylation of MDC1 physically links the MRE11–RAD50–NBS1 complex to damaged chromatin. The Journal of Cell Biology, 181 (2): 227–240. doi:10.1083/jcb.200709008.

Stehmeier, P. and Muller, S. (2009) Phospho-Regulated SUMO Interaction Modules Connect the SUMO System to CK2 Signaling. Molecular Cell, 33 (3): 400–409. doi:10.1016/j.molcel.2009.01.013.

Stephens, P.J., Greenman, C.D., Fu, B., et al. (2011) Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. Cell, 144 (1): 27–40. doi:10.1016/j.cell.2010.11.055.

Stewart, G.S., Wang, B., Bignell, C.R., et al. (2003) MDC1 is a mediator of the mammalian DNA damage checkpoint. Nature, 421 (6926): 961–966. doi:10.1038/nature01446.

Strickson, S., Emmerich, C.H., Goh, E.T.H., et al. (2017) Roles of the TRAF6 and Pellino E3 ligases in MyD88 and RANKL signaling. Proceedings of the National Academy of Sciences, 114 (17). doi:10.1073/pnas.1702367114.

Stucki, M., Clapperton, J.A., Mohammad, D., et al. (2005) MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks. Cell, 123 (7): 1213–1226. doi:10.1016/j.cell.2005.09.038.

Sumikawa, Y., Asada, H., Hoshino, K., et al. (2006) Induction of β-defensin 3 in keratinocytes stimulated by bacterial lipopeptides through toll-like receptor 2. Microbes and Infection, 8 (6): 1513–1521. doi:10.1016/j.micinf.2006.01.008.

Sunkel, C.E. and Glover, D.M. (1988) <i>Polo<i>, a mitotic mutant of Drosophila displaying abnormal spindle poles. Journal of Cell Science, 89 (1): 25–38. doi:10.1242/jcs.89.1.25.

SUZUKI, H., NAGAI, K., AKUTSU, E., et al. (1970) ON THE MECHANISM OF ACTION OF BLEOMYCIN. The Journal of Antibiotics, 23 (10): 473–480. doi:10.7164/antibiotics.23.473.

Tatham, M.H., Geoffroy, M.-C., Shen, L., et al. (2008) RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. Nature Cell Biology, 10 (5): 538–546. doi:10.1038/ncb1716.

Tatham, M.H., Jaffray, E., Vaughan, O.A., et al. (2001) Polymeric Chains of SUMO-2 and SUMO-3 Are Conjugated to Protein Substrates by SAE1/SAE2 and Ubc9. Journal of Biological Chemistry, 276 (38): 35368–35374. doi:10.1074/jbc.M104214200.

Taylor, S.S. and Kornev, A.P. (2011a) Protein kinases: evolution of dynamic regulatory proteins. Trends in Biochemical Sciences, 36 (2): 65–77. doi:10.1016/j.tibs.2010.09.006.

Taylor, S.S. and Kornev, A.P. (2011b) Protein kinases: evolution of dynamic regulatory proteins. Trends in Biochemical Sciences, 36 (2): 65–77. doi:10.1016/j.tibs.2010.09.006.

Thorslund, T., Ripplinger, A., Hoffmann, S., et al. (2015) Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. Nature, 527 (7578): 389–393. doi:10.1038/nature15401.

Tickle, I.J., Flensburg, C., Keller, P., et al. (n.d.) STARANISO. pp. 2018–2022.

Torres Esteban, M., Stewart, M. J., Bragginton, E., Meroni, A., Pellizzari, A., Jeanrenaud, A., Smerdon, S. J., & Stucki, M. (2025). MDC1 mediates Pellino recruitment to sites of DNA double-strand breaks. *Life Science Alliance*, *8*(5), e202403074. https://doi.org/10.26508/lsa.202403074

Tran, J.B., Padjasek, M. and Krężel, A. (2022) Relations between Structure and Zn(II) Binding Affinity Shed Light on the Mechanisms of Rad50 Hook Domain Functioning and Its Phosphorylation. International Journal of Molecular Sciences, 23 (19): 11140. doi:10.3390/ijms231911140.

Trujillo, K.M. and Sung, P. (2001) DNA Structure-specific Nuclease Activities in theSaccharomyces cerevisiae Rad50·Mre11 Complex. Journal of Biological Chemistry, 276 (38): 35458–35464. doi:10.1074/jbc.M105482200.

Truong, K., Lee, T.D., Li, B., et al. (2012) Sumoylation of SAE2 C Terminus Regulates SAE Nuclear Localization. Journal of Biological Chemistry, 287 (51): 42611–42619. doi:10.1074/jbc.M112.420877.

Tseng, P.-H., Matsuzawa, A., Zhang, W., et al. (2010) Different modes of ubiquitination of the adaptor TRAF3 selectively activate the expression of type I interferons and proinflammatory cytokines. Nature Immunology, 11 (1): 70–75. doi:10.1038/ni.1819.

Tunyasuvunakool, K., Adler, J., Wu, Z., et al. (2021) Highly accurate protein structure prediction for the human proteome. Nature, 596 (7873): 590–596. doi:10.1038/s41586-021-03828-1.

Usui, T., Ohta, T., Oshiumi, H., et al. (1998) Complex Formation and Functional Versatility of Mre11 of Budding Yeast in Recombination. Cell, 95 (5): 705–716. doi:10.1016/S0092-8674(00)81640-2.

Uziel, T. (2003) Requirement of the MRN complex for ATM activation by DNA damage. The EMBO Journal, 22 (20): 5612–5621. doi:10.1093/emboj/cdg541.

Uzunova, K., Göttsche, K., Miteva, M., et al. (2007) Ubiquitin-dependent Proteolytic Control of SUMO Conjugates. Journal of Biological Chemistry, 282 (47): 34167–34175. doi:10.1074/jbc.M706505200.

Vabulas, R.M., Ahmad-Nejad, P., da Costa, C., et al. (2001) Endocytosed HSP60s Use Toll-like Receptor 2 (TLR2) and TLR4 to Activate the Toll/Interleukin-1 Receptor Signaling Pathway in Innate Immune Cells. Journal of Biological Chemistry, 276 (33): 31332–31339. doi:10.1074/jbc.M103217200.

Vertegaal, A.C.O., Andersen, J.S., Ogg, S.C., et al. (2006) Distinct and Overlapping Sets of SUMO-1 and SUMO-2 Target Proteins Revealed by Quantitative Proteomics. Molecular & Cellular Proteomics, 5 (12): 2298–2310. doi:10.1074/mcp.M600212-MCP200.

Vigneron, S., Sundermann, L., Labbé, J.-C., et al. (2018) Cyclin A-cdk1-Dependent Phosphorylation of Bora Is the Triggering Factor Promoting Mitotic Entry. Developmental Cell, 45 (5): 637-650.e7. doi:10.1016/j.devcel.2018.05.005.

Vilenchik, M.M. and Knudson, A.G. (2003) Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. Proceedings of the National Academy of Sciences, 100 (22): 12871–12876. doi:10.1073/pnas.2135498100.

Vonrhein, C., Flensburg, C., Keller, P., et al. (2011) Data processing and analysis with the autoPROC toolbox. Acta Crystallographica Section D Biological Crystallography, 67 (4): 293–302. doi:10.1107/S0907444911007773.

van Vugt, M.A.T.M., Gardino, A.K., Linding, R., et al. (2010) A Mitotic Phosphorylation Feedback Network Connects Cdk1, Plk1, 53BP1, and Chk2 to Inactivate the G2/M DNA Damage Checkpoint. PLoS Biology, 8 (1): e1000287. doi:10.1371/journal.pbio.1000287.

Walser, F., Mulder, M.P.C., Bragantini, B., et al. (2020) Ubiquitin Phosphorylation at Thr12 Modulates the DNA Damage Response. Molecular Cell, 80 (3): 423-436.e9. doi:10.1016/j.molcel.2020.09.017.

Walsh, M.C., Kim, G.K., Maurizio, P.L., et al. (2008) TRAF6 Autoubiquitination-Independent Activation of the NFκB and MAPK Pathways in Response to IL-1 and RANKL. PLoS ONE, 3 (12): e4064. doi:10.1371/journal.pone.0004064.

Wang, C., Deng, L., Hong, M., et al. (2001) TAK1 is a ubiquitin-dependent kinase of MKK and IKK. Nature, 412 (6844): 346–351. doi:10.1038/35085597.

