# Structural insights into human SNF2/SWI2 chromatin remodeler SMARCAD1 and its role in DNA repair




Antonio Jose Biasutto
Medical Sciences Doctoral Training Centre
Lincoln College, University of Oxford

A Thesis submitted in partial fulfilment of the requirements
for the degree of *Doctor of Philosophy in Structural Biology*

Trinity Term 2016

# Structural insights into human SNF2/SWI2 chromatin remodeler SMARCAD1 and its role in DNA repair

*A thesis submitted for the degree of DPhil in Structural Biology*

## Abstract

ATP-dependent chromatin remodelers have been proposed to act sequentially, and to a certain extent non-redundantly, in the priming stages of the DNA Damage Response pathways by establishing chromatin in lesion sites ready to act as a scaffold for repair factors or to be displaced in order to allow DNA repair. Among remodeling factors proposed to play a role in DNA repair is SMARCAD1, a poorly characterized, non-canonical member of the SWR1-like family of SNF2/SWI2 superfamily of ATPases, which has recently been identified as a potential target for ATM/ATR phosphorylation at canonical and non-canonical sites upon DNA damage.

The actual mechanism for SMARCAD1 recruitment and involvement in DNA remodeling is still unknown, and unlike most other chromatin remodelers, SMARCAD1 does not contain DNA- or histone-binding domains frequently accompanying such proteins. Instead, in addition to the core ATPase domain, only two CUE domains (a type of helical ubiquitin-binding domain) have been identified.

This thesis presents the findings of an investigation intended to structurally characterize SMARCAD1 by dissecting and identifying its domain architecture, and examining the activity and ligand selectivity of its binding domains in the functional context of DNA damage repair. The solution NMR structure of the CUE1 domain is presented, describing a triple helix bundle consistent with other members of the family. Furthermore, a novel SUMO interacting motif was identified and through a combination of NMR titrations and phosphor-proteomics analysis, shown to be constitutively phosphorylated which excludes the possibility of DNA damage dependent ATM targeting as the recruitment mechanism for DNA repair.

Additionally, it is demonstrated that both CUE domains are poor binders of mono-ubiquitin, however CUE1 specifically mediates the high affinity binary interaction with the transcriptionally repressive master regulator KAP1. This interaction was shown to be independent of post-translational ubiquitylation but rather sustained through direct interaction with the dimeric RBCC domain of KAP1. Finally, mass spectrometry profiling of domain-dependent interactions (based on differential abundance relative to changes due to chemically induced DNA damage) suggests SMARCAD1 may be involved in p53 transcriptional regulation through interactions maintained with CUE1 prior to DNA damage, whereas the SIM domain selectively targets protein interactions upon DNA damage that simultaneously activate p53 transcriptional control and recruit SMARCAD1 to DNA damage repair pathways.

Antonio J Biasutto
Lincoln College

Trinity Term 2016

# Acknowledgments

I have been incredibly fortunate to find my amazing supervisors, Erika Mancini and Christina Redfield, to who (and for whom) I am incredibly grateful. Erika has been a source of renewing energy and curiosity, effectively recharging me after every meeting. I am so appreciative of the commitment and support she showed even when she was not down the hall anymore; she never made it seem hard. Christina was a surprising breath of fresh air; I can't count how many times she has shared laughter and wisdom with me. Thank you both for taking me on and becoming true role models of the balance between excellence and kindness.

Looking back, joining STRUBI was the greatest decision I could have made for my career after biting the bullet and daring to apply to Oxford. To the many people who shared space, knowledge, time and hardship with me, thank you. I owe a special note to the (mostly past) members of the Mancini and Gilbert groups: Luke Yates, David Aparicio, Mika Abu, Kai Deng, Xiaoyang Wang and Monica Mantri were teachers and friends. Thank you to Tao Ni, Dimple Karia and Max Renner for sharing the trip of navigating what sometimes felt to be impossible odds. Thank you to Margaret Jones, Weixian Lu, Yuguang Zhao and Tom Walter for help and assistances too many times to count. My deepest gratitude to Robert Gilbert for offering me shelter and the continuity of bi-weekly meetings at ungodly hours of the morning; they made all the difference. Finally, thank you to Prof Jason Schnell and Dr Patrick Varga-Weisz for a surprisingly enjoyable *viva voce* experience.

At the end of four years there are two instances in which "thank you" doesn't quite cover it. To the Wellcome Trust DPhil Programme committee, who took a chance on me and literally changed the outlook of my professional future, my sincerest gratitude. Finally, to my amazing, brilliant, beautiful wife, Maria Alejandra: through thick and thin, close together and far apart, you have been the source of every moment worth remembering. I would not be here, nor would I have achieved this, without your support, patience and love. Words cannot describe, but since that is all I have:

thank you.

# Contents

**List of Figures ..... viii**

**List of Tables ..... xiii**

**1 INTRODUCTION ..... 1**

1.1 Chromatin structure and organisation ..... 1

1.2 ATP dependent chromatin remodeling ..... 6

1.3 Ubiquitin and ubiquitin-like SUMO modifications in DNA repair ..... 17

1.4 Aims of this thesis ..... 23

**2 MATERIALS AND METHODS ..... 25**

2.1 DNA Handling and Sub-cloning ..... 25

2.1.1 Oligonucleotide Primer Design ..... 26

2.1.2 DNA Amplification by Polymerase Chain Reaction (PCR) ..... 27

2.1.3 Ligation-independent Cloning ..... 28

2.1.4 Mutagenesis ..... 33

2.1.5 Plasmid DNA amplification ..... 36

2.1.6 Plasmids ..... 36

2.2 Protein Expression and Purification ..... 40

2.2.1 Bacterial expression of unlabelled proteins ..... 40

2.2.2 Bacterial expression of uniformly labelled proteins ..... 41

2.2.3 Baculovirus-mediated expression of unlabelled proteins in insect cells ..... 42

2.2.4 Protein expression in cultured mammalian cells ..... 43

2.2.5 Purification of bacterial His-tagged recombinant proteins ..... 45

2.2.6 Purification of bacterial untagged recombinant Ubiquitin ..... 46

2.2.7 Purification of KAP1 RBCC expressed in insect cells ..... 46
2.2.8 Purification of SMARCAD1 from mammalian cells for mass spectrometry analysis ..... 47
2.3 Solution State NMR ..... 49
2.3.1 Sample Preparation ..... 49
2.3.2 Data Collection Parameters ..... 49
2.3.3 Residual Dipolar Coupling Measurement ..... 50
2.3.4 NMR Titrations ..... 50
2.3.5 Structure Calculation ..... 51
2.4 Protein X-ray Crystallography ..... 57
2.5 Protein-Protein Interaction Characterisation ..... 58
2.5.1 Co-Immunoprecipitation of GFP-SMARCAD1 mutants upon DNA damage ..... 58
2.5.2 Denaturing pull-down of ubiquitinated proteins using IMAC ..... 60
2.5.3 Size Exclusion Chromatography / Multiple Angle Laser Light Scattering ..... 61
2.5.4 Isothermal Calorimetry ..... 61
2.5.5 Co-Immunoprecipitation of GFP-KAP1 truncation mutants ..... 62
2.5.6 Proximity ligation assay coupled immunolocalisation ..... 63

**3 STRUCTURAL ELUCIDATION OF SMARCAD1 BINDING DOMAINS ..... 67**

3.1 Sequence analysis and construct design ..... 68
3.2 Protein Expression and Purification ..... 73
3.3 Confirmation of SMARCAD1 SUMO interacting motif ..... 75
3.4 Backbone and sidechain sequence assignment ..... 78
3.5 Backbone dynamics ..... 88
3.6 Collection of structural restraints ..... 93
3.7 Structural calculations. ..... 100
3.8 Comparison of SMARCAD1 CUE1 with other CUE Domains ..... 104

**4 CHARACTERISATION OF SMARCAD1 PROTEIN INTERACTIONS MEDIATED BY THE SMARCAD1 SIM AND CUE1 DOMAINS ..... 106**

4.1 SUMO1 interaction with SMARCAD1 SIM ..... 107

4.2 Potential phosphorylation mediated regulation of SIM-SUMO1 interaction ..... 112
4.3 Mass spectrometry analysis of SMARCAD1 phosphorylation state upon chemically induced DNA damage ..... 124
4.4 SUMO1 interaction with a SMARCAD1 SIM phosphomimic ..... 127
4.5 Ubiquitin interaction with SMARCAD1 tandem CUE domain ..... 132
4.6 Mass spectrometry profiling of CUE1 and SIM mediated interactions ..... 137
4.7 Conclusions and future work ..... 153

**5 BIOCHEMICAL AND BIOPHYSICAL ANALYSIS OF SMARCAD1-KAP1 PROTEIN COMPLEX ..... 158**

5.1 Ubiquitylation state of KAP1 ..... 159
5.2. Structural consequences of SMARCAD1 CUE1 F169A mutation ..... 164
5.3 Identification of SMARCAD1 binding domain of KAP1 ..... 166
5.4 Recombinant protein expression and oligomeric state determination ..... 172
5.5 Binary protein-protein interaction between SMARCAD1 CUE1 and KAP1 RBCC domains ..... 178
5.6 Crystallisation of the protein complex between SMARCAD1 CUE1 and KAP1 RBCC ..... 184
5.7 Immunofluorescence analysis of SMARCAD1-KAP1 interaction in chromatin ..... 187
5.8 Conclusions and future work ..... 194

**6 DISCUSSION ..... 197**

6.1 SMARCAD1 CUE1 domain adopts a triple-helix bundle conformation ..... 197
6.2 Novel SMARCAD1 SIM is largely flexible and constitutively phosphorylated ..... 199
6.3 SMARCAD1 CUE1 domain is a poor mono-ubiquitin binder but mediates a high affinity interaction with KAP1 dimeric RBCC domain ..... 201
6.4 SMARCAD1 regulates p53 transcriptional repression ..... 204

**References ..... 211**

# List of Figures

Figure 1.1 Organisation hierarchy of DNA within the chromatin structure ..... 2
Figure 1.2 Histone tail post-translational modifications ..... 5
Figure 1.3 Classification of ATP-dependent Chromatin Remodelers ..... 7
Figure 1.4 Double-strand break repair pathways ..... 10
Figure 1.5 SMARCAD1 mediated re-establishment of post-replicative heterochromatin. ..... 13
Figure 1.6 SMARCAD1 promotes end resection after genotoxic insults in U2OS cells ..... 15
Figure 1.7 The nucleosome remodeler SMARCAD1 favours resection and 53BP1 eviction downstream of BRCA1–BARD1 ligase activity ..... 16
Figure 1.8 Ubiquitin and SUMO ..... 19
Figure 2.1 Domain architecture and recombinant construct design of human SMARCAD1 and human KAP1/Trim28/TIF1-β. ..... 26
Figure 2.2 Ligation-independent fragment assembly strategy for SMARCAD1 SIM substitution mutants design. ..... 35
Figure 3.1 Secondary structure and disorder prediction plot. Bioinformatics analysis of SMARCAD1 primary sequence to assess its propensity for disorder and defined secondary structure elements ..... 69
Figure 3.2 Sequence homology of SMARCAD1 CUE domains ..... 70
Figure 3.3 Sequence homology of SMARCAD1 SIM domains ..... 71
Figure 3.4 Design and initial expression screening for recombinant SMARCAD1 SIM/CUE1 constructs ..... 74
Figure 3.6 Confirmation of novel SIM in SMARCAD1 ..... 76
Figure 3.7 Identification of a novel SIM in SMARCAD1 ..... 77

Figure 3.8 Assignment and sequence connection of Cα and Cβ in SMARCAD1 SIM/CUE1 ..... 80
Figure 3.9 Assignment and sequence connection of Hα and Hβ in SMARCAD1 SIM/CUE1 ..... 82
Figure 3.10 Assignment and sequence connection of $^{13}C_{CO}$ in SMARCAD1 SIM/CUE1 ..... 84
Figure 3.11 Assignment of sidechain resonances beyond Cβ in SMARCAD1 SIM/CUE1 ..... 86
Figure 3.12 {$^{1}$H}-$^{15}$N Heteronuclear NOE of SMARCAD1 SIM/CUE1 ..... 89
Figure 3.13 Torsion angles predicted by TALOS+ for SMARCAD1 SIM/CUE1 domains ..... 91
Figure 3.14 HAHN coupling constant determination for SMARCAD1 SIM/CUE1 ..... 92
Figure 3.15 $^{1}$H,$^{15}$N HSQC spectra of SMARCAD1 constructs ..... 95
Figure 3.16 Long-range distance restraints in SMARCAD1 CUE1 domain ..... 96
Figure 3.17 Residual Dipolar Couplings for SMARCAD1 CUE1 domain from HSQC-IPAP Spectra ..... 98
Figure 3.18 Residual dipolar couplings for SMARCAD1 CUE1 domain ..... 99
Figure 3.19 SMARCAD1 CUE1 low-energy ensemble ..... 102
Figure 3.20 Overlay of SMARCAD1 CUE1 structure with alternative CUE domains ..... 105
Figure 4.1 NMR titration experiment of $^{15}$N eCUE1 with unlabelled SUMO1 ..... 109
Figure 4.2 Chemical shift perturbations for SMARCAD1 SIM/CUE ($^{15}$N eCUE1) bound to a 50-fold excess of SUMO1 ..... 110
Figure 4.3 Relative chemical shift changes from SUMO1 and SMARCAD1 SIM/CUE1 interaction ..... 112
Figure 4.4 NMR titration experiment of $^{15}$N SUMO1 with unlabelled SMARCAD1 SIM peptide without modified phosphorylation sites ..... 114
Figure 4.5 NMR titration experiment of $^{15}$N SUMO1 with unlabelled SMARCAD1 SIM peptide phosphorylated on Ser124 ..... 115
Figure 4.6 NMR titration experiment of $^{15}$N SUMO1 with unlabelled SMARCAD1 SIM peptide phosphorylated on Ser127 ..... 116
Figure 4.7 NMR titration experiment of $^{15}$N SUMO1 with unlabelled SMARCAD1 SIM peptide phosphorylated on Ser132 ..... 117

Figure 4.8 Chemical shift perturbations for $^{15}$N SUMO1 bound to a 10-fold excess of SMARCAD1 SIM differentially phosphorylated peptides ..... 119
Figure 4.9 Relative chemical shift changes from SUMO1 and unmodified SMARCAD1 SIM peptide interaction ..... 122
Figure 4.10 Relative chemical shift changes from SUMO1 and pSer124 SMARCAD1 SIM peptide interaction ..... 122
Figure 4.11 Relative chemical shift changes from SUMO1 and pSer127 SMARCAD1 SIM peptide interaction. ..... 123
Figure 4.12 Relative chemical shift changes from SUMO1 and pSer132 SMARCAD1 SIM peptide interaction ..... 123
Figure 4.13 Pull down purification of transiently expressed recombinant GFP-SMARCAD1 from HEK 293T cells alternatively exposed to chemically induced DNA damage ..... 125
Figure 4.14 Phosphorylation profiling of SMARCAD1 differentially exposed to chemically induced DNA damage ..... 126
Figure 4.15 NMR titration experiment of $^{15}$N eCUE1/$S_3$E with unlabelled SUMO1 ..... 128
Figure 4.16 Chemical shift perturbations for SMARCAD1 SIM/CUE1 $S_3$E phosphomimic ($^{15}$N eCUE1/$S_3$E) bound to a 50-fold excess of SUMO1 ..... 130
Figure 4.17 Relative peak height intensity changes from SUMO1 and SMARCAD1 SIM/CUE1 $S_3$E phosphomimic interaction ..... 132
Figure 4.18 NMR titration experiment of $^{15}$N edCUE with unlabelled Ubiquitin ..... 133
Figure 4.19 Chemical shift perturbations for SMARCAD1 SIM/CUE1/CUE2 ($^{15}$N edCUE) bound to a 50-fold excess of Ubiquitin ..... 134
Figure 4.20 Normalised chemical shift changes from Ubiquitin and SMARCAD1 CUE1 interaction ..... 136
Figure 4.21 Normalised chemical shift changes from Ubiquitin and SMARCAD1 CUE2 interaction ..... 136
Figure 4.22 SMARCAD1 SIM and CUE1 binding domain mutants Schematic representation of functionally abrogating mutant constructs for the SIM (SIM8A) and CUE1 (F169A) from SMARCAD1 ..... 139