Wang, H., Bloom, O., Zhang, M., et al. (1999) HMG-1 as a Late Mediator of Endotoxin Lethality in Mice. Science, 285 (5425): 248–251. doi:10.1126/science.285.5425.248.

Wang, H., Meng, H., Li, X., et al. (2017a) PELI1 functions as a dual modulator of necroptosis and apoptosis by regulating ubiquitination of RIPK1 and mRNA levels of c-FLIP. Proceedings of the National Academy of Sciences, 114 (45): 11944–11949. doi:10.1073/pnas.1715742114.

Wang, J. and Chen, Y. (2010) Role of the Zn2+ Motif of E1 in SUMO Adenylation. Journal of Biological Chemistry, 285 (31): 23732–23738. doi:10.1074/jbc.M110.114660.

Wang, J., Taherbhoy, A.M., Hunt, H.W., et al. (2010) Crystal Structure of UBA2ufd-Ubc9: Insights into E1-E2 Interactions in Sumo Pathways. PLoS ONE, 5 (12): e15805. doi:10.1371/journal.pone.0015805.

Wang, L., Qiao, Q., Ferrao, R., et al. (2017b) Crystal structure of human IRAK1. Proceedings of the National Academy of Sciences, 114 (51): 13507–13512. doi:10.1073/pnas.1714386114.

Warren, C. and Pavletich, N.P. (2022) Structure of the human ATM kinase and mechanism of Nbs1 binding. eLife, 11. doi:10.7554/eLife.74218.

Williams, C.J., Headd, J.J., Moriarty, N.W., et al. (2018) MolProbity: More and better reference data for improved all-atom structure validation. Protein Science, 27 (1): 293–315. doi:10.1002/pro.3330.

Williams, R.S., Lee, M.S., Hau, D.D., et al. (2004) Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. Nature Structural & Molecular Biology, 11 (6): 519–525. doi:10.1038/nsmb776.

Williams-Ashman, H.G. and Kennedy, E.P. (1952) Oxidative Phosphorylation Catalyzed by Cytoplasmic Particles Isolated from Malignant Tissues. Cancer Research, 12 (6): 415–421.

Winkles, J.A. and Alberts, G.F. (2005) Differential regulation of polo-like kinase 1, 2, 3, and 4 gene expression in mammalian cells and tissues. Oncogene, 24 (2): 260–266. doi:10.1038/sj.onc.1208219.

Winston, J.T., Strack, P., Beer-Romero, P., et al. (1999) The SCFbeta -TRCP-ubiquitin ligase complex associates specifically with phosphorylated destruction motifs in Ikappa Balpha and beta -catenin and stimulates Ikappa Balpha ubiquitination in vitro. Genes & Development, 13 (3): 270–283. doi:10.1101/gad.13.3.270.

Winter, G., Waterman, D.G., Parkhurst, J.M., et al. (2018) DIALS : implementation and evaluation of a new integration package. Acta Crystallographica Section D Structural Biology, 74 (2): 85–97. doi:10.1107/S2059798317017235.

Wu, S.-Y. and Cheng, C.-M. (2009) p53 sumoylation: Mechanistic insights from reconstitution studies. Epigenetics, 4 (7): 445–451. doi:10.4161/epi.4.7.10030.

Xu, R., Xu, Y., Huo, W., et al. (2018) Mitosis-specific MRN complex promotes a mitotic signaling cascade to regulate spindle dynamics and chromosome segregation. Proceedings of the National Academy of Sciences, 115 (43). doi:10.1073/pnas.1806665115.

Xu, Y., Plechanovová, A., Simpson, P., et al. (2014) Structural insight into SUMO chain recognition and manipulation by the ubiquitin ligase RNF4. Nature Communications, 5 (1): 4217. doi:10.1038/ncomms5217.

Xu, Y., Tao, X., Shen, B., et al. (2000) Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. Nature, 408 (6808): 111–115. doi:10.1038/35040600.

XU, Z. and AU, S.W.N. (2005) Mapping residues of SUMO precursors essential in differential maturation by SUMO-specific protease, SENP1. Biochemical Journal, 386 (2): 325–330. doi:10.1042/BJ20041210.

Yaffe, M.B. and Cantley, L.C. (2000) Mapping specificity determinants for protein-protein association using protein fusions and random peptide libraries. In pp. 157–170. doi:10.1016/S0076-6879(00)28397-3.

Yaffe, M.B., Rittinger, K., Volinia, S., et al. (1997) The Structural Basis for 14-3-3:Phosphopeptide Binding Specificity. Cell, 91 (7): 961–971. doi:10.1016/S0092-8674(00)80487-0.

Yamamoto, M., Sato, S., Hemmi, H., et al. (2003) TRAM is specifically involved in the Toll-like receptor 4–mediated MyD88-independent signaling pathway. Nature Immunology, 4 (11): 1144–1150. doi:10.1038/ni986.

Yan, L., Cui, Y. and Feng, J. (2023) Biology of Pellino1: a potential therapeutic target for inflammation in diseases and cancers. Frontiers in Immunology, 14. doi:10.3389/fimmu.2023.1292022.

Yan, W., Shao, Z., Li, F., et al. (2011) Structural basis of γH2AX recognition by human PTIP BRCT5-BRCT6 domains in the DNA damage response pathway. FEBS Letters, 585 (24): 3874–3879. doi:10.1016/j.febslet.2011.10.045.

Yang, S.-H., Galanis, A., Witty, J., et al. (2006) An extended consensus motif enhances the specificity of substrate modification by SUMO. The EMBO Journal, 25 (21): 5083–5093. doi:10.1038/sj.emboj.7601383.

Yasugi, T. (1996) Identification of the structural and functional human homolog of the yeast ubiquitin conjugating enzyme UBC9. Nucleic Acids Research, 24 (11): 2005–2010. doi:10.1093/nar/24.11.2005.

Yata, K., Lloyd, J., Maslen, S., et al. (2012) Plk1 and CK2 Act in Concert to Regulate Rad51 during DNA Double Strand Break Repair. Molecular Cell, 45 (3): 371–383. doi:10.1016/j.molcel.2011.12.028.

Yim, H. and Erikson, R.L. (2009) Polo-Like Kinase 1 Depletion Induces DNA Damage in Early S Prior to Caspase Activation. Molecular and Cellular Biology, 29 (10): 2609–2621. doi:10.1128/MCB.01277-08.

Yin, Q., Lin, S.-C., Lamothe, B., et al. (2009) E2 interaction and dimerization in the crystal structure of TRAF6. Nature Structural & Molecular Biology, 16 (6): 658–666. doi:10.1038/nsmb.1605.

YOKOYAMA, N. and MILLER, W.T. (2015) Molecular characterization of WDCP, a novel fusion partner for the anaplastic lymphoma tyrosine kinase ALK. Biomedical Reports, 3 (1): 9–13. doi:10.3892/br.2014.374.

You, Z., Chahwan, C., Bailis, J., et al. (2005) ATM Activation and Its Recruitment to Damaged DNA Require Binding to the C Terminus of Nbs1. Molecular and Cellular Biology, 25 (13): 5363–5379. doi:10.1128/MCB.25.13.5363-5379.2005.

Yu, Z., Vogel, G., Coulombe, Y., et al. (2012) The MRE11 GAR motif regulates DNA double-strand break processing and ATR activation. Cell Research, 22 (2): 305–320. doi:10.1038/cr.2011.128.

Yuan, Z., Zhang, X., Sengupta, N., et al. (2007) SIRT1 Regulates the Function of the Nijmegen Breakage Syndrome Protein. Molecular Cell, 27 (1): 149–162. doi:10.1016/j.molcel.2007.05.029.

Zemla, A. (2003) LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Research, 31 (13): 3370–3374. doi:10.1093/nar/gkg571.

Zemla, A., Zhou, C.E., Slezak, T., et al. (2005) AS2TS system for protein structure modeling and analysis. Nucleic Acids Research, 33 (Web Server): W111–W115. doi:10.1093/nar/gki457.

Zhang, B., Li, M., Chen, L., et al. (2009) The TAK1-JNK cascade is required for IRF3 function in the innate immune response. Cell Research, 19 (4): 412–428. doi:10.1038/cr.2009.8.

Zhang, H., Cao, X., Tang, M., et al. (2021) A subcellular map of the human kinome. eLife, 10. doi:10.7554/eLife.64943.

Zhang, J., Clark, K., Lawrence, T., et al. (2014) An unexpected twist to the activation of IKKβ: TAK1 primes IKKβ for activation by autophosphorylation. Biochemical Journal, 461 (3): 531–537. doi:10.1042/BJ20140444.

Zhang, M.S., Brunner, S.F., Huguenin-Dezot, N., et al. (2017) Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nature Methods, 14 (7): 729–736. doi:10.1038/nmeth.4302.

Zhang, S., Chang, L., Alfieri, C., et al. (2016) Molecular mechanism of APC/C activation by mitotic phosphorylation. Nature, 533 (7602): 260–264. doi:10.1038/nature17973.

Zhou, S. (1993) SH2 domains recognize specific phosphopeptide sequences. Cell, 72 (5): 767–778. doi:10.1016/0092-8674(93)90404-E.

Zhu, J., Zhu, S., Guzzo, C.M., et al. (2008) Small Ubiquitin-related Modifier (SUMO) Binding Determines Substrate Recognition and Paralog-selective SUMO Modification. Journal of Biological Chemistry, 283 (43): 29405–29415. doi:10.1074/jbc.M803632200.

Zhu, P., Mehl, R. and Cooley, R. (2022) Site-specific Incorporation of Phosphoserine into Recombinant Proteins in Escherichia coli. BIO-PROTOCOL, 12 (21). doi:10.21769/BioProtoc.4541.

Zwart, P.H., Grosse-Kunstleve, R.W. and Adams, P.D. (2005) Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation.

Appendix I – Primers

Primer	Gene	Sequence (5'-3')	Use(s)
pGEX_4T1_GST_PLK1_PBD_for	PLK1	TAATAGAATTCGGTGAGGTGGTCGACTGCC	Traditional cloning
pGEX_4T1_GST_PLK1_PBD_rev	PLK1	TAATACTCGAGTCAGGAGGCCTTGAGACGG	Traditional cloning
pGEX_4T1_GST_PLK1_PBD_H538Afor	PLK1	CTTCTTCCAGGATGCCACCAAGCTCATCTTG	Site-directed mutagenesis
pGEX_4T1_GST_PLK1_PBD_H538Arev	PLK1	CAAGATGAGCTTGGTGGCATCCTGGAAGAAG	Site-directed mutagenesis
pGEX_4T1_GST_PLK1_PBD_K540Mfor	PLK1	CCAGGATCACACCATGCTCATCTTGTGCC	Site-directed mutagenesis
pGEX_4T1_GST_PLK1_PBD_K540Mrev	PLK1	GGCACAAGATGAGCATGGTGTGATCCTGG	Site-directed mutagenesis
PLK1-Seq1for	PLK1	GCAACCTTTTCCTGAATGAAGATC	DNA sequencing
PLK1-Seq2for	PLK1	CAGCTGCACAGTGTCAATGCCTCC	DNA sequencing
PLK1-Seq3for	PLK1	GCTCCGCTACGCCCGCACTATGGTG	DNA sequencing
PLK1-E401Efor	PLK1	CAAGAGGAGGCTGAAGATCCTGCCTGC	Site-directed mutagenesis (BamHI-resistant)
PLK1-E401Erev	PLK1	GCAGGCAGGATCTTCAGCCTCCTCTTG	Site-directed mutagenesis (BamHI-resistant)
pcDNA3.1-3xHA-PBDfor	PLK1	TAATAAAGCTTGGTGAGGTGGTCGACTGC	Traditional cloning
PBD-R557Stopfor	PLK1	CATCGACGAGAAGTAGGACTTCCGCACATAC	Site-directed mutagenesis
PBD-R557Stoprev	PLK1	GTATGTGCGGAAGTCCTACTTCTCGTCGATG	Site-directed mutagenesis
NdeI-PBD-for	PLK1	TAATACATATGGGTGAGGTGGTCGACTGC	Traditional cloning
KpnI-PBD-for	PLK1	TAATAGGTACCGGTGAGGTGGTCGACTGC	Traditional cloning
HindIII-PBD-rev	PLK1	TAATAAAGCTTGGAGGCCTTGAGACGGTTG	Traditional cloning
XhoI-MBP-for	MBP-PLK1-His6	TAATACTCGAGATGAAAATCGAAGAAGGTAAACTGG	Traditional cloning
XhoI-His-rev	MBP-PLK1-His6	TAATACTCGAGTTAATGATGGTGGTGGTGGTGAAGC	Traditional cloning
Rad50_Seq1for	Rad50	ATGTCCCGGATCGAAAAGATGAG	DNA sequencing
Rad50_Seq2for	Rad50	ACTGATGAGCAACTAAATGACTTATATC	DNA sequencing
Rad50_Seq3for	Rad50	AGGAACTAGCTTCATCTGAGCAG	DNA sequencing
Rad50_Seq4for	Rad50	TCAGTCTTTGTACAGAGAGATAAAGG	DNA sequencing
Rad50_Seq5for	Rad50	CGAGGACGATGCAGTGCTG	DNA sequencing
Rad50T690Afor	Rad50	GAGAGTTTTTCAGGCAGAGGCTGAGTTAC	Site-directed mutagenesis
Rad50T690Arev	Rad50	GTAACTCAGCCTCTGCCTGAAAAACTCTC	Site-directed mutagenesis
Rad50T690Efor	Rad50	GAGAGTTTTTCAGGAAGAGGCTGAGTTAC	Site-directed mutagenesis
Rad50T690Erev	Rad50	GTAACTCAGCCTCTTCCTGAAAAACTCTC	Site-directed mutagenesis
Rad50T690Sfor	Rad50	GAGAGTTTTTCAGTCAGAGGCTGAGTTAC	Site-directed mutagenesis
Rad50T690Srev	Rad50	GTAACTCAGCCTCTGACTGAAAAACTCTC	Site-directed mutagenesis
Rad50T690TAGfor	Rad50	GAGAGTTTTTCAGTAGGAGGCTGAGTTAC	Site-directed mutagenesis
Rad50T690TAGrev	Rad50	GTAACTCAGCCTCCTACTGAAAAACTCTC	Site-directed mutagenesis
pMBP-3C-ZCCT690TAG-His6for	Rad50	ATAATAAGCTTAAAGAAATAAATCAAACAAGAGATAGGC	Traditional cloning
pMBP-3C-ZCCT690TAG-His6rev	Rad50	ATAATAAGCTTTTAGTGATGATGGTGGTGGTG	Traditional cloning
pRSF1030C-Rad50ZCC-His6for	Rad50	TAATAGAATTCATGAAAGAAATAAATCAAAC	Traditional cloning
pRSF1030C-Rad50ZCC-His6rev	Rad50	TAATAGGTACCTTAGTGATGATGGTGGTGGTGGTTCTTCAGAC	Traditional cloning
pcDNA5FRT-GFP-Rad50FL-BamHIf	Rad50	TAATAGGATCCATGTCCCGGATCGAAAAG	Traditional cloning
pcDNA5FRT-GFP-Rad50FL-XhoI-rev	Rad50	TAATACTCGAGTTAATGAACATTGAATCCC	Traditional cloning
pGEX4T1-Bub3for	Bub3	ATTATGAATTCATGACCGGTTCTAACGAG	Traditional cloning
pGEX4T1-Bub3rev	Bub3	ATTATCTCGAGTCAAGTACATGGTGACTTG	Traditional cloning
pGEX6P-1-N4BP2-PNKfor	N4BP2	TAATAGGATCCGGACTAGTTCTTGTTCTTC	Traditional cloning
pGEX6P-1-N4BP2-PNKrev	N4BP2	TAATACTCGAGTTAAATTTTCTCTGGAACCGAAG	Traditional cloning
pGEX6P-1-N4BP2-SMRfor	N4BP2	TAATAGGATCCTCTTTCCAGGACTTTGAG	Traditional cloning
pGEX6P-1-N4BP2-SMRrev	N4BP2	TAATACTCGAGTTACTTTAGCATGACTTTC	Traditional cloning
MRFAP1_BamHI_for	MRFAP1	TAATAGGATCCATGAGGCCCTTAGATATAG	Traditional cloning
MRFAP1_XhoI_rev	MRFAP1	TAATACTCGAGTCAGCTCGATTCGCTTTTTTC	Traditional cloning
pGEX6P1-CDK2-for	CDK2	TAATAGGATCCATGGAGAACTTCCAAAAGGTGG	Traditional cloning
pGEX6P1-CDK2-rev	CDK2	TAATACTCGAGTTAGAGTCGAAGATGGGG	Traditional cloning
pGEX6P1-CDK2AP1-for	CDK2AP1	TAATAGGATCCATGTCATATAAACCCAATCTAGC	Traditional cloning
pGEX6P1-CDK2AP1-rev	CDK2AP1	TAATACTCGAGTTAGCTACGCGCATTACGC	Traditional cloning
pGEX6P-1-CyclinF-for	Cyclin F	TAATAGGATCCATCTTGAGTCTCCCCGAA	Traditional cloning
pGEX6P-1-CyclinF-rev	Cyclin F	TAATACTCGAGTTATTGTGTCACTCCTAATGC	Traditional cloning
AC6-Sae1-KpnIfor	Sae 1	TAATAGGTACCGTGGAGAAGGAGGAGGCTG	Traditional cloning
AC6-Sae1-HindIIIrev	Sae 1	TAATAAAGCTTCTTGGGGCCAAGGCACTC	Traditional cloning