Figure 4.23 Volcano plot of differential abundance proteins from a co-immunoprecipitated pair of wild-type GFP-SMARCAD1 samples alternatively treated with NCS ..... 148
Figure 4.24 Volcano plots of differentially abundant proteins from a co-immunoprecipitated set of wild type and SIM mutant GFP-SMARCAD1 samples alternatively treated with NCS ..... 149
Figure 4.25 Volcano plots of differentially abundant proteins from a co-immunoprecipitated set of wild type and CUE1 mutant GFP-SMARCAD1 samples alternatively treated with NCS ..... 152
Figure 5.1 Post-translational modifications profiling of KAP1 differentially exposed to chemically induced DNA damage and co-immunoprecipitated with GFP-SMARCAD1 ..... 160
Figure 5.2 Cobalt TALON® affinity purification of Nt $MAH_6G$/Ubiquitin modified proteins from HEK 293T cells under denaturing conditions ..... 163
Figure 5.3 $^{1}H,^{15}N$ HSQC analysis of SMARCAD1 SIM/CUE1 F169A ..... 165
Figure 5.4 Identification of SMARCAD1 binding domain in KAP1 ..... 168
Figure 5.5 Identification of SMARCAD1 binding subdomain in KAP1 RBCC ..... 171
Figure 5.6 Heterologous expression of KAP1 RBCC fusion proteins in baculovirus infected Sf9 insect cells ..... 173
Figure 5.7 Strategies for KAP1 RBCC N-terminal hexahistidine fusion protein overexpressed in Sf9 insect cells ..... 174
Figure 5.8 Multiple angle laser light scattering analysis of KAP1 RBCC NtHis. Chromatograms from the differential refractive index detector are shown at five protein concentrations ..... 176
Figure 5.9 Isothermal titration calorimetry (ITC) analysis for binary binding between KAP1 RBCC NtHis and SMARCAD1 CUE1 CtHis ..... 179
Figure 5.10 Crystallisation of SMARCAD1 CUE1 and KAP1 RBCC protein complex. Images of Nanoliter crystallisation experiments using sitting drops and the vapour diffusion method ..... 185
Figure 5.11 Data collection of initial crystal hits for the SMARCAD1 CUE1 and KAP1 RBCC protein complex ..... 186

Figure 5.12 Immunofluorescence microscopy co-localisation of SMARCAD1 and Kap1 on fixed and permeabilized HeLa cells ..... 188
Figure 5.13 Proximity Ligation Assay of the SMARCAD1-KAP1 complex..... 190
Figure 5.14 HeLa cells analysed by immunofluorescence co-localisation of PLA®-detected KAP1-SMARCAD1 complex with γH2A.X and H3K9me3 ..... 192
Figure 5.15 U2OS cells analysed by immunofluorescence co-localisation of PLA®-detected KAP1-SMARCAD1 complex with γH2A.X and H3K9me3 ..... 193
Figure 6.1 SMARCAD1 CUE1 low-energy ensemble ..... 198
Figure 6.2 Comparison between SMARCAD1 CUE1 domain solution structure models..... 199
Figure 6.3 Solution structure of KAP1 homologue TRIM25..... 203
Figure 6.4 Alternative pathways involved in SMARCAD1 response to chemically induced DNA damage ..... 205
Figure 6.5 SMARCAD1 bridges the concerted eviction of KAP1 and 53BP1 downstream of the ubiquitin-ligase activity of RNF4 and BRCA1-BARD1. ..... 208
Figure 6.6 SMARCAD1 nuclear subpopulations enforce autonomous functions associated with KAP1 interaction and 53BP1 chromatin eviction. ..... 210

# List of Tables

Table 2.1 Thermo cycling conditions used for PCR ..... 27
Table 2.2 Oligonucleotides used as for PCR construct design throughout this work .... 29
Table 2.3 Vectors used as linearized targets for InFusion™ cloning ..... 33
Table 2.4 Thermo cycling conditions used for QuickChange II SDM ..... 34
Table 2.5 Oligonucleotides used for mutagenesis ..... 36
Table 2.6 Description of plasmids used for protein expression ..... 37
Table 2.7 M9 minimal media components for uniformly labelled bacterial expression ..... 41
Table 2.8 Acquisition Parameters for Triple Resonance NMR Experiments used in Backbone Assignment ..... 53
Table 2.9 Acquisition Parameters for $^{13}$C-edited NMR Experiments ..... 53
Table 2.10 Acquisition Parameters for $^{15}$N-edited NMR Experiments ..... 54
Table 2.11 Protein/Ligand pairings of NMR titration experiments ..... 55
Table 2.12 Antibodies used for immunofluorescence and PLA analysis ..... 66
Table 3.1 Extent of backbone and sidechain assignment for SMARCAD1 SIM/CUE1 CtHis construct spanning amino acids 109-206. ..... 87
Table 3.2 NMR Structure refinement statistics for SMARCAD1 CUE1 low-energy ensemble ..... 103
Table 4.1 Dissociation constants for SUMO1 interaction with synthetic SAMARCAD1 SIM peptides, either unmodified or mono-phosphorylated at S124, S127 or S132. ... 124
Table 4.2 Differentially enriched SMARCAD1 binding partners, identified through label free quantitation mass spectrometry with $p \geq 0.05$ ..... 142

Chapter 1

# Introduction

## 1.1 Chromatin structure and organisation

Eukaryotic cells, ranging from yeasts to humans, package their genomic DNA together with an equal amount of protein to form chromatin, a highly complex molecular ensemble which also integrates additional indirectly associated protein and RNA molecules including DNA-binding factors, the basal transcription machinery with its nascent transcripts, and multi-protein complexes that copy and maintain DNA during replication and repair [1, 2]. All components work in concert, under strict regulation by overlapping and redundant signalling pathways that incorporate endogenous and external stimuli allowing the cell to adapt and survive.

The repeating unit of chromatin, at the centre of this overarching structure, is the nucleosome [3]. Formed by a segment ~147bp of dsDNA organized around a histone octamer (consisting of two copies of each highly conserved H2A, H2B, H3, and H4 proteins), they are connected by short stretches of linker DNA into nucleosomal arrays, to resemble “beads on a string”. This is the first level of DNA compaction which allows a single human cell to fit approximately 1.8m of DNA in length into the cell nucleus, less than 10μm in diameter, while remaining accessible for essential cellular processes [4].

Linear arrays of nucleosomes are folded and compacted into three-dimensional (3D) higher-order heteromorphic structures thought to consist of 30nm fibres in loose zigzag or solenoid conformations intermingling to form a “polymer melt” [2, 5]. This plastic organization is eventually compacted further, during the cell cycle transition

between interphase and metaphase, into condensed chromosomes. A schematic representation of the hierarchy of chromatin compaction is illustrated in Figure 1.1

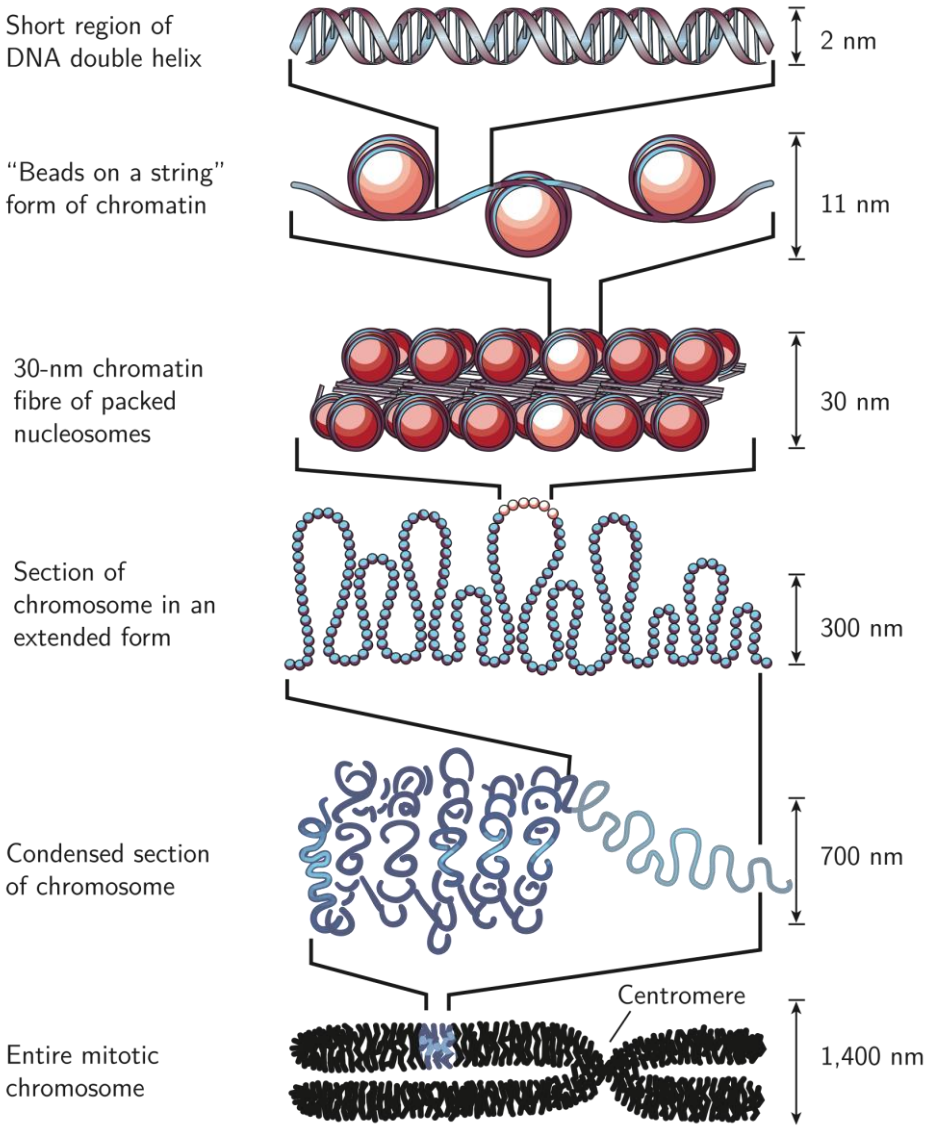



**Figure 1.1 Organisation hierarchy of DNA within the chromatin structure**. Two superhelical turns of DNA are wound around core histone octamers to produce nucleosomes. These, in turn, are connected by short stretches of histone H1 bound linker DNAin a string of nucleosomes, which further folded into 30nm fibre structures. Higher-order compaction eventually forms chromosomes, though molecular understanding of structures beyond the nucleosome remains uncertain. Adapted by permission from Macmillan Publishers Ltd: Nature [6], copyright 2003.

Though easily the best recognisable form of chromatin structures are metaphase chromosomes at the endpoint of chromatin compaction in dividing cells, outside of mitosis, chromatin exists as a dense heterogeneous distribution of regions with variable degrees of compaction [2, 5]. Traditionally, these are roughly categorised as gene-rich, transcriptionally active euchromatin and transcriptionally repressed heterochromatin, further classified as constitutive if enriched in repetitive, gene-poor, late replicating DNA sequences as present in centromeres and telomeres; or facultative if consisting of genomic regions differentially expressed under strict temporal and spatial contexts through development but then silenced [7-9]. Despite the fact that this model has now been proved to be a clear over simplification of the chromatin organisation in the nucleus (increasing evidence suggests a continuum of large-scale chromatin compaction, with less than two-fold difference in accessibility between euchromatin and repressive heterochromatin) [1, 10], it remains a useful distinction to cluster chromatin domains depending on whether they contain active or inactive genes.

Chromatin is increasingly recognised as a dynamic structure and not simply a passive scaffold, required only to package the entire eukaryotic genome into the confines of the nucleus. It is in fact an active regulator of DNA accessibility for gene transcription, replication and repair [11, 12]. Among the mechanisms employed by the cell to assure chromatin plasticity are histone exchange and post-translational modifications (PTMs). In the former, nucleosome composition is altered in a DNA replication independent manner through specific histone chaperones and ATP-dependent chromatin remodeling factors [2, 12, 13]. Alternative histone variants, encoded by genes that are distinct from those encoding their major-type counterparts, are expressed and incorporated into nucleosomes throughout the cell cycle. Most histone variants have been identified for histones H2 and H3, they are specifically recruited to precise genomic locations and their substitution alters the biochemical properties of the nucleosome by affecting PTMs, protein interactions or higher-order chromatin structure [2, 12, 14, 15]. Histone variants are highly conserved between different species, indicating that they have evolved to fulfil important functions that cannot be

accomplished by their canonical counterparts, as has been demonstrated for H2A.X, H2A.Z, and CenH3/CENPA [16-18].

The second mechanism, histone tail post-translational modification, is exploited to modify and regulate interactions between neighbouring nucleosomes as well as the accessibility to the genome by nuclear factors [2, 19, 20]. The function of PTMs in regulating nearby transcription or repair is attributed to their ability to act as scaffolds to which effector proteins may bind. Conversely, PTMs may also prevent proteins from binding to chromatin through steric hindrance or electrostatic repulsion. Depending on their effect, most marks are classified as activating or repressing. Histone tail modifications are for the most part reversible, which allows for temporal and spatial regulation of these marks, and the misregulation of which is at the root of many genomic diseases, including cancer [21-24].

Histones contain many amino acids that may be post-translationally modified simultaneously, with virtually all known types of protein modifications found on histones (including acetylation, phosphorylation, methylation, ubiquitylation and SUMOylation), and new types of modifications as well as new locations on histone proteins are discovered at a rapid pace, although in many instances their biological significance awaits confirmation [12, 19, 22] (Figure 1.2). The current histone code model incorporates the extraordinary diversity of epigenetic chemical marks and proposes the multivalent binding of combinatorial modification resulting in distinct downstream events, different from what individual modifications or an alternate pattern might invoke [11, 25]. One important example, relevant to chromatin compaction, is the different modifications decorating euchromatin (hyperacetylation of histone tails and trimethylation of Lys4 on Histone H3 – H3K4me3) and heterochromatin (histone hypoacetylation, trimethylation of Lys27 on Histone H3 – H3K27me3, and di-/trimethylation of Lys9 on Histone H3 – H3K9me2/H3K9me3), responsible for the functional distinction between the two [11, 26].