Appendix II – Expression Vectors

Plasmid	Antibiotic resistance	Affinity Tags and Protease Cleavage Sites	Origin	Promoter	Use(s)	Source
pGEX-4T-1	Ampicillin	N-terminal GST and thrombin cleavage.	pBR322	Tac	Recombinant protein expression.	Pre-existing in Smerdon lab.
pGEX-6P-1	Ampicillin	N-terminal GST and HRV 3C cleavage.	pBR322	Tac	Traditional cloning, site-directed mutagenesis, and recombinant protein expression.	Pre-existing in Smerdon lab.
pET-28b(+)	Kanamycin	N-terminal His_6, thombin cleavage, and C-terminal His_6.	pBR322	T7	Recombinant protein expression.	Pre-existing in Smerdon lab.
pRSF1030C	Chloramphenicol	C-terminal His_6 added by traditional cloning.	RSF	Ara	Traditional cloning, recombinant protein expression, co-expression, and site-directed phosphoserine/threonine incorporation.	A gift from John Cronan (University of Illinois).
W40 (pRSF-MBP-3C-His_6)	Ampicillin	N-terminal MBP, HRV 3C cleavage, C-terminal His_6.	RSF	Ara	Traditional cloning, recombinant protein expression, co-expression, and site-directed phosphoserine/threonine incorporation.	A gift from Jesse Rinehart (Yale University).
AC6 (optimised pRSF-MBP-3C-His_6)	Ampicillin	N-terminal MBP, HRV 3C cleavage, C-terminal His_6.	RSF	Ara	Traditional cloning, recombinant protein expression, co-expression, and site-directed phosphoserine/threonine incorporation.	A gift from Jesse Rinehart (Yale University).
pKW2-SepRS(2)/pSertRNA(B4)/EF-Sep	Chloramphenicol	-	pBR322	Tac	OTS components for site-directed phosphoserine incorporation.	A gift from Jason Chin (Addgene plasmid #173897).
pSerOTS-C1* (V70; pKD-pSerRS-EF-pSer)	Kanamycin	-	ColE1	glnS* and proK	OTS components for site-directed phosphoserine incorporation.	A gift from Jesse Rinehart (Addgene plasmid #188537).
pUC_pThrRS_SeptRNA$^{v2.0}_{CUA}$-OXB20-PduX_EFSep	Kanamycin	-	pMB1 derivative	OXB20	OTS components for site-directed phosphothreonine incorporation.	A gift from Jason Chin (Addgene plasmid #173899).
pThrOTS$_{Zeus}$ (T69)	Kanamycin	-	pBR322	TRC and proK	OTS components for site-directed phosphothreonine incorporation.	A gift from Jesse Rinehart.
pThrOTS$_{Zeus}$ Rop$^-$ (AE64)	Kanamycin	-	pBR322	TRC and proK	OTS components for site-directed phosphothreonine incorporation.	A gift from Jesse Rinehart.
pBK-PylRS	Kanamycin	-	ColE1	Tac	OTS components for site-directed pyrrolysine incorporation.	A gift from Jason Chin.
pCDF-PylT	Spectinomycin	-	pBR322	Tac	OTS components and expression plasmid for site-directed pyrrolysine incorporation.	A gift from Jason Chin.
pcDNA5-FRT-TO-GFP	Ampicillin/Hygromycin	N-terminal GFP.	pMB1 derivative	CMV/TetO2	Traditional cloning and site-directed mutagenesis, stable monoclonal cell line generation for mammalian overexpression.	Pre-existing in Smerdon lab.

Appendix III – Cloned Plasmids

Plasmid	Antibiotic resistance	Affinity Tags and Protease Cleavage Sites	Origin	Promoter	Use(s)	Source
pGEX-4T-1	Ampicillin	N-terminal GST and thrombin cleavage.	pBR322	Tac	Recombinant protein expression.	Pre-existing in Smerdon lab.
pGEX-4T-1-PLK1$_{367-603}$	“”	“”	“”	“”	Recombinant protein expression of PLK1-PBD domain proteins.	Cloned by Matthew Stewart.
pGEX4T-1-SUMO1$_{1-97}$	“”	“”	“”	“”	Recombinant protein expression of mature SUMO1.	A gift from Dr Alex Lanz.
pGEX4T-1-SUMO1$_{1-97\ S9CC52A}$	“”	“”	“”	“”	Recombinant protein expression of thiol-reactive mature SUMO1.	A gift from Dr Alex Lanz.
pGEX-4T-1-SUMO2$_{1-93}$	“”	“”	“”	“”	Recombinant protein expression of mature SUMO1.	A gift from Dr Alex Lanz.
pGEX-4T-1-SUMO2$_{1-93\ A2CC48A}$	“”	“”	“”	“”	Recombinant protein expression of thiol-reactive mature SUMO1.	A gift from Dr Alex Lanz.
pGEX-4T-1-HRV-3C	“”	“”	“”	“”	Recombinant protein expression of HRV 3C protease.	Pre-existing in Smerdon lab.
pGEX-6P-1	Ampicillin	N-terminal GST and HRV 3C cleavage.	pBR322	Tac	Traditional cloning, site-directed mutagenesis, and recombinant protein expression.	Pre-existing in Smerdon lab.
pGEX-6P-1-Peli 1$_{13-273}$	“”	“”	“”	“”	Recombinant protein expression of the Peli 1 FHA domain.	Cloned by Dr Éilís Bragginton.
pGEX-6P-1-Peli 2$_{15-275}$	“”	“”	“”	“”	Recombinant protein expression of the Peli 2 FHA domain.	Cloned by Dr Éilís Bragginton.
pGEX-6P-1-MDC1$_{1891-2086}$	“”	“”	“”	“”	Recombinant protein expression of the MDC1 tBRCT domain.	Cloned and gifted by Stucki lab (ETH Zurich).
pGEX-6P-1-PLK1$_{367-603}$	“”	“”	“”	“”	Recombinant protein expression of PLK1-PBD domain proteins. Site-directed mutagenesis to generate H538A-K540M mutant.	Cloned by Matthew Stewart.
pGEX-6P-1-Rad50$_{585-766}$	“”	“”	“”	“”	Recombinant protein expression of Rad50-ZCC fragment. Site-directed mutagenesis to generate T690/S/A/E/* mutants and traditional cloning.	Cloned by Matthew Stewart.
pGEX-6P-1-Bub3	“”	“”	“”	“”	Recombinant protein expression of Bub3.	Cloned by Matthew Stewart.
pGEX-6P-1-N4BP2$_{439-591}$	“”	“”	“”	“”	Recombinant protein expression of N4BP2 PNK domain. Cloning of D473A-D474A kinase-dead mutant	Cloned by Matthew Stewart.
pGEX-6P-1-N4BP2$_{1610-1770}$	“”	“”	“”	“”	Recombinant protein expression of N4BP2 SMR domain. Cloning of SMR-dead mutant.	Cloned by Matthew Stewart.
pGEX-6P-1-MRFAP1	“”	“”	“”	“”	Recombinant protein expression of MRFAP1.	Cloned by Matthew Stewart.
pGEX-6P-1-CDK2	“”	“”	“”	“”	Recombinant protein expression of CDK2.	Cloned by Matthew Stewart.
pGEX-6P-1-CDK2AP1	“”	“”	“”	“”	Recombinant protein expression of CDK2AP1.	Cloned by Matthew Stewart.
pGEX-6P-1-Cyclin F$_{32-544}$	“”	“”	“”	“”	Recombinant protein expression of structured Cyclin F fragment.	Cloned by Matthew Stewart.
pDB1/pGEX-3X-CK2a	Ampicillin	N-terminal GST and Factor Xa cleavage.	“”	“”	Recombinant protein expression of CK2 catalytic subunit.	A gift from Dr David Litchfield (Addgene plasmid #27083).
pAB46/pET-28a-CK2b	Kanamycin	N-terminal His_6 and thrombin cleavage.	pBR322	T7	Recombinant protein expression of CK2 regulatory subunit.	A gift from Dr David Litchfield (Addgene plasmid #27085).
pET-28b(+)	Kanamycin	N-terminal His_6, thrombin cleavage, and C-terminal His_6.	pBR322	T7	Recombinant protein expression.	Pre-existing in Smerdon lab.
pET-28b-PLK1$_{367-603}$	“”	“”	“”	“”	Recombinant protein expression of PLK1-PBD domain and cloning of H538A-K540M mutant.	Cloned by Matthew Stewart.
pET-28b-Sae2	“”	“”	“”	“”	Recombinant protein expression of Sae2.	A gift from Dr Alex Lanz.
pRSF1030C	Chloramphenicol	C-terminal His_6 added by traditional cloning.	RSF	Ara	Traditional cloning, recombinant protein expression, and site-directed phosphoserine/threonine incorporation.	A gift from Professor John Cronan (University of Illinois).
pRSF1030C-Rad50$_{585-766\ T690*}$-His_6	“”	“”	“”	“”	Recombinant protein expression of Rad50 ZCC fragment and site-directed phosphoserine/threonine incorporationat T690 position.	Cloned by Matthew Stewart.
W40/pRSF-MBP-3C-His_6	Ampicillin	N-terminal MBP, HRV 3C cleavage, C-terminal His_6.	RSF	Ara	Traditional cloning, recombinant protein expression, co-expression, and site-directed phosphoserine/threonine incorporation.	A gift from Professor Jesse Rinehart (Yale University).
W40/pRSF-MBP-3C-Chk2210-531 T383*-His6	""	""	""	""	Site-directed phosphothreonine incorporation.	A gift from Professor Jesse Rinehart (Yale University).
W40/pRSF-MBP-3C-Rad50$_{585-766\ T690*}$-His_6	“”	“”	“”	“”	Recombinant protein expression of Rad50 ZCC fragment, co-expression with CK2a subunit, and site-directed phosphoserine/threonine incorporationat T690 position.	Cloned by Matthew Stewart.