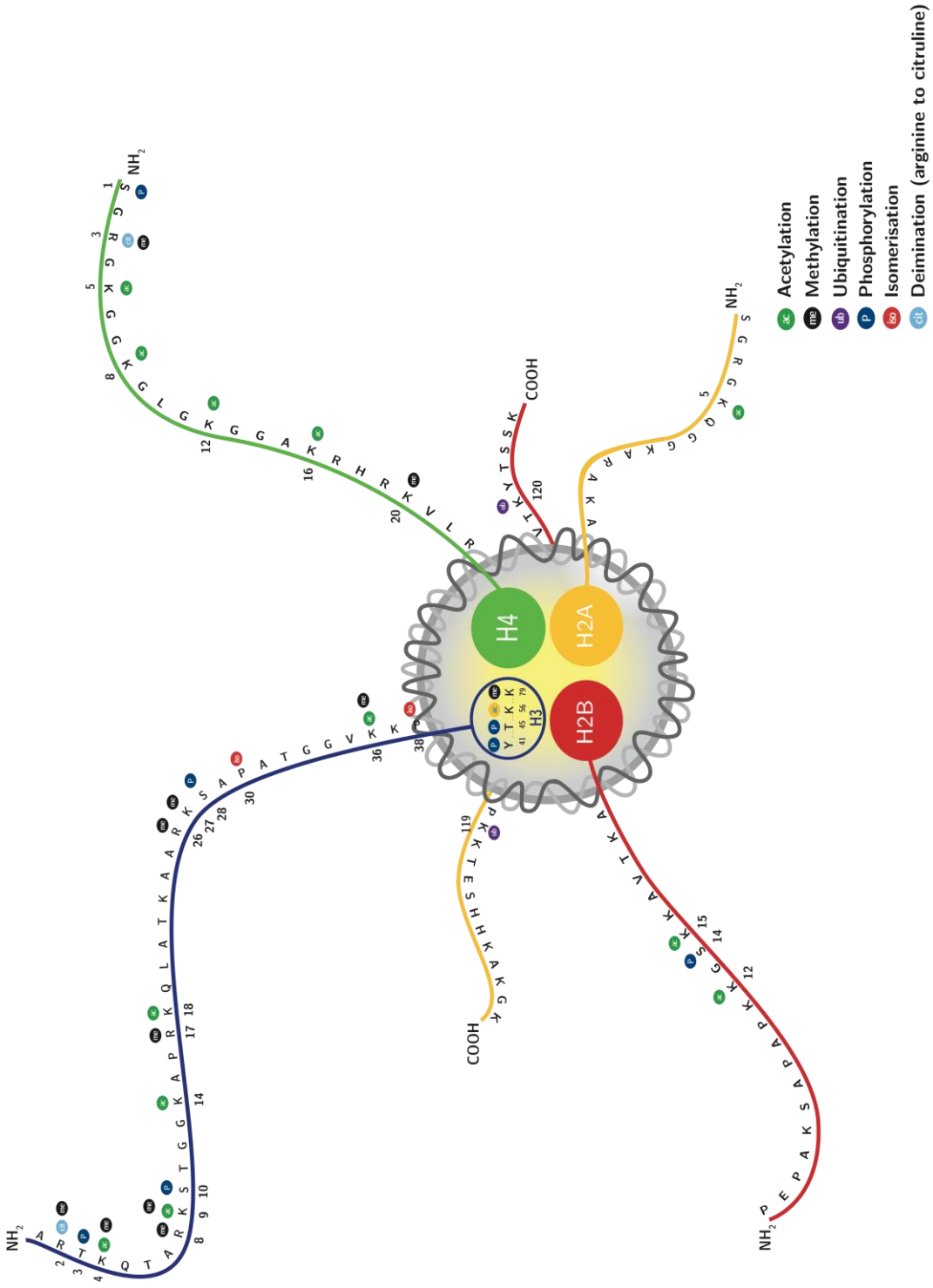



**Figure 1.2 Histone tail post-translational modifications.** Schematic diagram of a nucleosome with protruding histone tails and sites of various known modifications. Histone marks are critical to dynamic modulation of chromatin structure and function, contributing to the cellular gene expression programme. Adapted from Huang et al 2014 [27].

## 1.2 ATP dependent chromatin remodeling

As mentioned above, chromatin itself plays a regulatory role on crucial cellular processes (such as replication, transcription and repair) by physically limiting access of DNA binding effectors to target sequences. Moreover, epigenetic markings including histone PTMs and GpC methylation patterns are central to appropriate recruitment within a cellular spatial-temporal context to ensure the required downstream events for cell survival [28]. These processes are necessarily coupled to extensive remodeling of the genome landscape to allow access and must be followed by the re-establishment of the proper chromatin organisation once completed. To this end, a number of histone modifiers and ATP-dependent chromatin remodelers, usually associated with additional protein partners in multi-subunit complexes, will be recruited to induce local chromatin rearrangement through the recognition of structural features such as nucleosome architecture, DNA sequences and most importantly PTMs both on histones and interacting bound factors [29, 30].

Chromatin remodelers, at the centre of these multi-subunit complexes, confer conformational rearrangements to the target nucleosome by means of chemical energy consumption. In general these remodelers, members of the Swi/Snf2 family of proteins and multiprotein assemblies that disrupt or remodel protein-nucleic acid complexes (most notably nucleosomes, but also transcription factor and RNA Polymerase) are classified based on their ATPase domain sequence conservation and domain architecture [31, 32]. Along with two RecA-like helicase domains a diverse combination of DNA, nucleosome and PTM binding ancillary domains are part of these proteins though each subfamily cluster does exhibit a certain predisposition for a set combination of said domains, as illustrated in Figure 1.3.

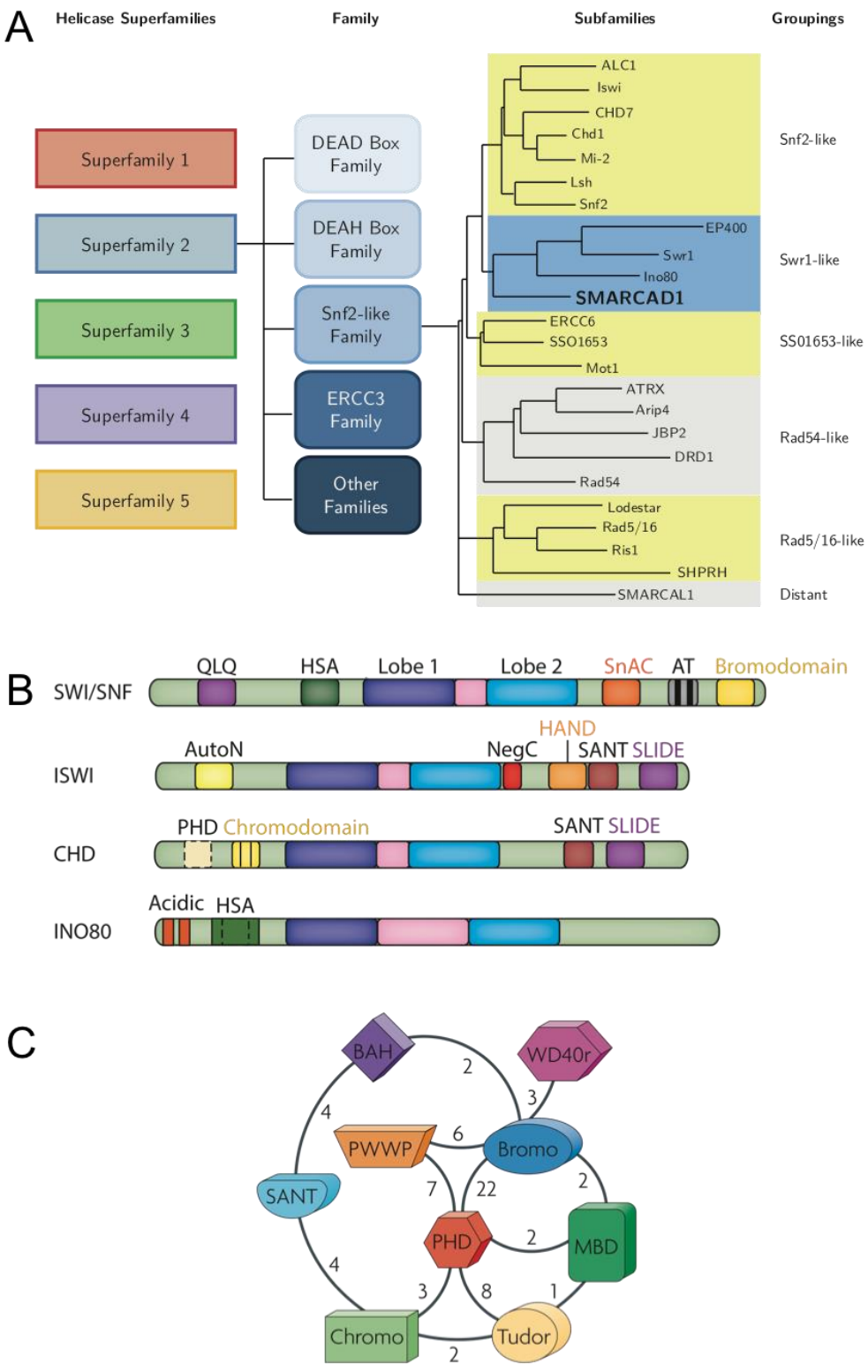



**Figure 1.3 Classification of ATP-dependent Chromatin Remodelers (Legend on next page).**

**Figure 1.3 (Previous page) Classification of ATP-dependent Chromatin Remodelers.** A) ATP-dependent chromatin remodelers can be categorised according to the sequence of their ATPase domain. Swi2/Snf2- like family helicases, are distinct from other SF2 members in that they contain additional linker sequence between two RecA domains. Based on phylogenetic analysis, the Snf-2 family has been further split into 24 subfamilies (or 6 groupings partially correlated with known function). Mammalian SMARCAD1 and its S. cerevisiae homologue FUN30 (Function UNknown 30) belong to the SMARCAD1/Etl1 subfamily, which is in the Swr1-like grouping, shown in blue. B) Domain organisation of ATPase subunit of selected remodelers, C) Schematic depiction of combined effector module domains within a remodeler, with the number of known instances depicted along the linker connecting two domains. BAH: bromo-adjacent homology domain, Bromo: bromodomain, Chromo: chromodomain, MBD: methyl-CpG binding domain, PHD: plant homeodomain, PWWP: PWWP motif protein of the Royal superfamily, Tudor: Tudor binding domain, SANT: Switching-Defective Protein 3 (Swi3), Adaptor 2 (Ada2), nuclear receptor co-repressor (N-CoR) and transcription factor (TF)IIIB, WD40r: WD40 repeat. A and B adapted from Clapier and Cairn 2009, and Bartholomew 2014 [11, 29]; C adapted by permission from Macmillan Publishers Ltd: Nature [31], copyright 2007.

Once bound, remodelers are able to reshape the chromatin landscape by modifying nucleosome distribution in the immediate vicinity, alternatively by sliding along the DNA, ejecting nucleosomes altogether or exchanging histone variants which could result in indirect chromatin restructuring, as mentioned before [32, 33]. Furthermore as part of multi-subunit complexes chromatin remodelers can rearrange chromatin, with long range effects, by altering the PTM profile of neighbouring nucleosomes [29, 34].

Through means of these remodeling proteins, processes such as replication and repair may enjoy access to substantial stretches of naked or near-naked dsDNA. It would be misleading, however, to assume a static nature of these events. On the contrary, chromatin reorganization is a dynamic effort throughout replication or DNA lesion repair, and in fact many of the same players are involved in both chromatin

modulation processes, with the distinct difference that faithful restoration of the original chromatin structure is required during repair to avoid unscheduled epigenetic changes [15, 35].

Though replication is an incredible complex sequence of events, highly coordinated and strictly regulated, DNA repair presents particular challenges in the sense that it is a generalised term for a large number of specific cellular mechanisms to correct a variety of DNA lesions, from base adducts to double-strand breaks (DSBs) caused endogenously (through replication fork collapse, metabolic reactive oxygen species or programmed V(D)J and classical switch recombination) or through exogenous agents (UV light, ionizing radiation, and intercalating agents) [26, 36, 37]. Although DSBs arise less frequently than DNA single strand breaks and many base alterations, they represent the greatest genotoxic insult as they pose the risk of extensive chromosomal rearrangements if improperly repaired [38, 39].

There are two repair mechanisms available to resolve DSBs: error-prone non-homologous end joining (NHEJ) and error-free homologous recombination (HR). The former is available to cells throughout the cell cycle and involves the fusion of DNA broken ends through direct ligation. It is classified as canonical or alternative according to the degree of DNA resection (that is, the eviction of nucleosomes immediately flanking the DSB and nuclease degradation of the 5'-end to create a ssDNA strand) accompanying the repair cascade (Figure 1.4). On the other hand, HR makes use of a sister chromatid as a template to reproduce the original genome sequence, though consequently it is necessarily confined to the S/G2 phases of the cell cycle. It is characterised by extensive resection and remodeling due to recombination with the sister chromatid [37, 40, 41].

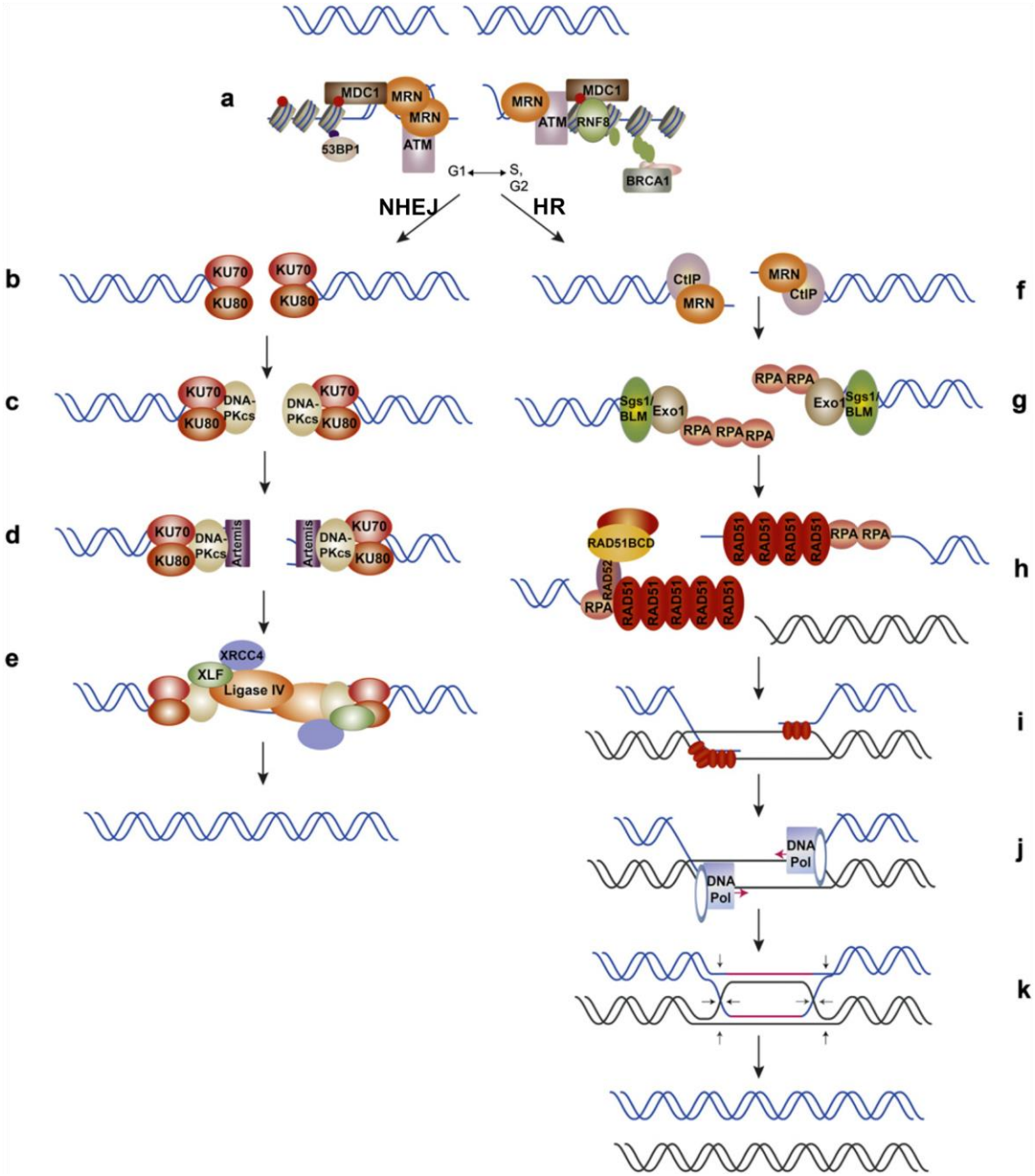



**Figure 1.4 Double-strand break repair pathways (Legend on next page).**

**Figure 1.4 (Previous page) Double-strand break repair pathways.** (a) Double-strand DNA breaks are recognized by the MRN complex, which catalyses the activation of ATM. The choice between the HR repair and NHEJ is established at the step of DNA end resection–its initiation prevents NHEJ and promotes HR repair. This is dependent on the cell cycle phase, DNA end modifications and end binding complexes. (b-e) Following limited end resection by the MRN complex, the Ku70/Ku80 and DNA-PKcs proteins initiate NHEJ by recruiting the nucleases Artemis and APLF to process broken ends, which are then joined by Ligase IV. In HR repair (f-k), broken ends are first processed to long single-stranded DNA (ssDNA) tails covered by RPA through a multi-stage process that involves MRN complex, CtIP and EXO1 and/or BLM and DNA2. RPA covering the ssDNA overhang is later replaced by Rad51, which in turn facilitates sister chromatid invasion, template DNA extension and Holliday Junction resolution trough recombination, restoring the original genomic sequence without loss of information. Adapted by permission of Elsevier Publishing Group from Gospodinov and Herceg, 2013 [40].

An interesting corollary of chromatin remodeling in cellular processes such as replication and repair, is their execution within a transcriptionally repressive, highly condense heterochromatin context. Again, repair of DNA lesions, and specifically DSBs, pose particular challenges in heterochromatic regions. The extensive degree of compaction influences the accessibility of effector proteins, though it has been demonstrated that at least damage sensor are still capable of detecting lesion with equivalent efficiency to euchromatin [26, 38].

However, damaged heterochromatin does require relaxation and remodeling for repair [26, 38]. At DSBs a seeming sequence of events, following initial lesion recognition and ATM activation has been proposed, with heterochromatin binding factors (HP1) being recruited to DSBs (possibly to stabilise flanking chromatin regions and reduce the risk of chromosomal rearrangements), which also enriches HP1-binding Krüppel-associated box (KRAB)-associated co-repressor (KAP1), an ATM target at Ser824 resulting in disruption of CHD1 SIM-KAP1 SUMO interaction and local

chromatin relaxation [42-44]. Furthermore, HP1 is phosphorylated by ATM-activated pChk2 (cell cycle Checkpoint Kinase 2) at Ser473 and irradiation induced CK2 at Thr51 resulting in disruption of KAP1 and H3K9me3 associations, respectively [26, 42, 45], and eviction from DSB flanking chromatin. Histone acetyl transferase Tip600 is then able to bind H3K9me3 and, in turn, hyper-acetylate histone H4, which further enhances chromatin relaxation. Tip600 also acetylates histone H2A.X pSer139 (γH2A.X), which is a pre-requisite for its subsequent ubiquitylation and eviction to restore the chromatin landscape once repair is fulfilled [26, 46].

Local relaxation around DSBs is coupled to the use of flanking chromatin regions as scaffolding platforms to concentrate and regulate recruitment of DDR effector proteins, through PTMs both on histones and previously bound factors, including MCM1, 53BP1, RNF8, RFN168, BRAC1 and RAD51 [34, 39, 47]. An additional consideration should be made regarding the influence that the intrinsic repressive structure of heterochromatin might impose on DSB repair pathway choice [48]. Not only is constitutive heterochromatin rich in repetitive sequences, which can mistakenly be used as homologue templates for HR, but the physical requirement for extensive relaxation prior to HR and the rapid binding of abundant Ku70/Ku80 factors to broken DNA ends might bias the propensity of DSB repair to NHEJ, which can be selectively activated even before relaxation of heterochromatin through ATM/KAP1/HP1 signalling occurs. At any rate, the limited need for remodeling throughout ligation mediated repair as established by canonical NHEJ is likely to be beneficial for maintaining genomic stability [38, 49].

On the other hand, replication of heterochromatic regions is accompanied by the specific re-establishment of its epigenetic landscape. Histone eviction and deposition around the progression replication fork requires a careful and concerted effort to recognise recycled histones (bearing the original repressive marks, including H3K9me2/3 and H3K27me3) and recruit a heterochromatin specific deacetylase/methyl-transferase complex, which would in turn process newly deposited histones (constitutively acetylated). This complex combines heterochromatin factors such as HP1 and KAP1, along with KAP1-bound chromatin

remodeler SMARCAD1 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1), which specifically interacts with PCNA [50, 51], as illustrated in Figure 1.5. In this manner, a core ATPase remodeling complex brings together histone deacetylases HDAC1/2 and histone H3K9 methyl-transferase G9a/GLP to the site of newly deposited histones and propagates the repressive epigenome.

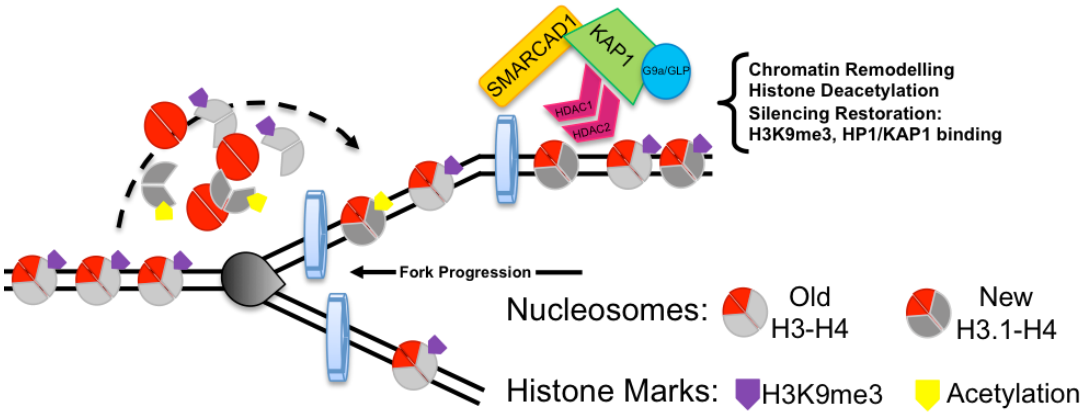



**Figure 1.5 SMARCAD1 mediated re-establishment of post-replicative heterochromatin.** Newly deposited acetylated nucleosomes are silenced by the SMARCAD1/KAP1 complex recruited to replication sites by PCNA and primed by recycled H3K9me3 methylated nucleosomes. Associated factors bound to SMARCAD1 or KAP1, HDAC1/2 and G9a/GLP, alter histone post-translational modifications. KAP1-associated HP1 is then recruited to modified nucleosomes, helping maintain heterochromatic packaging. Adapted from Mermoud et al 2011 [50].

Given the crucial role of the binary core complex between heterochromatin bound master regulator KAP1 (also known as Tripartite Motif TRIM28, and traditionally as Transcriptional Intermediary Factor TIF1β) and its bona fide partner SMARCAD1, it is particularly pertinent to consider if any of the functions associated with KAP1 throughout heterochromatic DNA repair could involve SMARCAD1 mediated remodeling.

SMARCAD1 (also known as Etl-1, HEL-1 or KIAA1122 [52-54]) is a non-canonical ATPase remodeler from the Swr1-like subfamily, unique in its domain organization by

encoding two CUE domains (mono-ubiquitin binding domains) along with a classical SNF2 helicase ATP-dependent motor. In humans, two SMARCAD1 isoforms have been reported, which differ in length and sites of transcription initiation [55]. The larger of the two corresponds to the full-length protein discussed thus far and has been shown to be expressed ubiquitously at low levels. On the other hand, a shorter isoform, which omits the initial 430 residues bearing the tandem CUE domains, is mainly expressed in skin fibroblasts, keratinocytes and the oesophagus. Interestingly, this latter isoform has been linked to adermatoglyphia (the absence of ridges on the skin on the pads of the fingers and toes, as well as on the palms of the hands and soles of the feet) and Basan syndrome, originating from the disruption of a single conserved donor splice site by splice variants resulting in reduced mRNA stability and suggesting a tissue specific role during dermatoglyph ontogenesis [55-58].

Full-length SMARCAD1 is well conserved from yeast to humans, and many of its reported potential roles in DNA repair are functional extrapolations from the well-characterised *S. cerevisiae* ortholog FUN30 [59, 60]. Two recent reports, however, do specifically link SMARCAD1 with DSB resection in HR. First, Costello *et al* [61] show that SMARCAD1 exhibits a similar phenotype to EXO1 when depleted by siRNA knockdown in cultured human U2OS cells (Figure 1.6). In addition, immunofluorescence staining shows that SMARCAD1 co-localises with γH2A.X at laser-induced DNA damage sites and at DSBs generated by FokI nuclease. Finally, GFP-SMARCAD1 was recruited to laser micro-irradiation-induced lesions before GFP-tagged RPA and with kinetics similar to that of GFP-tagged EXO1. These results indicate that chromatin remodelling driven by SMARCAD1 facilitates long-range resection (either directly or indirectly) along with recombinational DSB repair.

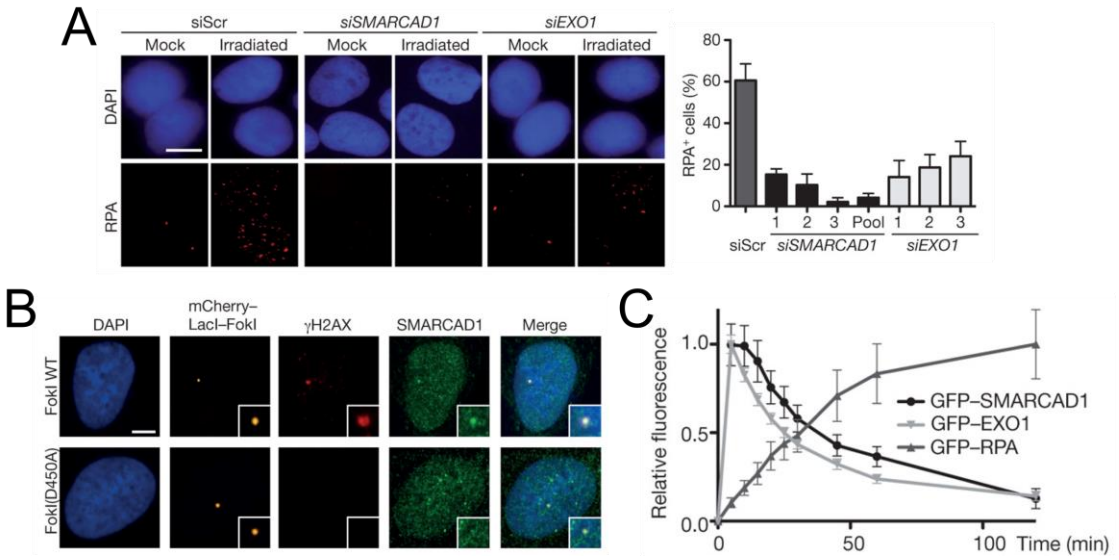



**Figure 1.6 SMARCAD1 promotes end resection after genotoxic insults in U2OS cells.** A) Immunofluorescence analysis of RPA foci after ionising radiation treatment. Quantification of RPA foci shows a similar effect of siRNA-mediated knockdown of EXO1 dsDNA nuclease and chromatin remodeler SMARCAD1 B) Immunofluorescence staining of SMARCAD1 and γH2A.X at DSBs. The induced FokI endonuclease creates a single DSB, simplifying the analysis of co-localisation studies. From this it can be inferred that SMARCAD1 is recruited to γH2A.X foci, C) Quantification of GFP-SMARCAD1, GFP-EXO1 and GFP-RPA accumulation at sites of laser micro-irradiation in live cells. Adapted by permission from Macmillan Publishers Ltd: Nature [61], copyright 2012.

Additionally, very recently, Densham *et al* [62] have proposed a model in which SMARCAD1 is recruited to DSBs downstream of BRCA1-BARD1 C-terminal ubiquitylation of histone H2A and regulates 53BP1 eviction to ensure complete resection of DSBs during HR. BRCA1 (Breast Cancer Susceptibility Gene 1) and BARD1 (BRCA1-associated RING-domain Protein) form a heterodimer that functions as an E3 ubiquitin-ligase in the transfer from E2 ubiquitin-conjugating enzymes to target proteins [63]. Inheritance of mutations in BRCA1 is associated with high risk of breast and ovarian cancer, with tumours presenting excessive genome instability. BRCA1 has been implicated in several aspects of genome stability including favouring DSB repair through homologous recombination by opposing the block on DNA resection established by the p53-binding protein 53BP1 and its effector proteins [64, 65]. The

authors demonstrate that the tandem CUE domains (dCUE) in SMARCAD1 selectively recognize ubiquitylated H2A (H2A-Ub) and that these domains are also responsible for SMARCAD1 accumulation on γH2A.X decorated laser induced DNA damage sites (Figure 1.7). BARD1 depletion by siRNA knockdown displayed a similar phenotype on SMARCAD1 recruitment to that observed upon abrogating mutagenesis of dCUE. Moreover, repositioning of 53BP1 (quantified through immunofluorescence co-localisation on ionising radiation induced foci, IRIF) failed on SMARCAD1 depleted cells at S/G2 phase, and was selectively rescued by WT SMARCAD1 but not abrogating mutants of dCUE or the ATPase domain, indicating that these domains in concert play a role in promoting 53BP1 eviction through remodeling of damage sites.

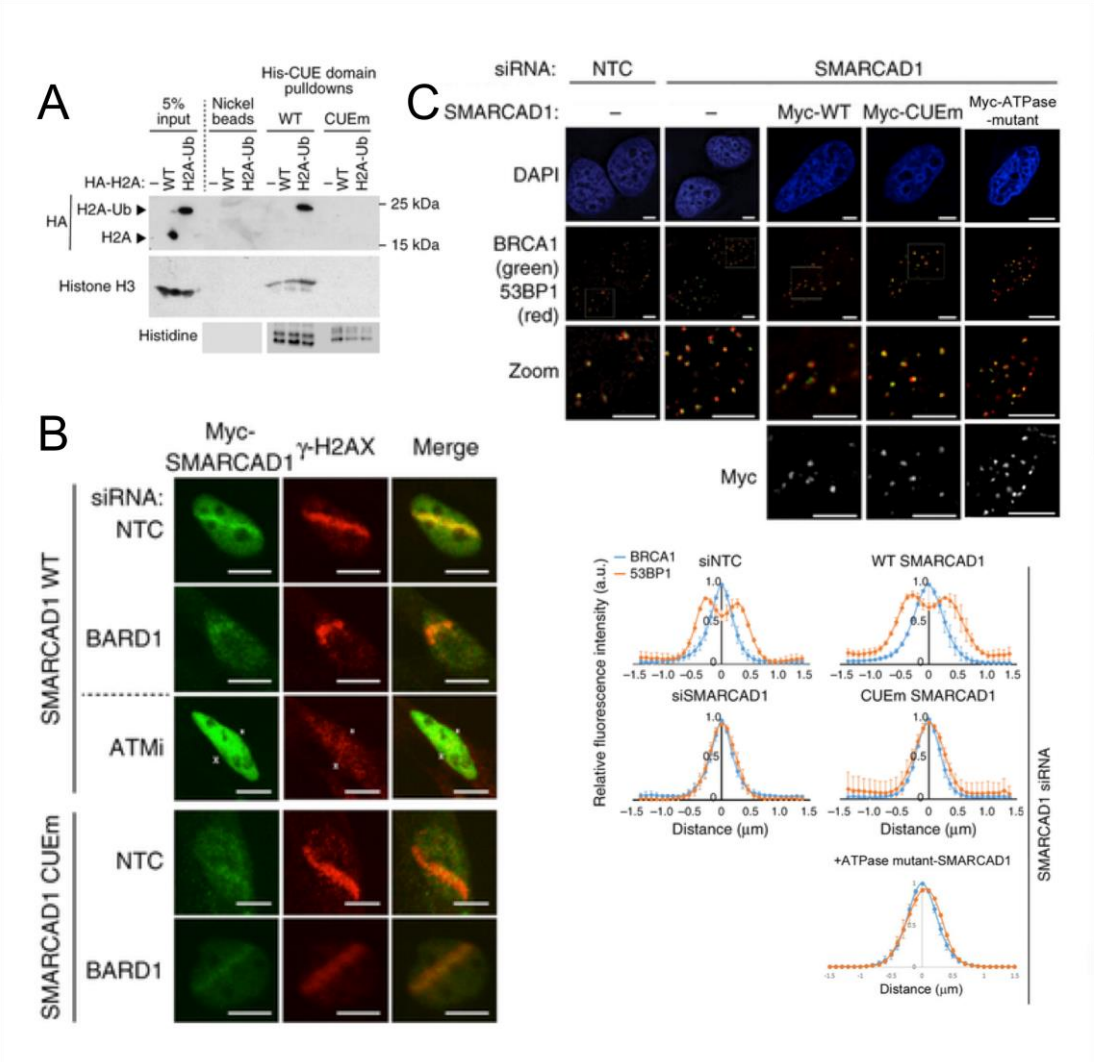



**Figure 1.7 The nucleosome remodeler SMARCAD1 favours resection and 53BP1 eviction downstream of BRCA1–BARD1 ligase activity (Legend on next page).**

**Figure 1.7 (previous page) The nucleosome remodeler SMARCAD1 favours resection and 53BP1 eviction downstream of BRCA1–BARD1 ligase activity.** A) Western blot analysis of H2A NtHA or H2A-Ub NtHA stably expressed in HeLa cells and pulled down with purified WT or mutant SMARCAD1 dCUE (aa 98–318, CUEm) bound to Nickel-affinity beads. B) Representative images of γH2A.X colocalisation in cells expressing siRNA-resistant SMARCAD1 variants treated with SMARCAD1 siRNA or co-depleted with BARD1 siRNA and bearing laser-line-induced DNA damage. ATM inhibitor (10 µM KU55933, 4h prior to damage) was used as negative control. C) Representative images of cells transfected with siRNA targeting SMARCAD1 in cells stably expressing siRNA-resistant WT, dCUE mutant (CUEm) or ATPase mutant (K528R) as myc-SMARCAD1 fusion proteins. Cells were stained for BRCA1 and 53BP1 after ionising radiation. Adapted by permission from Macmillan Publishers Ltd: Nature Structural and Molecular Biology [62], copyright 2016.