AC6/optimised pRSF-MBP-3C-His_6	Ampicillin	N-terminal MBP, HRV 3C cleavage, C-terminal His_6.	RSF	Ara	Traditional cloning, recombinant protein expression, co-expression, and site-directed phosphoserine/threonine incorporation.	A gift from Professor Jesse Rinehart (Yale University).
AC6/pRSF-MBP-3C-$Sae1_{S185}$-His_6	“”	“”	“”	“”	Recombinant protein expression of Sae1, and site-directed phosphoserine incorporation at S185 position.	Cloned by Matthew Stewart.
AC6/pRSF-MBP-3C-$Chk2_{T68*T383}$-His_6	“”	“”	“”	“”	Recombinant protein expression of Chk2, and double site-directed phosphothreonine incorporation at T68 and T383 positions.	Cloned by Matthew Stewart.
pKW2-SepRS(2)/pSertRNA(B4)/EF-Sep	Chloramphenicol	-	pBR322	Tac	OTS components for site-directed phosphoserine incorporation.	A gift from Professor Jason Chin (Addgene plasmid #173897).
pSerOTS-C1* (V70; pKD-pSerRS-EF-pSer)	Kanamycin	-	ColE1	glnS* and proK	OTS components for site-directed phosphoserine incorporation.	A gift from Professor Jesse Rinehart (Addgene plasmid #188537).
pUC_pThrRS_$SeptRNA^{v2.0}_{CUA}$_OXB20-PduX_EFSep	Kanamycin	-	pMB1 derivative	OXB20	OTS components for site-directed phosphothreonine incorporation.	A gift from Professor Jason Chin (Addgene plasmid #173899).
$pThrOTS_{2xu5}$(T69)	Kanamycin	-	pBR322	TRC and proK	OTS components for site-directed phosphothreonine incorporation.	A gift from Professor Jesse Rinehart (Yale University).
$pThrOTS_{2xu5}$ Rop^- (AE64)	Kanamycin	-	pBR322	TRC and proK	OTS components for site-directed phosphothreonine incorporation.	A gift from Professor Jesse Rinehart (Yale University).
pBK-PylRS	Kanamycin	-	ColE1	Tac	OTS components for site-directed pyrrolysine incorporation.	A gift from Professor Jason Chin (University of Cambridge).
pCDF-PylT	Spectinomycin	-	pBR322	Tac	OTS components and expression plasmid for site-directed pyrrolysine incorporation.	A gift from Jason Chin (University of Cambridge).
pCDF-PylT-MBP-3C-$PLK1_{367\text{-}603\ R557}$-$His_6$	“”	N-terminal MBP, 3C protease cleavage, and C-terminal His_6.	“”	“	OTS components, recombinant expression of PLK1-PBD domain for site-directed incorporation of pyrrolysine at R557 position.	Cloned by Matthew Stewart.
pcDNA5-FRT-TO-GFP	Ampicillin/Hygromycin	N-terminal GFP.	pMB1 derivative	CMV/TetO2	Traditional cloning and site-directed mutagenesis, stable monoclonal cell line generation for mammalian overexpression.	A gift from Professor Grant Stewart.
pcDNA5-FRT-TO-GFP-Rad50	“”	“”	“”	“”	Stable cell line generation for tetracycline/doxycycline overexpression of GFP-tagged Rad50.	Cloned by Matthew Stewart.
pcDNA3.1(+)-$PLK1_{367\text{-}603}$-HA_3	Ampicillin	C-terminal HA_3	pBR322 and SV40	CMV and SV40	Transient overexpression of HA-tagged proteins and traditional cloning.	A gift from Professor Grant Stewart.
pcDNA3.1(+)-$PLK1_{367\text{-}603}$-HA_3	“”	“”	“”	“”	Transient overexpression of HA-tagged PLK1 PBD domain and traditional cloning of H538A-K540M mutant.	Cloned by Matthew Stewart.