Finally, SMARCAD1 has been identified as a potential target for ATM/ATR phosphorylation at canonical (Ser/Thr-Gln) sites, as well as at non-canonical sites, in response to genotoxic insults [66].

## 1.3 Ubiquitin and ubiquitin-like SUMO modifications in DNA repair

Among the varied post-translational modifications involved in DNA repair, post-translational incorporation of ubiquitin and ubiquitin-like SUMO have emerged, along with phosphorylation, as crucial regulators of almost every aspect of cellular responses to DNA damage. Recent proteomic studies have identified hundreds of damage-induced targets of both ubiquitylation and SUMOylation [67-69] with the complexities of combinatorial effects of these moieties exacerbated by evidence suggesting that it is the bulk SUMOylation of a group of proteins rather than any one specific target that stimulates DSB repair [70]. Despite these challenges, these modifications remain central to the recruitment of the diverse effector proteins required to collectively activate repair mechanisms and cell cycle checkpoints [70, 71].

Despite their lack of sequence similarity, ubiquitin and SUMO are small, single-domain

proteins that share the same β-GRASP fold [72] (Figure 1.8). However, their specific surface residues and charge distributions are significantly different, supporting the distinct functions of SUMO and ubiquitin, mediated by separate sets of enzyme and substrate partners. Conjugation of ubiquitin and SUMO to target substrates typically relies on the coordinated activity of a catalytic E1–E2–E3 triad [73], with the SUMO conjugation machinery being the less complex of the two. SUMO becomes activated by the dimeric UBA2/AOS1 complex and is subsequently transferred to Ubc9, the only known SUMO E2 [74]. Although Ubc9 is able to transfer SUMO directly to substrates, it typically interacts with a range of SUMO E3-ligases that optimize the orientation of SUMO-bound E2 relative to the substrate and allows for efficient SUMOylation [73]. Among the better characterized SUMO E3-ligases are Ran-binding protein 2 (RanBP2) and the human family of protein inhibitor of STAT (PIAS) proteins. In contrast to SUMOylation, the activation and conjugation of ubiquitin moieties involves at least two E1 enzymes, dozens of E2 enzymes, and hundreds of different ubiquitin E3-ligases that confer substrate specificity [73, 75].

Ubiquitin and SUMO can be attached to substrates via isopeptide linkages between their C-termini and the side chain $\varepsilon NH_2$ group of Lysine residues on target proteins [72, 76]. While only a single, highly conserved ubiquitin isoform is found in lower and higher eukaryotes, four SUMO forms (SUMO 1–4) have been identified in humans (Figure 1.8). Among these, only SUMO1–3 are conjugated to substrates. SUMO2 and SUMO3 share 97% sequence identity and show ~50% amino acid identity with SUMO1. SUMO2/3 are able to form linear poly-SUMO chains primarily through a single consensus SUMOylation motif, residue Lys11, while modification by SUMO1 preferentially occurs as a linear mono-modification [73, 75, 77]. In the case of ubiquitin, seven Lysine residues may be used to potentially form a diverse range of linear and branched poly-ubiquitin (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63) in addition to linkages through its N-terminus to form linear chains [70, 76].

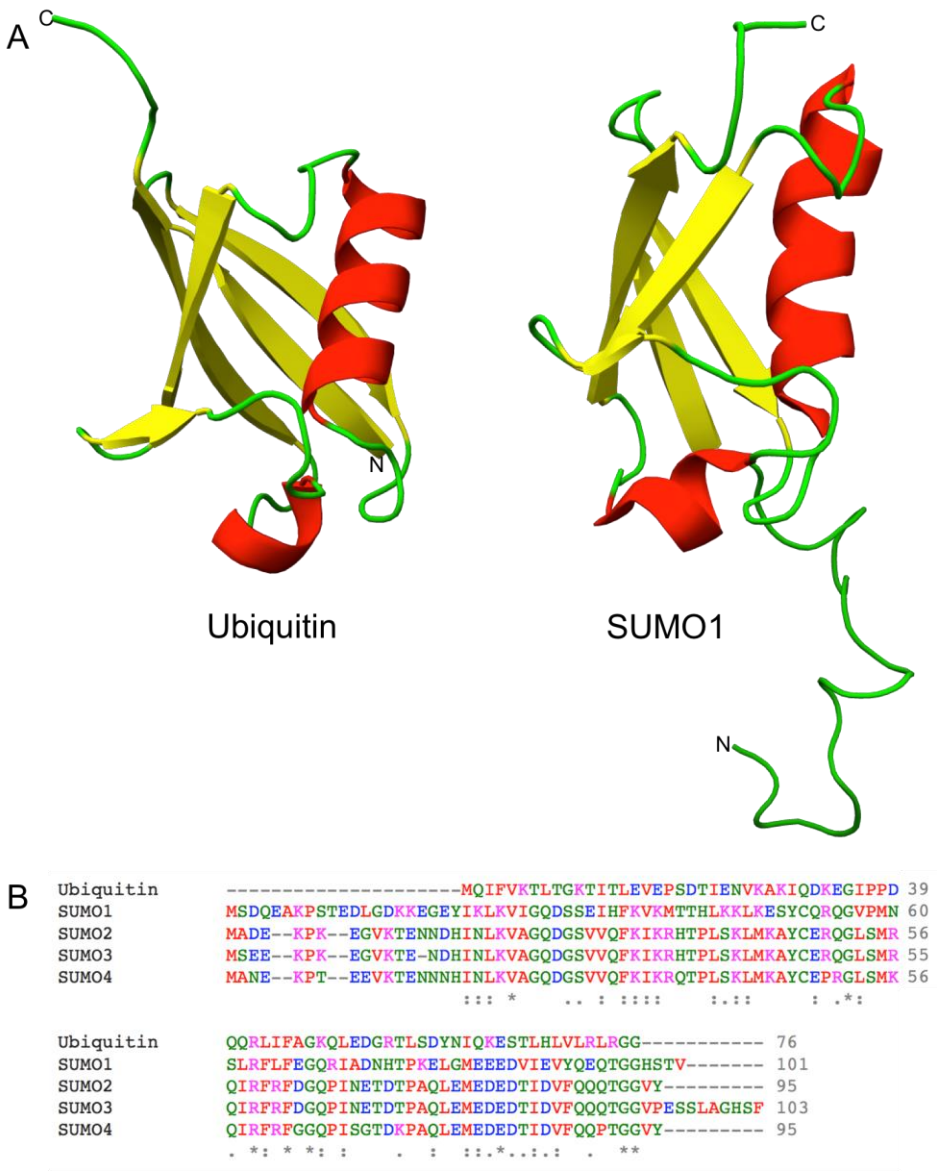



**Figure 1.8 Ubiquitin and SUMO.** A) Structural models for Ubiquitin (PDB: 1UBQ) and SUMO1 (PDB: 2KQS), B) EBI Clustal Omega sequence alignment of human ubiquitin and all four SUMO variants. Residues are colour-coded according to polarity (apolar hydrophobic residues in red, polar hydrophilic residues in green) and charge (acidic residues in blue, basic residues in pink). Sequence conservation is denoted as fully conserved (*), conservative mutations (:), semi-conservative mutations (.) and non-conservative mutations ( ).

Like other post-translational modifications, SUMOylation and ubiquitylation are reversible. While SUMO isopeptide bonds are cleaved by a small family of peptidases (SENP1– SENP3 and SENP5–SENP7), there are over 100 deubiquitylating enzymes (also known as deubiquitylases or DUBs) [75, 78]. DUBs are grouped into five families (UCH, USP, OTU, Josephins and JAMM/MPN+), four of which are Cysteine proteases and one a $Zn^{2+}$-dependent metalloprotease [76, 79, 80].

Although best known for mediating protein degradation, ubiquitin can convey a variety of non-proteolytic signals. This is due to its ability to form polymeric chains of different geometries, recognized by highly chain-selective effector proteins [71]. Similarly, SUMO has been identified as a highly diverse modification [81], which may also trigger its targets for degradation by forming polySUMO2/3 chains that in turn recruit a class of enzymes known as SUMO-targeted ubiquitin ligases (STUbLs) [82, 83]. Hence, both proteolytic and non-proteolytic contributions must be considered when assessing the effects of ubiquitin and SUMO on cellular processes, including DNA repair.

A link between Lys63-ubiquitin chains and the DDR in mammalian cells remained elusive until several groups identified RAP80 as a binding partner of breast cancer 1, early onset (BRCA1) [70]. RAP80 contains a tandem ubiquitin-interacting motif (UIM) that binds with high affinity to Lys63 linkages *in vitro* and associates with them *in vivo* following DNA damage [84]. Importantly, mutations in the RAP80 UIM and in the BRCA1/RAP80 associated protein Abraxas [85, 86] have been described in familial breast cancer cases where the *BRCA1* and *BRCA2* genes are not affected. These disease associated mutations highlight the importance of this specific ubiquitin interaction for genome stability.

Furthermore, RING finger protein 8 (RNF8) identified as the first Ubiquitin E3-ligase to catalyse poly-ubiquitin Lys63 linkages at DSBs in mammals, is recruited by mediator of DNA damage checkpoint 1 (MDC1) after ATM phosphorylation resulting in histone H1 ubiquitylation via the UBC13 E2 ligase, and thus in turn recruits a second E3-ligase, RNF168, to ubiquitylate histone H2A at positions K13 and K15 [87, 88]. Collectively these ubiquitylations establish chromatin changes that facilitate the recruitment of

other DSB response factors including 53BP1 and BRCA1.

As mentioned above, SUMOylation has been demonstrated to be part of a concerted and synergistic targeting effort against a protein group, rather than a specific substrate. In this regulated and spatially restricted SUMO "wave", generalized SUMOylation results in stabilization cellular protein complexes [69]. Among the extensive and diverse targets demonstrated to be SUMOylated following DNA damage are BM1, HERC2, RNF168, 53BP1, BRAC1, RAP80, MDC1, RPA70, CtIP and Exo1 [75, 76, 89]

Although conceptually it might be convenient to treat ubiquitylation and SUMOylation separately, it is important to recognize that they can act in a combinatorial fashion, through the range of SUMO-targeted E3 ubiquitin ligase (STUbL) [70]. RNF4, one such STUbL, localizes to SUMO-modified MDC1 at DSB sites where its ubiquitin ligase activity is required for effective RAP80-BRCA1 recruitment. RAP80 itself contains a SIM domain adjacent to its tandem UIM domains and all cooperate in the productive association of RAP80-BRCA1 to damage sites. Importantly, BRCA1 itself is an ubiquitin E3-ligase that can catalyze poly-ubiquitin Lys6 linked chains *in vitro*, with a stimulated activity upon BRCA1 SUMOylation [90, 91].

An important aspect of ubiquitin and SUMO signalling is that they are recognized and coupled to specific cellular responses by interactions with effector proteins containing specialized binding domains. Ubiquitin binding domains (UBDs) are diverse modules in a protein that can bind, and often distinguish, between different types of ubiquitin modifications. Over 20 different families of UBDs have been identified to date, with diverging structures and molecular mechanisms for ubiquitin recognition [92].

Most UBDs use α-helical structures to bind a hydrophobic patch in the β-sheet of ubiquitin. The ubiquitin-interacting motif (UIM), inverted UIM (MIU) and ubiquitin-binding zinc finger (UBZ) bind this region of ubiquitin with a single α-helix, oriented either parallel or antiparallel to the central β-strand [92, 93]. The ubiquitin associated (UBA) and coupling of ubiquitin conjugation to endoplasmic reticulum degradation

(CUE) domains have structural homology, with a shared three-helical bundle architecture and the ability to bind ubiquitin through two discontinuous α-helices. The common use of the Ile44-containing surface by UBDs causes their binding to monoubiquitin to be mutually exclusive [92, 94]. It is possible that such regulation is an important component of effective ubiquitin signalling, as it can prevent each ubiquitin modification from stimulating multiple processes, leading to discordance within the cell.

There is no evidence that the binding affinity for ubiquitin can be predicted on the basis of the UBD structural fold, with a broad range typically between 1μM and 500 μM, typically [94-96]. As a crude generalization, binding domains in enzymes with activities involved in or regulated by ubiquitination tend to have affinities in the low micromolar range, though the largest class of characterized UBDs (including IUM, UBA and CUE domains) do have mono-ubiquitin affinities which might seem surprisingly low for a physiological interaction [92, 94]. However, such relatively weak interactions are physiologically relevant, as abrogating mutagenesis have been shown to be detrimental *in vivo*. Ubiquitin interactions might be relatively weak because they represent readily reversible, transitory protein networks in which clustering of high-specificity, low-affinity interactions creates a network that is fluid, can undergo rapid assembly and disassembly, and is only transiently stable following the formation of several modular interactions. Such a “dynamically unstable” systems allow for regulatory plasticity and are susceptible to redundant and highly adjustable regulation of the many cumulative interactions in the network [93]. Finally, it is important to note that UBDs not only recognize ubiquitin conjugates but can also function as readers for proteins, which contain ubiquitin-like domains [73, 97-99].

Similar to what has been observed for ubiquitin, signalling by SUMO often relies on the recognition of the post-translational mark by a specialized interaction module, termed the SUMO interacting motif (SIM) [100]. In many cases SIM-mediated interactions direct the assembly of protein complexes, like the SIM-dependent targeting of Daxx to Promyelotic Leukemia (PML) nuclear bodies, respectively [101]. SIMs have been identified in a wide range of proteins and are typically characterized by a short

consensus sequence of hydrophobic residues, which folds into a small β-strand that becomes inserted between helix 1 and β strand 2, extending the SUMO β-sheet [74, 100].

Notably, acidic residues are flanking the hydrophobic patch with NMR studies suggesting that the acidic stretch mediates electrostatic interactions with a surface area formed by basic residues of SUMO [100, 102]. Importantly, negative charges that stabilize SUMO-SIM interactions can also be provided by phosphorylation of serine or threonine residues in close proximity to the hydrophobic core [74, 100]. Thus, it has been demonstrated that in a subset of SIM-containing proteins (including RAP80 [90], PML [101], and PIAS1 [103, 104]) phosphorylation by the serine/threonine kinase 2 (CK2) within a phosphorylated SIM module enhances the binding to SUMO paralogs. This mechanism could act as a molecular switch that modulates the dynamics of SIM interactions.

## 1.4 Aims of this thesis

SMARCAD1 remains a poorly understood yet central actor in the key cellular processes like DNA replication, DNA repair and heterochromatin maintenance. Evidence firmly demonstrates that despite its non-canonical architecture, SMARCAD1 is capable of establishing a multitude of protein-protein interactions that regulate and recruit it to its diverse cellular roles.

Recent reports associate the sole known ancillary CUE domains to the resection-stimulating recognition of ubiquitylated histone H2A downstream of BRCA1 in the homologous recombination pathway for double-strand break repair, yet don't account for the well-established and robust interaction with transcriptionally repressive master regulator KAP1, nor has it been addressed what potential role might SMARCAD1 have in the repair of heterochromatin DNA lesions. To this end, the known binding domains of SMARCAD1 have been structurally and biochemically characterized, assessing the potential role they may have on DNA repair in heterochromatin regions.

1. Structural elucidation of SMARCAD1 CUE1 domain.

   Despite previous efforts, the structure of CUE1 from SMARCAD1 is still not fully characterized. Previous work in our group attempted to solve the solution structure of the tandem CUE domain, but due to technical limitations only CUE2 was satisfactorily characterized. Renewed efforts to fully define its secondary and tertiary structure through solution NMR were carried out, resulting in a much improved model, which resolved previously ambiguous features and is highly consistent with the family fold.

2. Identification of a novel SUMO interacting motif

   A new SIM has been identified and confirmed through solution NMR. Moreover, its relationship with previously reported ATM/ATR phosphorylation sites was characterized through NMR and mass spectrometry, resulting in a model which presents biologically relevant affinity and is in agreement with SIM mediated interactions with SUMO.

3. Identification of CUE1- and SIM-mediated DDR specific protein interactions

   Characterization of protein profiles was carried out with co-immunoprecipitation of SMARCAD1 functional mutants expressed in human cultured cells and analysed by quantitative mass spectrometry to assess relative changes in abundance upon chemically induced DNA damage

4. Biochemical and Biophysical characterisation of CUE1 mediated interaction with KAP1

   Detailed characterisation of the CUE1 specific interaction with KAP1 through complementary biochemical and biophysical assays

Chapter 2

# Materials and Methods

This chapter comprises a detailed description of all materials and methods used throughout my research. It is organized to follow the sequential steps to produce an overexpressed recombinant protein from a synthetic gene sequence and its subsequent structural or functional characterisation. There is particular emphasis on general, standardised protocols that have been modified to overcome a particular obstacle.

## 2.1 DNA Handling and Sub-cloning.

A variety of standard DNA handling methods were employed during this work, the reader is directed to the detailed description found in *Molecular Cloning: A Laboratory Manual* [105]. The protein sequences of human SMARCAD1 and KAP1 were analysed using secondary structure and disorder prediction software, the result of which was used along with details of reported functional domains to define variable domain boundaries and led to libraries of truncation constructs in combination with multiple solubility tags as fusion proteins for each protein. Oligonucleotides were designed for the use of ligation independent cloning (LIC)-complementary vectors of the pOPIN suite from OPPF [106] (Section 2.1.1) and constructs were amplified using the Polymerase Chain Reaction (PCR) (Section 2.1.2). The purified PCR fragments were transferred to linearized vectors using LIC (Section 2.1.3), amplified in chemically competent DH5α *E.coli* cells (Section 2.1.4), purified and sequence verified before being used for heterologous expression. Furthermore, a number of point mutants were designed in order to probe the functional role of specific sequence regions (Section 2.1.5).

2.1.1 Oligonucleotide Primer Design

Oligonucleotides with complementary sequences to template cDNA were designed using the OPINER primer utility [107], optimised for downstream use of InFusion™ (Takara-Clontech) LIC. Domain boundaries used in designing truncations were derived from a combination of amino acid secondary structure predictions from PredictProtein [108] and ProteinCCD [109] together with domain designations shown in the Uniprot Knowledge protein database. Additionally, low complexity regions as predicted by RONN [110] were avoided in the constructs whenever possible (Figure 2.1).

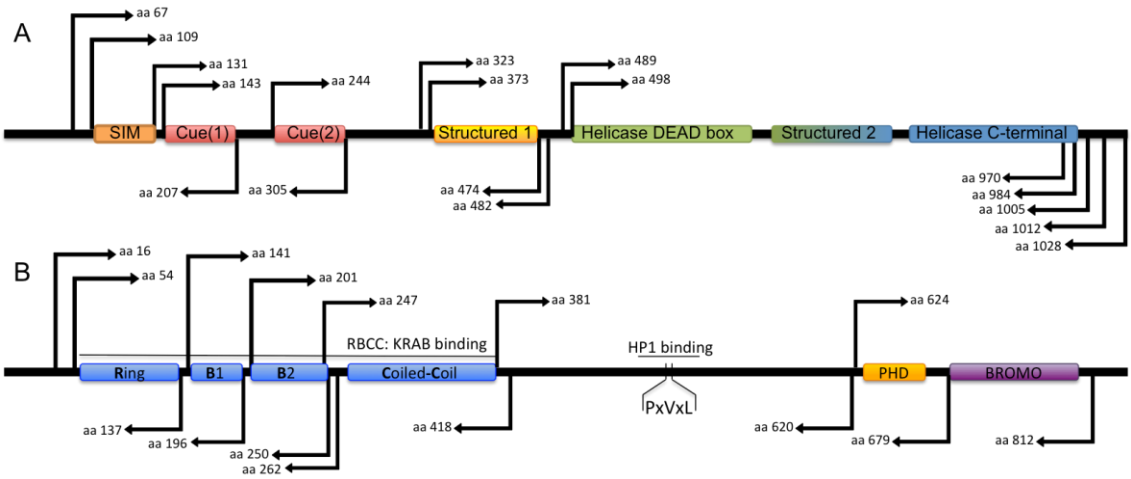



**Figure 2.1 Domain architecture and recombinant construct design of A) human SMARCAD1 and B) human KAP1/Trim28/TIF1-β.**

All fragments were amplified using full-length human SMARCAD1 cDNA (IOH26772, Invitrogen ORF collection) or human KAP1/Trim28/TIF1-β cDNA (IOH5106 Invitrogen ORF collection) as template, kindly donated by Dr. Patrick Varga-Weisz (Babraham Institute, University of Cambridge). These correspond to UniprotKB/SwissProt entries Q9H4L7-2 and Q13263, respectively. A total of 67 forward and reverse primers were designed complementary to 5'- and 3'-end cDNA template sequences for each construct, all with an annealing temperature of ~60-65°C and an average GC content of (50±10)%. In addition to the template specific sequence, the OPINER suite annexes a 15-nucleotide sequence homologous to the linearized ends of the target pOPIN vector, for the subsequent LIC cloning step. Primers were synthesized by Eurofins MWG Operon (Germany), delivered at a standardized concentration of 100µM in 10mM Tris

HCl pH8.0 and stored at -20°C. Table 2.2 includes all oligonucleotide sequences used as primers in this work.

### 2.1.2 DNA Amplification by Polymerase Chain Reaction (PCR)

DNA fragments corresponding to the various truncation constructs designed were amplified by PCR in 50µl reaction mixes using Phusion Flash High-Fidelity 2X MasterMix (Thermo) polymerase according to the manufacturer's instructions, with 25pmol of each forward and reverse primer, ~1ng of template plasmid cDNA per reaction and 1% DMSO. Mixtures were prepared on ice, using thin walled PCR tubes, 8-well strings or skirted 96-well plates (Thermo). PCR products were obtained after 35 cycles in a T100™ thermal cycler (BioRad) using the temperature programme detailed in Table 2.1

**Table 2.1 Thermo cycling conditions used for PCR**

| PCR Step | Cycles | Time | Temperature (°C) |
|---|---|---|---|
| Denaturation | 1 | 30s | 98 |
| Denaturation | 35 | 30s | 98 |
| Annealing | | 20s | 65 |
| Elongation | | 15s/kb | 72 |
| Elongation | 1 | 10min | 72 |

Amplicons were analysed using agarose gel electrophoresis by mixing a sample volume of the PCR product with GelPilot DNA loading dye 5x (QIAgen) in a 5:1 ratio, and separated with a 1-1.5%$^{w}/_{v}$ agarose gel in 1x Tris-Borate-EDTA buffer (10x, Sigma-Aldrich-Aldrich) containing Sybr® Safe or Sybr® Gold (1x final concentration, Invitrogen). DNA bands were visualized using a Chemidoc MP Imager (BioRad) with a UV trans-illuminator. PCR products were treated with 20-40 units of DpnI enzyme (New England Biolabs) by incubation at 37°C for 1h to remove methylated template cDNA, purified using a commercial silica-membrane purification kit (QIAquick PCR Purification, QIAgen) and eluted with DNAse-free water (Invitrogen).

### 2.1.3 Ligation-independent Cloning

Cloning was carried out through a ligation-independent protocol employing the commercially available InFusion™ enzyme (Takara-Clontech), capable of inserting a PCR fragment bearing a minimum 15-nucleotide homologous dsDNA sequence at each end that matches the linearized ends of a target vector in a process akin to homologous recombination. All pOPIN vectors were linearized with the appropriate restriction enzyme pair and purified with a silica-membrane commercial kit (QIAquick PCR Purification, QIAgen). 100-150ng of linearized vector were mixed with 50ng of PCR fragment and diluted with DNAse-free water (Invitrogen) to 10µL; this solution was used to resuspend the lyophilised enzyme pellet and briefly mixed until homogeneity. The reaction was sealed and incubated at 42°C for 30min in a T100™ thermal cycler (BioRad), followed by immediate dilution with four volumes of TE buffer. The diluted mixture was immediately transformed into chemically competent DH5α *E.coli* bacteria (Library efficiency, Invitrogen) and plated on LB/Agar supplemented with 50µg/mL Carbenicillin (Melford). For empty vectors that include the β-Galactosidase gene, as the pOPIN suite does, the plates were further supplemented with 1mM IPTG (Melford) and 1000x dilution of 20%$^{w}/_{v}$ X-Gal stock in DMSO for blue/white screening. Table 2.3 includes all vectors used in this study.

**Table 2.2 Oligonucleotides used as for PCR construct design throughout this work**

| Name | Template | Vector | Sequence (5' to 3') |
|---|---|---|---|
| hS1_aa67_NtHis_F | pENTR221/ SMARCAD1 IOH26772 | pOPIN F | AAG TTC TGT TTC AGG GCC CGA GTG TTC CAG AAA CTC CAG ATA ATG AAA GAA AAG C |
| hS1_aa109_NtHis_F | | | AAG TTC TGT TTC AGG GCC CGA CAG TTC AAG AGA AAA CAT TCA ACA AAG ATA CAG TG |
| hS1_aa131_NtHis_F | | | AAG TTC TGT TTC AGG GCC CGG AGT CCC AAG GCC TTC CTA CCA TGG |
| hS1_aa143_NtHis_F | | | AAG TTC TGT TTC AGG GCC CGG ATG ATA TTT CAG AAC TGG AAG ACC TTT CGG |
| hS1_aa207_NtHis_R | | | ATG GTC TAG AAA GCT TTA TTT CCT GGG CCC ACC ACC TG |
| hS1_aa244_NtHis_F | | | AAG TTC TGT TTC AGG GCC CGG AAT CTA GCA ATA ATT GGG AAA AGC AGG AAA G |
| hS1_aa305_NtHis_R | | | ATG GTC TAG AAA GCT TTA CTC ACT TTG TGA TGC ATA TTG CAT ATC TTG G |
| hS1_aa323_NtHis_F | | | AAG TTC TGT TTC AGG GCC CGG CAA CTA AAA CAA AAC TAA AAC AGA AAT TTT CAA TGA AAG |
| hS1_aa373_NtHis_F | | | AAG TTC TGT TTC AGG GCC CGG ACT ATA GTA GTG GTG AAG AAG TGA TGG AGG ATG G |
| hS1_aa474_NtHis_R | | | ATG GTC TAG AAA GCT TTA GGT AAC TTG TTT GGT CAA TTT ATT TGA AAT GTC TTC |
| hS1_aa482_NtHis_R | | | ATG GTC TAG AAA GCT TTA TCC ACC TCC ATT TCC AGT AAG CAT GG |
| hS1_aa489_NtHis_F | | | AAG TTC TGT TTC AGG GCC CGT CCA TTC TAA ACC AGA GTT TGT CAC TCA AGC |
| hS1_aa498_NtHis_F | | | AAG TTC TGT TTC AGG GCC CGA AGC CCT ATC AGA AGG TTG GTT TGA ATT G |
| hS1_aa970_NtHis_R | | | ATG GTC TAG AAA GCT TTA CTG GCC TAC TCT ATG GCA TCT ATC TTC TGC |
| hS1_aa984_NtHis_R | | | ATG GTC TAG AAA GCT TTA CGT CCC TTG GCT TAT TAG TTT TAT AAC TAG TGC TTC |
| hS1_aa1005_NtHis_R | | | ATG GTC TAG AAA GCT TTA TGT AGT CAT ATC CTG TTC TAG TTT CAA TTT CTG TTG G |
| hS1_aa1012_NtHis_R | | | ATG GTC TAG AAA GCT TTA CCC TTC ATC ACC TTC ATC TAC TGT AGT CAT ATC C |
| hS1_aa1028_NtHis_R | | | ATG GTC TAG AAA GCT TTA CAG GCC CAT TGA TGT TTT TAG TAA TGT GGC |

**Table 2.2 (Continued) Oligonucleotides used for construct design throughout this work**

| Name | Template | Vector | Sequence (5' to 3') |
|---|---|---|---|
| hS1_aa67_CtHis_F | pENTR221/ SMARCAD1 IOH26772 | pOPIN E | AGG AGA TAT ACC ATG AGT GTT CCA GAA ACT CCA GAT AAT GAA AGA AAA GC |
| hS1_aa109_CtHis_F | | | AGG AGA TAT ACC ATG ACA GTT CAA GAG AAA ACA TTC AAC AAA GAT ACA GTG |
| hS1_aa131_CtHis_F | | | AGG AGA TAT ACC ATG GAG TCC CAA GGC CTT CCT ACC ATG G |
| hS1_aa143_CtHis_F | | | AGG AGA TAT ACC ATG GAT GAT ATT TCA GAA CTG GAA GAC CTT TCG G |
| hS1_aa207_CtHis_R | | | GTG ATG GTG ATG TTT TTT CCT GGG CCC ACC ACC TG |
| hS1_aa244_CtHis_F | | | AGG AGA TAT ACC ATG GAA TCT AGC AAT AAT TGG GAA AAG CAG GAA AG |
| hS1_aa305_CtHis_R | | | GTG ATG GTG ATG TTT CTC ACT TTG TGA TGC ATA TTG CAT ATC TTG G |
| hS1_aa323_CtHis_F | | | AGG AGA TAT ACC ATG GCA ACT AAA ACA AAA CTA AAA CAG AAA TTT TCA ATG AAA G |
| hS1_aa373_CtHis_F | | | AGG AGA TAT ACC ATG GAC TAT AGT AGT GGT GAA GAA GTG ATG GAG GAT GG |
| hS1_aa474_CtHis_R | | | GTG ATG GTG ATG TTT GGT AAC TTG TTT GGT CAA TTT ATT TGA AAT GTC TTC |
| hS1_aa482_CtHis_R | | | GTG ATG GTG ATG TTT TCC ACC TCC ATT TCC AGT AAG CAT GG |
| hS1_aa489_CtHis_F | | | AGG AGA TAT ACC ATG TCC ATT CTA AAC CAG AGT TTG TCA CTC AAG C |
| hS1_aa498_CtHis_F | | | AGG AGA TAT ACC ATG AAG CCC TAT CAG AAG GTT GGT TTG AAT TG |
| hS1_aa970_CtHis_R | | | GTG ATG GTG ATG TTT CTG GCC TAC TCT ATG GCA TCT ATC TTC TGC |
| hS1_aa984_CtHis_R | | | GTG ATG GTG ATG TTT CGT CCC TTG GCT TAT TAG TTT TAT AAC TAG TGC TTC |
| hS1_aa1005_CtHis_R | | | GTG ATG GTG ATG TTT TGT AGT CAT ATC CTG TTC TAG TTT CAA TTT CTG TTG G |
| hS1_aa1012_CtHis_R | | | GTG ATG GTG ATG TTT CCC TTC ATC ACC TTC ATC TAC TGT AGT CAT ATC C |
| hS1_aa1028_CtHis_R | | | GTG ATG GTG ATG TTT CAG GCC CAT TGA TGT TTT TAG TAA TGT GGC |