Recombinant Protein	Expression System	Lysis Buffer	Affinity Beads 1	Elution 1	Affinity Beads 2	Elution 2	Theoretical Molecular Weight (Da)	Theoretical ε (M-1cm-1)
GST-Peli 1_{13-273}	BL21 (DE3), LB, 1mM IPTG, 18°C O/N	50mM Hepes pH 7.5, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 1% (v/v) recombinant GST-HRV-3C protease (yielding Peli 1_{H6}).	-	-	55531.55 (GST-tagged), 29118.84 (untagged)	67310 (GST-tagged, oxidised), 66810 (GST-tagged, reduced), 24200 (untagged, oxidised), 23950 (untagged, reduced)
GST-Peli 2_{15-275}	BL21 (DE3), LB, 1mM IPTG, 18°C O/N	50mM Hepes pH 7.5, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 1% (v/v) recombinant GST-HRV-3C protease (yielding Peli 2_{H6}).	-	-	55675.58 (GST-tagged), 28851.41 (untagged)	70290 (GST-tagged, oxidised), 69790 (GST-tagged, reduced), 27810 (untagged, oxidised), 26930 (untagged, reduced)
GST-MDC1$_{1891-2086}$	BL21 (DE3), LB, 1mM IPTG, 18°C O/N	50mM Hepes pH 7.5, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 1% (v/v) recombinant GST-HRV-3C protease (yielding MDC1$_{1891CT}$).	-	-	48310.15 (GST-tagged), 21897.45 (untagged)	56310 (GST-tagged, oxidised), 55810 (GST-tagged, reduced), 13200 (untagged, oxidised), 12950 (untagged, reduced)
GST-PLK1$_{367-603}$	BL21 (DE3), LB, 1mM IPTG, 18°C O/N	50mM Hepes pH 8.0, 500mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Option 1: Lysis Buffer supplemented with 1% (v/v) recombinant GST-HRV-3C protease (yielding PLK1PBD).	-	-	54433.65 (GST-tagged), 28020.94 (untagged)	79355 (GST-tagged, oxidised), 78730 (GST-tagged, reduced), 36245 (untagged, oxidised), 35870 (untagged, reduced)
				Option 2: Lysis Buffer supplemented with 25mM reduced L-glutathione (Sigma-Aldrich #G4251) (yielding GST-PLK1PBD).	-	-		
GST-Rad50$_{585-766}$	BL21 (DE3), LB, 1mM IPTG, 18°C O/N	50mM Hepes pH 8.0, 300mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Option 1: Lysis Buffer supplemented with 1% (v/v) recombinant GST-HRV-3C protease (yielding Rad50ZCC).	-	-	48064.24 (GST-tagged), 21537.43 (untagged)	46340 (GST-tagged, oxidised), 45840 (GST-tagged, reduced), 3230 (untagged, oxidised), 2980 (untagged, reduced)
				Option 2: Lysis Buffer supplemented with 25mM reduced L-glutathione (Sigma-Aldrich #G4251) (yielding GST-Rad50ZCC).	-	-		
GST-CK2a	BL21 (DE3), 2TY, 1mM IPTG, 18°C O/N	50mM Hepes pH 8.0, 300mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 25mM reduced L-glutathione (Sigma-Aldrich #G4251) (yielding GST-CK2a).	Optional for CK2 tetramer purification: TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM imidazole (yielding CK2 tetramer).	71547.26 (GST-tagged)	105005 (GST-tagged, oxidised), 104630 (GST-tagged, reduced)
GST-Bub3	BL21 (DE3), LB/TB/2TY, 1mM IPTG screened	50mM Hepes pH 8.0, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Both options failed, despite high expression. Appears likely that co-purification with BubR1 is required.	-	-	63978.94 (GST-tagged), 37566.24 (untagged)	97470 (GST-tagged, oxidised), 96720 (GST-tagged, reduced), 54360 (untagged, oxidised), 53860 (untagged, reduced)
GST-SUMO1_{1-97}	BL21 (DE3), 2TY, 1mM IPTG, 18°C O/N	50mM Hepes pH 8, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 1% (v/v) thrombin (gifted by Dr Alex Lanz) (yielding mature SUMO1).	-	-	37425 (GST-tagged), 11276.66 (untagged)	47580 (GST-tagged, oxidised), 47330 (GST-tagged, reduced), 4470 (untagged)
SUMO$1_{1-97 59C C52A}$	BL21 (DE3), 2TY, 1mM IPTG, 18°C O/N	50mM Hepes pH 8, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 1% (v/v) thrombin (gifted by Dr Alex Lanz) (yielding mature thiol-reactive SUMO1).	-	-	**	**
SUMO2_{1-93}	BL21 (DE3), 2TY, 1mM IPTG, 18°C O/N	50mM Hepes pH 8, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 1% (v/v) thrombin (gifted by Dr Alex Lanz) (yielding mature SUMO2).	-	-	36901.47 (GST-tagged), 10753.04 (untagged)	44600 (GST-tagged, oxidised), 44350 (GST-tagged, reduced), 1490 (untagged)
SUMO$2_{1-93 A2CC48A}$	BL21 (DE3), 2TY, 1mM IPTG, 18°C O/N	50mM Hepes pH 8, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 1% (v/v) thrombin (gifted by Dr Alex Lanz) (yielding mature thiol-reactive SUMO2).	-	-	**	**
GST-MRFAP1	BL21 (DE3), LB, 1mM IPTG, 18°C O/N	50mM Hepes pH 7.5, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Single attempt where aggregates formed on gel-filtration column.	-	-	41473.78 (GST-tagged), 15061.07 (untagged)	50100 (GST-tagged, oxidised), 49850 (GST-tagged, reduced), 6990 (untagged)
GST-Cyclin F	BL21 (DE3), LB/TB/2TY/LB Autoinduction media (Formedium # AIMLB0205), 1mM IPTG, 18°C O/N	50mM Hepes pH 7.5, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Expression very low. Plasmid should be re-cloned to include unstructured regions or different expression vectors.	-	-	84873.16 (GST-tagged)	127085 (GST-tagged, oxidised), 125710 (GST-tagged, reduced)
GST-N4BP$2_{439-591}$	BL21 (DE3), LB, 1mM IPTG, 18°C O/N	50mM Hepes pH 8.0, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 1% (v/v) recombinant GST-HRV-3C protease (yielding N4BP2_{H6}).	-	-	44694.79 (GST-tagged), 18282.09 (untagged)	75540 (GST-tagged, oxidised), 75290 (GST-tagged, reduced), 32430 (untagged)
GST-N4BP$2_{1610-1770}$	BL21 (DE3), LB, 1mM IPTG, 18°C O/N	50mM Hepes pH 8.0, 150mM NaCl, 0.5mM TCEP	Glutathione Sepharose™ 4B (Cytiva #GE17-0756-01)	Lysis Buffer supplemented with 1% (v/v) recombinant GST-HRV-3C protease (yielding N4BP2_{SH6}).	-	-	45337.43 (GST-tagged), 18924.72 (untagged)	55155 (GST-tagged, oxidised), 54780 (GST-tagged, reduced), 12045 (untagged, oxidised), 11920 (untagged, reduced)
His_6-PLK1$_{367-603}$	BL21 (DE3), LB, 1mM IPTG, 18°C O/N	50mM Hepes pH 7.5, 300mM NaCl, 10mM Imidazole, 0.5mM TCEP	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole.	-	-	29530.59 (His6-tagged)	36245 (His6-tagged, oxidised), 35870 (His6-tagged, reduced)
His_6-CK2b	BL21 (DE3), 2TY, 1mM IPTG, 18°C O/N	50mM Hepes pH 8.0, 300mM NaCl, 0.5mM TCEP	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole.	-	-	29701.66 (His6-tagged)	30745 (His6-tagged, oxidised), 30730 (His6-tagged, reduced)
His_6-Sae2	BL21 (DE3), LB/2TY/TB, 1mM IPTG, 18°C O/N (mix pellets with MBP-3C-Sae1_{5185}-His_6 for maximum yield)	50mM Hepes pH 8.0, 300mM NaCl, 10mM Imidazole, 0.5mM TCEP	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole. Protein is very unstable without Sae1.	-	-	73518.15 (His6-tagged)	50015 (His6-tagged, oxidised), 49390 (His6-tagged, reduced)
Rad50$_{585-766}$-His_6	BL21 (DE3), LB, 1mM IPTG, 18°C O/N (good expression for non-phosphorylated proteins)	50mM Hepes pH 8.0, 300mM NaCl, 10mM Imidazole, 0.5mM TCEP	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole. Attempts at gel-filtration and ion-exchange chromatography failed to remove *E. coli* endogenous His_6-containing contaminants for phospho-incorporated proteins.	-	-	22194.11 (His6-tagged full-length), 12292.72 (T690 truncation)	3230 (His6 tagged full-length, oxidised), 2980 (His6-tagged full-length, reduced)
	BL21 DserB + pKW2-SepRS(2)/pSertRNA(B4)/EF-Sep LB/TB/2TY, 1mM IPTG, 1%/2% (w/v) L-arabinose, 2mM O-Phospho-L-Serine, 37°C/18°C, O/N and 4 hours				-	-		
	BL21 DserC DycdX + pUC_pThrRS_SeptRNAv2.0CUA_OXB20-PduX_EFSep, 1mM IPTG, 1%/2% (w/v) L-arabinose, 2mM O-Phospho-L-Serine, 37°C/18°C, O/N and 4 hours				-	-		
	C321 .DA.exp DycdX DserC DpphA + pThrOTSZeus (T69), LB/TB/2TY, 1mM IPTG, 1%/2% (w/v) L-arabinose, 2mM O-Phospho-L-Serine, 37°C/18°C, O/N and 4 hours (good expression for site-specific incorporation of pT690).				-	-		
	BL21(DE3) DycdX DserC DpphA + pDB1-CK2a, LB/TB, 1mM IPTG, 2% (w/v) L-arabinose (good co-expression)				-	-		
(W40) MBP-3C-Chk2$_{210-531\ T383*}$-His_6	C321 .DA.exp DycdX DserC DpphA + pThrOTS$_{Zeus}$ (T69) LB, 1mM IPTG, 2% (w/v) L-arabinose, 37°C, 4 hours (good expression for site-specific incorporation of pT383).	50mM Hepes pH 8.0, 300mM NaCl, 10mM Imidazole, 0.5mM TCEP	Amylose Resin (New England Biolabs #E8021S)	Lysis Buffer supplemented with 10mM D-(+)-Maltose monohydrate (Sigma-Aldrich #M5895) (yielding Chk2$_{210-531\ T383*}$-His_6).	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole (yielding Chk2$_{210-531\ T383*}$-His_6).	79039.47 (MBP-3C-His6-tagged, full-length), 61284.35 (MBP-3C-tagged, 383 truncation)	99740 (MBP-3C-His6-tagged, full-length, oxidised), 99240 (MBP-3C-His6-tagged, full-length, reduced)
(W40) MBP-3C-Rad50$_{585-766}$-His_6	C321 .DA.exp DycdX DserC DpphA + pThrOTS$_{Zeus}$ (T69)/pThrOTS$_{Zeus}$ Rop- (AE64), LB/TB/2TY/NZY/MDA5052, 1mM IPTG, 1%/2% (w/v) L-arabinose, 37°C/18°C, O/N and 4 hours (good expression for site-specific incorporation of pT690).	50mM Hepes pH 8.0, 300mM NaCl, 10mM Imidazole, 0.5mM TCEP	Amylose Resin (New England Biolabs #E8021S)	Option 1: Lysis Buffer supplemented with 1% (v/v) recombinant GST-HRV-3C protease (yielding Rad50ZCC-His6).	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole (yielding MBP-3C-Rad50$_{ZCC-p5/T690}$-His_6).	64356.97 (MBP-3C-His6-tagged, full-length), 54427.57 (MBP-3C-tagged, T690 truncation)	69580 (MBP-3C-His6-tagged, full-length, oxidised), 69330 (MBP-3C-His6-tagged, full-length, reduced)
	BL21(DE3) DycdX DserC DpphA + pThrOTS$_{Zeus}$ (T69)/pThrOTS$_{Zeus}$ Rop- (AE64), LB/TB/2TY/NZY/MDA5052, 1mM IPTG, 1%/2% (w/v) L-arabinose, 37°C/18°C, O/N and 4 hours (good expression for site-specific incorporation of pT690).			Option 2: Lysis Buffer supplemented with 10mM D-(+)-Maltose monohydrate (Sigma-Aldrich #M5895) (yielding MBP-3C-Rad50ZCC-His6).	-	-		
	BL21(DE3) DycdX DserC DpphA + pDB1-CK2a, LB/TB, 1mM IPTG, 2% (w/v) L-arabinose (good co-expression).				-	-		