**Table 2.2 (Continued) Oligonucleotides used for construct design throughout this work**

| Name | Template | Vector | Sequence (5' to 3') |
|---|---|---|---|
| hS1_aa109_CtGFP_3xNLS_F | pENTR221/ SMARCAD1 IOH26772 | pOPIN GFP/CHis | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC ACA GTT CAA GAG AAA ACA TTC AAC AAA GAT ACA GTG |
| hS1_aa143_CtGFP_3xNLS_F | | | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC GAT GAT ATT TCA GAA CTG GAA GAC CTT TCG G |
| hS1_aa244_CtGFP_3xNLS_F | | | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC GAA TCT AGC AAT AAT TGG GAA AAG CAG GAA AG |
| hS1_aa305_CtGFP_R | | | CAG AAC TTC CAG TTT CTC ACT TTG TGA TGC ATA TTG CAT ATC TTG G |
| hS1_aa2_NtGFP_pHAGE_F | | pHAGE NtGFP | GTA CAA GGC TCG AGG TAA TCT TTT CAA CCT GGA CCG TTT TC |
| hS1_aa1028_NtGFP_pHAGE_R | | | GGG GCT GCT AAA GCG CCA GGC CCA TTG ATG TTT TTA GTA ATG TGG C |
| Ubiquitin_NtMAHisG_F | pT3a/Ubiquitin | pOPIN F | TCA AAG GAG ATA TAC CAT GGC ACA CCA TCA CCA CCA TCA CGG TCA GAT CTT CGT CAA GAC GTT AAC CGG |
| Ubiquitin_NtMAHisG_R | | | ATG GTC TAG AAA GCT TTA ACC ACC TCT TAG TCT TAA GAC AAG ATG TAA GGT CG |
| SUMO1_NtMAHisG_F | pET15b/NHis SUMO1 | | TCA AAG GAG ATA TAC CAT GGC ACA CCA TCA CCA CCA TCA CGG TTC TGA CCA GGA GGC CAA ACC TTC |
| SUMO1_Q92R_NtMAHisG_R | | | ATG GTC TAG AAA GCT TTA ACC CCC CGT TTG TTC CCG ATA AAC TTC |
| Kap1_aa16_CtGFP_3xNLS_F | pENTR221/Kap-1 IOH5106 | pOPIN GFP/CHis | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC GCC TCT GGC AGC CCG G |
| Kap1_aa54_CtGFP_3xNLS_F | | | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC GGC GGC GCC GAG GC |
| Kap1_aa137_CtGFP_R | | | CAG AAC TTC CAG TTT ATC ACG CAT GAA ATA ATT CTC CAC GAT G |
| Kap1_aa141_CtGFP_3xNLS_F | | | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC AAG GCT GCC ACC GAC GC |
| Kap1_aa196_CtGFP_R | | | CAG AAC TTC CAG TTT CCC AGT AGA GCG CAC AGT ATG GTC C |
| Kap1_aa201_CtGFP_3xNLS_F | | | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC CGG GAT GGT GAA CGT ACT GTC TAT TG |

**Table 2.2 (Continued) Oligonucleotides used for construct design throughout this work**

| Name | Template | Vector | Sequence (5' to 3') |
|---|---|---|---|
| Kap1_aa247_CtGFP_3xNLS_F | pENTR221/Kap-1 IOH5106 | pOPIN GFP/CHis | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC GAT GCA GTG AGG AAC CAG CGC |
| Kap1_aa250_CtGFP_R | | | CAG AAC TTC CAG TTT CCT CAC TGC ATC CTC TAA GAA CTG GTA CTG |
| Kap1_aa381_CtGFP_3xNLS_F | | | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC GAT CCC GTG GAG CCA CAT GG |
| Kap1_aa418_CtGFP_R | | | CAG AAC TTC CAG TTT TGT TGA GTT AGT GCC AGG ACG CTC |
| Kap1_aa620_CtGFP_R | | | CAG AAC TTC CAG TTT GGT TCC CGG GCC ACC AC |
| Kap1_aa624_CtGFP_3xNLS_F | | | AGG AGA TAT ACC ATG GCA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GAT CCA AAA AAG AAG AGA AAG GTA GGA TCC ACC AGT GCC ACC ATT TGC CGT GTC |
| Kap1_aa679_CtGFP_R | | | CAG AAC TTC CAG TTT ATC CTC CTC CTT CAG GTC AGG GAG |
| Kap1_aa812_CtGFP_R | | | CAG AAC TTC CAG TTT GGG CTC CAC CAG CAC AGC AG |
| Kap1_aa16_NtHis_F | | pOPIN F | AAG TTC TGT TTC AGG GCC CGG CCT CTG GCA GCC CGG |
| Kap1_aa418_NtHis_R | | | ATGG TCT AGA AAG CTT TAT GTT GAG TTA GTG CCA GGA CGC TC |
| Kap1_aa16_CtHis_F | | pOPIN E | AGG AGA TAT ACC ATG GCC TCT GGC AGC CCG G |
| Kap1_aa418_CtHis_R | | | GTG ATG GTG ATG TTT TGT TGA GTT AGT GCC AGG ACG CTC |

**Table 2.3 Vectors used as linearized targets for InFusion™ cloning**

| Name | Description | Source |
|---|---|---|
| pOPIN F | 5531bp, In-Fusion enabled, pTriEx2-based vector for expression of HHHHHH-(3C)-POI in E. coli, HEK 293T, Baculovirus, AmpR | OPPF |
| pOPIN E | 5534bp, In-Fusion enabled, pTriEx2-based vector for expression of POI-KHHHHHH in E. coli, HEK 293T, Baculovirus, AmpR | OPPF |
| pOPIN GFP/KHis | 6265bp, In-Fusion enabled, pTriEx2-based vector for expression of POI-(3C)-eGFP-KHHHHHH in E. coli, HEK 293T, Baculovirus, AmpR | OPPF |
| pHAGE CMV eGFP W | 7149bp, Lentiviral enabled, HIV-1-based backbone expressing eGFP fluorescent marker from CMV promoter, with WPRE, CccdB/CamR bacterial amplification, AmpR | Harvard Medical School |

### 2.1.4 Mutagenesis

Substitution mutations of specific amino acids were performed to assess their contribution to protein function. Both site-directed mutagenesis and ligation-independent fragment assembly were employed to create the plasmids bearing mutated sequences of human SMARCAD1. These were later directly used for expression or as templates for truncation constructs.

Site-directed mutagenesis (SDM) was employed to introduce the single point Phe169Ala mutation in SMARCAD1 using the commercially available QuickChange II kit (Agilent) and following the recommended protocol. Two overlapping primers with opposite directionality and carrying the appropriate codon change were designed using the design utility recommended by the manufacturer [111, 112] and optimised for the application with an annealing temperature >78°C and at least 15 bases to either side of the mutation bubble. PCR reactions in a total volume of 50µL were carried out with 125ng of both forward and reverse primers, ~50ng of template plasmid cDNA per reaction, 2.5U of *PfuUltra* HD DNA Polymerase and 1% DMSO. Mixtures were prepared on ice and PCR products were obtained after 18 cycles in a T100™ thermal cycler (BioRad) using the temperature programme detailed in Table 2.4.

**Table 2.4 Thermo cycling conditions used for QuickChange II SDM**

| PCR Step | Cycles | Time | Temperature (°C) |
|---|---|---|---|
| Denaturation | 1 | 30s | 95 |
| Denaturation | 18 | 30s | 95 |
| Annealing | | 20s | 60 |
| Elongation | | 60s/kb | 68 |
| Elongation | 1 | 10min | 72 |

Amplicons were treated with 10 units of DpnI enzyme by incubation at 37°C for 1h to remove methylated template DNA and immediately transformed into XL1-Blue chemically competent cells followed by plating on LB/Agar supplemented with 50µg/mL Carbenicillin (Melford). Colonies bearing mutated plasmids were amplified and purified (Section 2.1.5).

DNA fragment assembly was employed to create alternative mutant forms of human SMARCAD1 where several point mutations are simultaneously required, exploiting the ease of short DNA sequence commercial synthesis (GeneArt™ service, Invitrogen) and InFusion™ mediated ligation-independent cloning. The overall aim is to replace a section of the pHAGE/NtGFP WT SMARCAD1 plasmid (Table 2.6) with an equivalent sequence bearing the desired codon variations: a SUMO interacting motif ablation mutant 118DTVIIVSE125 to $A_8$ (SIM8A), a triple phosphoserine alanine substitution mutant S124A/S127A/S132A (S3A) and a triple phosphoserine mimic mutant S124E/S127E/S132E (S3E). Thus, due to the complex and extensive work it would otherwise require, traditional site-directed mutagenesis was not deemed suitable.

First, a sequence encoding the homologous region of the pHAGE/NtGFP WT SMARCAD1 cDNA between bases 3538 and 6272, framed by XhoI and KpnI restriction enzyme sites, served as a general insert for the requisite linearization of the target vector during InFusion™. This 2739bp sequence spans over 80% of the protein, including the extended SUMO interacting motif near the N-terminus, however the sequence flanking the region of interest in human SMARCAD1 had to be modified to introduce new unique restriction enzyme sites. Two sets of codon exchanges, as noted in Figure 2.2, introduced silent mutations coupled with novel NheI and SalI cut sites.

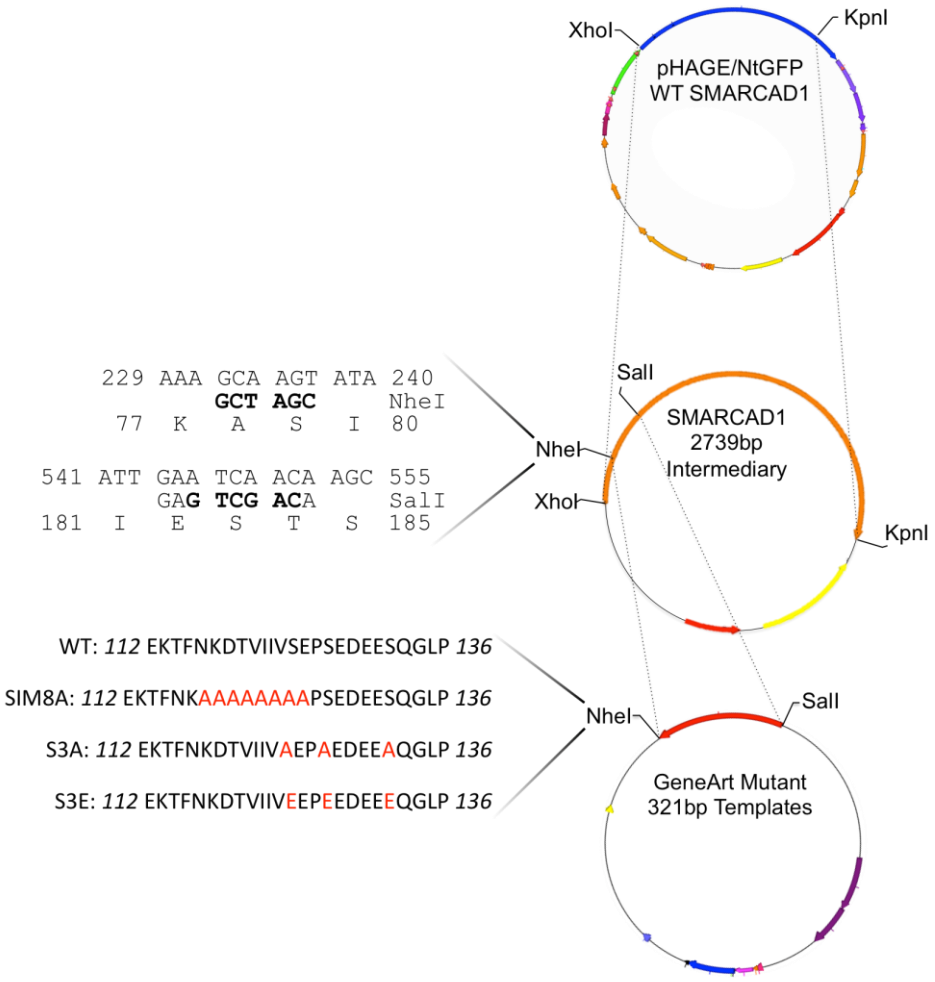



**Figure 2.2 Ligation-independent fragment assembly strategy for SMARCAD1 SIM substitution mutants design.**

In addition to this general intermediary plasmid sequence, a set of three short 321bp fragments with the mutated sequences between bases 232 and 552 were designed, each with one of the desired mutant sequences. These plasmids were used as templates for PCR amplification using primers designed for InFusion™ cloning. These amplicons were inserted into the linearized intermediary plasmid, creating a new set of templates for a second round of PCR and InFusion™ cloning, this time into the linearized pHAGE/GFP WT SMARCAD1 plasmid. All cloning steps followed the general guideline set above and were sequence verified. Oligonucleotides used for mutagenesis experiments, both for site-directed and fragment assembly, are shown in Table 2.5.

**Table 2.5 Oligonucleotides used for mutagenesis**

| Name | Sequence (5' to 3') |
|---|---|
| F169A SDM Sense | GATGCTAAACTTCAGACTTTGAAGGAACTTGCTCCACAAAGAAGTGAC |
| F169A SDM Antisense | GTCACTTCTTTGTGGAGCAAGTTCCTTCAAAGTCTGAAGTTTAGCATC |
| eSIMut NheI nt232 F | CTCCAGATAATGAAAGAAAAGCTAGCATATCATATTTCAAAAATCAAAGAGG |
| eSIMut SalI nt552 R | CAATTGCTCCATCCATAGTGCTTGTCGACTCAATCAACTTAAGTAAATCATTGTC |
| eSIMut XhoI nt3538 F | GCATGGACGAGCTGTACAAGGCTCGAGGTTATCAAACAAGTTTGTACAAAAAAGC |
| eSIMut KpnI nt6272 R | GAGTCTTTCCATCTAATCTGAGGTACCTATGCTGATGATGTTTTAATAGAACCTC |

2.1.5 Plasmid DNA amplification

Bacterial colonies were selected and used to inoculate small-scale (~15mL) cultures consisting of Luria Broth (LB) media (Millipore) supplemented with 50µg/mL of Carbenicillin (Melford), which were incubated at 37°C overnight while shaking at 160rpm. The cultures were centrifuged at 4500x*g* for 15min at 4°C and the plasmid DNA was purified from the resulting pellets using a commercial silica membrane extraction kit (QIAprep Spin Miniprep, QIAgen), following the recommended protocol. Purified plasmid is eluted with DNAse-free water (Invitrogen) and stored at -20°C. All plasmids were sequence verified with Source BioScience (United Kingdom) or GATC Biotech AG (Germany) using in-house universal primers against vector specific features or insert specific primers when necessary.

2.1.6 Plasmids

Table 2.6 denotes all plasmids used and described in this work for heterologous protein expression.