(AC6) MBP-3C-Rad50$_{635-766}$-His$_6$	C321 .DA.exp DycdX DserC DpphA + pThrOTS$_{2ass}$ (T69)/pThrOTS$_{2ass}$ Rop- (AE64), LB/TB/2TY/NZY/MDA5052, 1mM IPTG, 1%/2% (w/v) L-arabinose, 37°C/18°C, O/N and 4 hours (good expression for site-specific incorporation of pT690).	50mM Hepes pH 8.0, 300mM NaCl, 10mM Imidazole, 0.5mM TCEP	Amylose Resin (New England Biolabs #E8021S)	Option 1: Lysis Buffer supplemented with 1% (vV) recombinant GST-HRV-3C protease (Rad50ZCC-pS/T690-His6).	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole (yielding either MBP-3C-Rad50 $_{ZCC\text{-}pS/T690}$-His$_6$ or Rad50 $_{ZCC\text{-}pS/T690}$-His$_6$).	**	**
	BL21(DE3) DycdX DserC DpphA + pThrOTS$_{2ass}$ (T69)/pThrOTS$_{2ass}$ Rop- (AE64), LB/TB/2TY/NZY/MDA5052, 1mM IPTG, 1%/2% (w/v) L-arabinose, 37°C/18°C, O/N and 4 hours (good expression for site-specific incorporation of pT690).			Option 2: Lysis Buffer supplemented with 10mM D-(+)-Maltose monohydrate (Sigma-Aldrich #M5895) (yielding MBP-3C-Rad50ZCC-His6).	-	-		
	BL21(DE3) DycdX DserC DpphA + pDB1-CK2a, LB/TB, 1mM IPTG, 2% (w/v) L-arabinose (good co-expression)				-	-		
	rEcoli XpS + pSerOTS-C1* + 1mM IPTG + 2% (w/v) L-arabinose + 2mM O-Phospho-L-Serine, 2TY, 37°C, 4 hours (excellent expression for site-specific incorporation of pS690).				-	-		
(AC6) MBP-3C-Sae1$_{S185}$-His$_6$	rEcoli XpS + pSerOTS-C1* + 1mM IPTG + 2% (w/v) L-arabinose + 2mM O-Phospho-L-Serine, LB/2TY/TB, 37°C, 4 hours (excellent expression for site-specific incorporation of pS690). Mix cell pellets with His$_6$-Sae2 for maximum yield.	50mM Hepes pH 8.0, 300mM NaCl, 10mM Imidazole, 0.5mM TCEP	Amylose Resin (New England Biolabs #E8021S)	Option 1: Lysis Buffer supplemented with 1% (vV) recombinant GST-HRV-3C protease (Sae1pS185-His6:His6-Sae2).	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole (yielding Sae1$_{pS185}$-His$_6$:His$_6$-Sae2).	80922.11 (MBP-3C-His6-tagged, full-length), 61852.40 (MBP-3C-tagged, S185 truncation), 39668.20 (His6-tagged, full-length)	86790 (MBP-3C-His6-tagged, full-length, oxidised), 86290 (MBP-3C-His6-tagged, full-length, reduced), 20440 (His6-tagged, full-length, oxidised), 19940 (His6-tagged, full-length, reduced)
(W40) MBP-3C-Chk2$_{210-531\ T383*}$-His$_6$	C321 .DA.exp DycdX DserC DpphA + pThrOTS$_{2ass}$ (T69) LB, 1mM IPTG, 2% (w/v) L-arabinose, 37°C, 4 hours (good expression for site-specific incorporation of pT383).	50mM Hepes pH 8.0, 300mM NaCl, 10mM Imidazole, 0.5mM TCEP	Amylose Resin (New England Biolabs #E8021S)	Lysis Buffer supplemented with 10mM D-(+)-Maltose monohydrate (Sigma-Aldrich #M5895) (yielding Chk2$_{210-531\ T383*}$-His$_6$).	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole (yielding Chk2$_{210-531\ T383*}$-His$_6$)..	79039.47 (MBP-3C-His6-tagged, full-length), 61284.35 (MBP-3C-tagged, 383 truncation)	99740 (MBP-3C-His6-tagged, full-length, oxidised), 99240 (MBP-3C-His6-tagged, full-length, reduced)
(AC6) MBP-3C-Chk2$_{T68*-T383*}$-His$_6$	C321 .DA.exp DycdX DserC DpphA + pThrOTS$_{2ass}$ (T69)/pThrOTS$_{2ass}$ Rop- (AE64), LB/TB/2TY/NZY/MDA5052, 1mM IPTG, 1%/2% (w/v) L-arabinose, 37°C/18°C, O/N and 4 hours.	50mM Hepes pH 8.0, 300mM NaCl, 10mM Imidazole, 0.5mM TCEP	Amylose Resin (New England Biolabs #E8021S)	Lysis Buffer supplemented with 10mM D-(+)-Maltose monohydrate (Sigma-Aldrich #M5895) (yielding MBP-3C-Chk2$_{pT68-pT383*}$-His$_6$). Purification not an issue.	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole (yielding MBP-3C-Chk2$_{pT68-pT383*}$-His$_6$).	103470.09 (MBP-3C-His6-tagged, full-length), 48142.94 (MBP-3C-tagged, T68 truncation), 84053.47 (MBP-3C-tagged, pT8, T383 truncation)	126920 (MBP-3C-His6-tagged, full-length, oxidised), 126170 (MBP-3C-His6-tagged, full-length, reduced)
	BL21(DE3) DycdX DserC DpphA + pThrOTS$_{2ass}$ (T69)/pThrOTS$_{2ass}$ Rop- (AE64), LB/TB/2TY/NZY/MDA5052, 1mM IPTG, 1%/2% (w/v) L-arabinose, 37°C/18°C, O/N and 4 hours.				-	-		
	All conditions tested resulted in pT68 incorporation, but little to no pT383 incorporation. University of Birmingham advanced mass spectrometry facility unable to validate.				-	-		
MBP-3C-PLK1$_{367-603\ R557*}$-His$_6$	BL21 (DE3) + pBK-PylRS, LB/TB/2TY/NZY/MDA5052, 1mM IPTG, 1%/2% (w/v) L-arabinose, 0.5mM AbK	50mM Hepes pH 8.0, 500mM NaCl, 10mM Imidazole, 0.5mM TCEP	Amylose Resin (New England Biolabs #E8021S)	Lysis Buffer supplemented with 10mM D-(+)-Maltose monohydrate (Sigma-Aldrich #M5895) (yielding MBP-3C-PLK1$_{pBD\text{-}R557AbK*}$-His$_6$). Unclear from time constraints whether purification is an issue.	TALON® Superflow™ (Cytiva #28957499)	Lysis Buffer supplemented with 250mM Imidazole (yielding MBP-3C-Chk2$_{pT68-pT383*}$-His$_6$).	69708.34 (MBP-3C-His6-tagged, full-length), 63146.80 (MBP-3C-tagged, R557 truncation)	102595 (MBP-3C-His6-tagged, full-length, oxidised), 102220 (MBP-3C-His6-tagged, full-length, reduced)
	BL21(DE3) DycdX DserC DpphA + pBK-PylRS, LB/TB/2TY/NZY/MDA5052, 1mM IPTG, 1%/2% (w/v) L-arabinose, 0.5mM AbK				-	-		
	All conditions resulted in little R557AbK incorporation. University of Birmingham advanced mass spectrometry facility unable to validate.				-	-		

Appendix V – PBD Crystal Structure Attempts

Screen	Well	Composition	Cryoprotectant	Crystal Dataset
ProPlexä HT-96	A10	0.1 M Potassium chloride, 0.1 M Tris pH 8.0, and 15% (w/v) PEG 2000 MME	30% (v/v) Glycerol	
	F1	0.1 M Na HEPES pH 7.0 and 20% (w/v) PEG 8000	30% (v/v) Glycerol	
The LMB Crystallisation Screenä HT-96	D1	7.5% (w/v) PEG 20000, 7.5% (v/v) PEG 550 MME, 0.1 M Tris pH 7.5, and 80 mM Sodium formate	30% (v/v) Glycerol	99036844_fast_dp; P 1 2 1; 2.825-28.263Å
	D7	15% (w/v) PEG 3350 and 0.1 M MES pH 6.2	30% (v/v) Glycerol	99037090_fast_dp; P 1 2 1; 2.22-28.73Å
	F9	18% (w/v) PEG 5000 MME, 0.1 M MES pH 6.5, and 0.2 M Ammonium sulphate	30% (v/v) Glycerol and 10mM Biotin-Rad50-pT690	99039958_mx34438v6_xMSPDB690LMBG101_free; P 1 21 1; 2.355-93.16Å
	G10	10% (w/v) PEG 8000, 0.1 M HEPES pH 7.5, and 9% (v/v) Ethylene glycol	25% (v/v) Ethylene glycol	101401221_mx34438v80_xMattSPT690LMBF911_free; P 2 21 21; 2.209-59.763Å
Morpheus® HT-96	G3	0.1M Carboxylic acids (0.2M Sodium formate, 0.2M Ammonium acetate, 0.2M Sodium citrate tribasic dihydrate, 0.2M Potassium sodium tartrate tetrahydrate, and 0.2M Sodium oxamate), 0.1M Buffer System 1 pH 6.5 (Imidazole and MES), and 50% (v/v) Precipitant Mix 3 (40% (v/v) Glycerol and 20% (w/v) PEG4000)	N/A	
The BCS Screen HT-96	D3	0.15M Ammonium nitrate, 0.1M MES pH 6.0, 20% (v/v) PEG Smear Medium, and 5% (v/v) Ethylene glycol	25% (v/v) Ethylene glycol and 10mM Biotin-Rad50-pT690	
	D11	0.1M Ammonium sulfate, 50mM Magnesium sulfate heptahydrate, 0.1M Sodium citrate pH 5.5, and 22.5% (v/v) PEG Smear Medium	25% (v/v) Ethylene glycol and 10mM Biotin-Rad50-pT690	

Appendix VI – AlphaFold Primary Sequence

Model	Predictive Webserver	Primary Sequence Input	Comment
Pellino 1FHA:MDC1pT765 AFM	ColabFold	APVKYGELIVLGYNGSLPNGDRGRRKSRFALFKRPKANGV KPSTVHIACTPQAAKAISNKDQHSISYTLSRAQTVVVEYTHDS NTDMFQIGRSTESPIDFVVTDTVPGSQSNSDTQSVQSTISRF ACRIICERNPPFTARIYAAGFDSSKNIFLGEKAAKWKTSDGQ MDGLTTNGVLVMHPRNGFTEDSKPGIWREISVCGNVFSLR ETRSAQQRGKMVEIETNQLQDGSLIDLCGATLLWRTAEGL SHTPTVKHLEALRQ:EDSETQPFD	MDC1T765 phosphorylated in COOT.
Pellino 1FHA:MDC1pT765 AF3	AlphaFold3	APVKYGELIVLGYNGSLPNGDRGRRKSRFALFKRPKANGV KPSTVHIACTPQAAKAISNKDQHSISYTLSRAQTVVVEYTHDS NTDMFQIGRSTESPIDFVVTDTVPGSQSNSDTQSVQSTISRF ACRIICERNPPFTARIYAAGFDSSKNIFLGEKAAKWKTSDGQ MDGLTTNGVLVMHPRNGFTEDSKPGIWREISVCGNVFSLR ETRSAQQRGKMVEIETNQLQDGSLIDLCGATLLWRTAEGL SHTPTVKHLEALRQ:EDSEpTQPFD	-
PLK1PBD:Rad50pT690 AFM	ColabFold	GEVVDCHLSDMLQQLHSVNASKPSERGLVRQEEAEDPAC IPIFWVSKWVDYSDKYGLGYQLCDNSVGVLFNDSTRLILYN DGDSLQYIERDGTESYLTVSSHPNSLMKKITLLKYFRNYMS EHLLKAGANITPREGDELARLPYLRTWFRTRSAIILHLSNGS VQINFFQDHTKLILCPLMAAVTYIDEKRDFRTYRLSLLEEYG CCKELASRLRYARTMVDKLLSSRSASNRLKAS:RVFQpTEA EL	Rad50T690 phosphorylated in COOT. Model binned due to low iPTM score.
PLK1PBD:Rad50E690 AFM	ColabFold	GEVVDCHLSDMLQQLHSVNASKPSERGLVRQEEAEDPAC IPIFWVSKWVDYSDKYGLGYQLCDNSVGVLFNDSTRLILYN DGDSLQYIERDGTESYLTVSSHPNSLMKKITLLKYFRNYMS EHLLKAGANITPREGDELARLPYLRTWFRTRSAIILHLSNGS VQINFFQDHTKLILCPLMAAVTYIDEKRDFRTYRLSLLEEYG CCKELASRLRYARTMVDKLLSSRSASNRLKAS:RVFQEEA EL	-
PLK1PBD:Rad50pT690 AF3	AlphaFold3	GEVVDCHLSDMLQQLHSVNASKPSERGLVRQEEAEDPAC IPIFWVSKWVDYSDKYGLGYQLCDNSVGVLFNDSTRLILYN DGDSLQYIERDGTESYLTVSSHPNSLMKKITLLKYFRNYMS EHLLKAGANITPREGDELARLPYLRTWFRTRSAIILHLSNGS VQINFFQDHTKLILCPLMAAVTYIDEKRDFRTYRLSLLEEYG CCKELASRLRYARTMVDKLLSSRSASNRLKAS:RVFQpTEA EL	-