**Table 2.6 Description of plasmids used for protein expression throughout this work**

| Name | POI | Description |
|---|---|---|
| pOPINE/WT eCUE1 CtHis | SMARCAD1 SIM/CUE1, 11.83kDa C-terminal HexaHistidine | pOPINE linearized with NcoI/PmeI |
| pOPINE/WT eCUE1 WCtHis | SMARCAD1 SIM/CUE1, 12.15kDa Ct Triptophan/HexaHistidine | pOPINE linearized with NcoI/PmeI |
| pOPINE/WT CUE1 WCtHis | SMARCAD1 CUE1, 8.32kDa Ct Triptophan/HexaHistidine | pOPINE linearized with NcoI/PmeI |
| pOPINE/WT CUE2 CtHis | SMARCAD1 CUE2, 8.59kDa Ct HexaHistidine | pOPINE linearized with NcoI/PmeI |
| pOPINE/WT edCUE CtHis | SMARCAD1 SIM/CUE1/CUE2, 23.60kDa Ct HexaHistidine | pOPINE linearized with NcoI/PmeI |
| pOPINE/WT dCUE CtHis | SMARCAD1 CUE1/CUE2, 19.77kDa Ct HexaHistidine | pOPINE linearized with NcoI/PmeI |
| pOPINE/F169A eCUE1 WCtHis | SMARCAD1 F169A SIM/CUE1, 12.08kDa Ct Triptophan/HexaHistidine | pOPINE linearized with NcoI/PmeI |
| pOPINE/S3E eCUE1 WCtHis | SMARCAD1 S124E/S127E/S132E SIM/CUE1, 12.20kDa Ct Triptophan/HexaHistidine | pOPINE linearized with NcoI/PmeI |
| pOPINGFP/WT eCUE1 NLSx2 CtGFP | SMARCAD1 SIM/CUE1, 42.19kDa,Nt 2x SV40 NLS, Ct eGFP/HexaHistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/WT CUE1 NLSx2 CtGFP | SMARCAD1 CUE1, 38.17kDa,Nt 2x SV40 NLS, Ct eGFP/HexaHistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/WT CUE2 NLSx2 CtGFP | SMARCAD1 CUE2, 38.63kDa,Nt 2x SV40 NLS, Ct eGFP/HexaHistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/WT edCUE NLSx2 CtGFP | SMARCAD1 SIM/CUE1/CUE2, 53.64kDa,Nt 2x SV40 NLS, Ct eGFP/HexaHistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/WT dCUE NLSx2 CtGFP | SMARCAD1 CUE1/CUE2, 49.81kDa,Nt 2x SV40 NLS, Ct eGFP/HexaHistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/NLSx2 GFP | eGFP, 31.14kDa Nt 2x SV40 NLS, Ct Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |

**Table 2.6 (Continued) Description of plasmids used for protein expression throughout this work**

| Name | POI | Description |
|---|---|---|
| pOPINGFP/Kap1 RBCC NLSx3 CtGFP | Kap1 Ring/B1/B2/CoiledCoil, 74.07kDa, Nt 3x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 Δ2 NLSx3 CtGFP | Kap1 HP1 Binding domain ΔRBCC ΔPHD/Bromo, 54.24kDa, Nt 3x SV40 NLS, Ct eGFP/HexaHistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 PHDBromo NLSx3 CtGFP | Kap1 PHD/Bromo, 51.09kDa, Nt 3x SV40 NLS, Ct eGFP/HexaHistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 ΔPHDBromo NLSx3 CtGFP | Kap1 Ring/B1/B2/CoiledCoil/HP1 Binding Domain, 94.01kDa, Nt 3x SV40 NLS, Ct eGFP/HexaHistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 ΔBromo NLSx3 CtGFP | Kap1 Ring/B1/B2/CoiledCoil/HP1 Binding Domain/PHD, 100.64kDa, Nt 3x SV40 NLS, Ct eGFP/HexaHistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 Ring NLSx2 CtGFP | Kap1 Ring ONLY, 38.86kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 B1 NLSx2 CtGFP | Kap1 B1 ONLY, 36.14kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 B2 NLSx2 CtGFP | Kap1 B2 ONLY, 36.00kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 CC NLSx2 CtGFP | Kap1 CoiledCoil ONLY, 51.05kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 RB NLSx2 CtCGP | Kap1 Ring/B1, 46.20kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 RBB NLSx2 CtGFP | Kap1 Ring/B1/B2, 51.45kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 BB NLSx2 CtCGFP | Kap1 B1/B2, 42.44kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 BCC NLSx2 CtGFP | Kap1 B2/CoiledCoil, 56.53kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/Kap1 BBCC NLSx2 CtGFP | Kap1 B1/B2/CoiledCoil, 62.97kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |

**Table 2.6 (Continued) Description of plasmids used for protein expression throughout this work**

| Name | POI | Description |
|---|---|---|
| pOPINF/Kap1 RBCC NtHis | Kap1 Ring/B1/B2/CoiledCoil, 46.15kDa, Nt Hexahistidine | pOPINF linearized with KpnI/HindIII |
| pOPINE/Kap1 RBCC CtHis | Kap1 Ring/B1/B2/CoiledCoil, 45.09kDa, Ct Hexahistidine | pOPINE linearized with NcoI/PmeI |
| pET3a/Ubiquitin | Ubiquitin, 8.56kDa | Prof Susan Lea, Dunn School of Pathology, University of Oxford |
| pET15b/SUMO1-GG NtHis | SUMO1-GG, 13.96kDa, Nt HexaHistidine | Dr Hsiu-Ming Shih, Taiwan Institute of Biomedical Sciences |
| pOPINF/Ubiquitin Nt MAHx6G | Ubiquitin, 9.52kDa, Nt HexaHistidine | pOPIN F linearized with NcoI/HindIII |
| pOPINF/SUMO1 Q92R Nt MAHx6G | SUMO1-GG Q92R, 12.08kDa, Nt HexaHistidine | pOPIN F linearized with NcoI/HindIII |
| pHAGE/WT SMARCAD1 NtGFP | SMARCAD1 full length WT, 146.09kDa, Nt eGFP | pHAGE CMV eGFP W linearized with XhoI/SphI |
| pHAGE/F169A SMARCAD1 NtGFP | SMARCAD1 full length F169A, 146.01kDa, Nt eGFP | SDM of pHAGE/WT SMARCAD1 NtGFP |
| pHAGE/SIM8A SMARCAD1 NtGFP | SMARCAD1 full length 118DTVIIVSE125 to 118AAAAAAAA125, 145.80kDa, Nt eGFP | pHAGE/WT SMARCAD1 NtGFP linearized with XhoI/KpnI |
| pHAGE/S3A SMARCAD1 NtGFP | SMARCAD1 full length S124A/S127A/S132A, 146.04kDa, Nt eGFP | pHAGE/WT SMARCAD1 NtGFP linearized with XhoI/KpnI |
| pHAGE/S3E SMARCAD1 NtGFP | SMARCAD1 full length S124E/S127E/S132E, 146.22kDa, Nt eGFP | pHAGE/WT SMARCAD1 NtGFP linearized with XhoI/KpnI |
| pOPINGFP/WT SMARCAD1 NLSx2 CtGFP | SMARCAD1 full length WT, 146.90kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/F169A SMARCAD1 NLSx2 CtGFP | SMARCAD1 full length F169A, 146.83kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |
| pOPINGFP/SIM8A SMARCAD1 NLSx2 CtGFP | SMARCAD1 full length 118DTVIIVSE125 to 118AAAAAAAA125, 146.49kDa, Nt 2x SV40 NLS, Ct eGFP/Hexahistidine | pOPIN GFP/KHis linearized with NcoI/PmeI |

## 2.2 Protein Expression and Purification

Multiple expression systems were employed throughout this work in an effort to produce the required amounts of purified material for protein characterisation, both structural and functional. It was possible to express the majority of constructs in bacterial systems, where the conflicting differences in codon usage between host and plasmid-encoded human cDNA were overcome by the use of an *E.coli* strain carrying a "rare codon" plasmid. In the case of larger constructs, where bacterial expression was shown to be of limited yield or resulted in misfolded proteins, the use of baculovirus-mediated insect cell expression surmounted such limitations. Finally, mammalian cell culture expression was used to fully recreate the endogenous environment of protein expression, which is crucial when investigating of post-translational modification or the formation of transient protein interactions. The specific purification protocol used for each recombinant protein in this thesis is given after the generalised description of protein expression, itemized according the host system.

### 2.2.1 Bacterial expression of unlabelled proteins

Plasmids bearing a T7 promoter region (constructs in pOPIN vectors, pET15b or pT3a) were transformed into chemically competent Rosetta 2 (DE3) pLacI *E.coli* cells (Novagen) and plated on LB/Agar supplemented with 50µg/mL Carbenicillin (Melford) and 35µg/mL Chloramphenicol (Sigma-Aldrich). Colonies were used within 48h to inoculate medium-scale (~150mL) cultures consisting of LB media (Millipore) supplemented with 50µg/mL of Carbenicillin and 35µg/mL Chloramphenicol, which were incubated at 37°C overnight while shaking at 160rpm. Overnight pre-cultures were used to inoculate Carbenicillin/Chloramphenicol supplemented LB in volumes up to 20% of the total Erlenmeyer flask volume (normally 400mL in 2L flasks), with an initial Optical Density (O.D.) (λ= 600nm) of ~0.06-0.08. The flasks were shaken at 180-220 rpm at 37°C until O.D. (λ= 600nm) was between 0.6-0.8, cooled to 18°C and induced with 1mM IPTG (Melford) while shaking for a further 12-16 hours. Cultures were harvested by centrifugation at 4500x *g* for 30 minutes before discarding of waste media; the remaining pellets were rinsed once with Phosphate Buffered Saline (PBS) (Sigma-Aldrich) and stored at -20°C. Expression of individual culture batches was

checked by harvesting a 1mL aliquot of each culture and verifying the total cell lysate protein pattern by SDS-PAGE.

### 2.2.2 Bacterial expression of uniformly labelled proteins

Isotopically labelled proteins for solution NMR studies were expressed similarly to unlabelled proteins (Section 2.2.1) with the main exception that cultures were grown in M9 minimal media alternatively containing $^{15}$N-$NH_4Cl$ (Sigma-Aldrich) or $^{13}C_6$-Glucose (Cambridge Isotope Laboratories) as sole nitrogen and carbon sources, respectively. Table 2.7 presents the composition of the M9 minimal media enriched with thiamine and trace metals [113].

**Table 2.7 M9 minimal media components for uniformly labelled bacterial expression**

| Component | Volume (mL) | Final concentration |
|---|---|---|
| Ultra pure water | 857.9 | |
| 100x Trace Element stock | 10.0 | 215.8µM $Fe_2SO_4 \cdot 7H2O$, 58.0µM $MnCl_2$, 39.5µM $CoCl_2 \cdot 6H2O$, 24.3µM $ZnSO_4 \cdot 7H_2O$, 17.63µM $CuCl_2$, 3.2µM $H_3BO_3$, 2.0µM $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 171.1µM $EDTANa_2$ |
| 5mg/mL Thiamine | 6.0 | 0.03mg/mL |
| 1M $CaCl_2$ | 0.1 | 0.1mM |
| 1M $MgSO_4$ | 2.0 | 2.0mM |
| 20%$^{w}/_{v}$ Glucose | 20.0 | 0.4%w/v Glucose |
| 0.25mg/mL $NH_4Cl$ | 4.0 | 1µg/mL $NH_4Cl$ |
| 10x M9 Salts w/o $NH_4Cl$ | 100.0 | 0.478M $Na_2HPO_4$, 0.220M $KH_2PO_4$, 0.086M NaCl pH7.40 |

Freshly transformed colonies were used to inoculate a small-scale (~15mL) LB culture supplemented with 50µg/mL of Carbenicillin and 35µg/mL Chloramphenicol, which was incubated overnight at 37°C while shaking at 160rpm. The pellet collected after briefly centrifuging at 4500x*g* was resuspended in a medium-scale (~150mL) culture of M9 minimal media with the appropriate isotopically enriched nitrogen and carbon sources according to the desired labelled strategy and supplemented with 50µg/mL of Carbenicillin and 35µg/mL Chloramphenicol. The culture was incubated at 30°C overnight while shaking at 180-220rpm, and later used to inoculate a large-scale culture of minimal media for expression. Growth, induction and harvest were as described for unlabelled proteins.

### 2.2.3 Baculovirus-mediated expression of unlabelled proteins in insect cells

Expression of heterologous protein constructs in *Spodoptera frugiderpa* cells (Sf9) required the intermediate formation of a recombinant baculovirus capable of infecting these cultured cells. Generation and amplification of this recombinant virus is simple though time consuming, and a minimum of three sequential transfection/infection and harvest cycles are required before large-scale protein expression can be achieved [114]. Sf9 cells adapted to Sf-900 II SFM (Invitrogen) were cultured in GIBCO™ Sf-900 II media (Invitrogen) by growing them in flasks at 27°C while shaking at 100rpm. Cell density was calculated by manual counting with aid of a haemocytometer and light microscopy; it was kept between $0.5x10^6$cells/mL and $4.0x10^6$cells/mL by splitting and diluting the culture with fresh media every 72h.

#### 2.2.3.1 Recombinant baculovirus preparation

Plasmids derived from pOPIN vectors containing the appropriate baculovirus flanking recombination sequences (*left2* and *ORF1629*) as well as the baculoviral P10 promoter [106] were co-transfected with a linearized modified *Ac*MNPV bacmid [115] into freshly plated Sf9 cells to produce the first titre stock of recombinant baculovirus. First, $1.0x10^6$ cells/mL were diluted in GIBCO Sf-900 II media (Invitrogen), and then 1mL was dispensed into each well of a 6-well plate and allowed to adhere. Simultaneously and per transfection, 2µg of plasmid DNA was mixed with ~200ng of Bsu36I linearized $BAC10_{KO1629}$ bacmid DNA in 100µL of Sf-900 II media (Solution 1), while separately 5µL of FUGENE HD (Promega) was diluted in 100µL of Sf-900 II media (Solution 2). After 5min at room temperature, both were combined, mixed by gently flicking the tube and incubated at room temperature for 20min to form a lipid-DNA complex. This mixture was finally diluted to 1mL with Sf-900 II media and used to replace the media in each well with adherent cells. The plate was incubated at 27°C for 24h in a humidity chamber before adding an additional 1mL of media to each monolayer culture. Co-transfection of Sf9 cells with pOPIN vector plasmids and linearized $BAC10_{KO1629}$ (in which an essential viral gene, ORF1629, has been inactivated by the insertion of a chloramphenicol resistance gene resulting in an unviable baculovirus genome) repairs the bacmid DNA through transposition and inserts the designed ORF construct under the control of the polyhedron promoter [115]. After a total of 6 days, the virus stock

(P1) is harvested by clearing the cell pellet through centrifugation at 500x*g* for 5min and stored at -20°C until further analysis by SDS-PAGE. The supernatant is stored for up to 8 weeks at 4°C protected from light. This method of baculovirus generation results in 100% recombinant viruses and effectively dispenses the need for plaque purification [116].

2.2.3.2 Recombinant baculovirus amplification

For the amplification of virus stocks 50mL of a Sf9 suspension culture at $1.0x10^6$cells/mL – $1.2x10^6$cells/mL were transferred into a 2L flask and inoculated with 700uL of P1 viral stock. The culture was incubated at 27°C protected from light, while shaking at 100rpm for a total of 3 days. The cells were harvested by centrifugation at 1000x*g* for 5min and stored at -20°C and the virus stock supernatant (P2) stored at 4°C protected from light, for up to 4 weeks.

2.2.3.3 Protein expression in insect cells

Protein expression was achieved by infecting 400mL of Sf9 cells between $2.0x10^6$cells/mL – $2.1x10^6$cells/mL, supplemented with 1%$^v/_v$ heat inactivated FBS (Invitrogen) and 1%$^v/_v$ of Penicillin-Streptomycin antibiotic mix (Sigma-Aldrich), with 4mL of P2 baculovirus stock. The culture was incubated at 27°C protected from light, while shaking at 100rpm for a total of 3 days. The cells were harvested by centrifugation at 4500x*g* for 30min, washed once with PBS (Sigma-Aldrich) and stored at -80°C until further processed; while the virus stock supernatant (P3) was discarded.

2.2.4 Protein expression in cultured mammalian cells

Expression of human recombinant proteins in mammalian cells provides the optimum native environment for expression and post-translation modifications (PTMs), though it can be severely limited by poor yields. In this study its use was limited to highly sensitive assays (mass spectrometry, western blot and immunofluorescence) that circumvent this limitation, and allowed the analysis of native PTMs, protein interaction and cellular localisation. Mammalian adherent cells HEK 293T/17 (ATCC® CRL-11268™), HeLa and U2OS (Cancer Research UK Cell Services, both kindly provided by Dr Ross Chapman) were cultured in high glucose DMEM media (Sigma-Aldrich) supplemented

with 10%$^{v}/_{v}$ GIBCO FBS (Invitrogen), 1%$^{v}/_{v}$ MEM non-essential amino acids (Invitrogen), 1%$^{v}/_{v}$ L-Glutamine (Invitrogen) and maintained at 37°C in a humidified incubator with 5% $CO_2$. Cell densities were kept between $0.5x10^5$cells/mL and $4.0x10^6$cells/mL, depending on cell line, by splitting and diluting the culture with fresh media every 48-72h.

For experiments in which DNA damage was chemically induced, the media on transfected cells after 48h was supplemented with 5µM Camptothecin (Sigma-Aldrich) or 100µg/mL Neocarzinostatin (Sigma-Aldrich) and incubated for an additional 8h or 60min, respectively, before harvesting.

2.2.4.1 Large-scale expression in roller bottles

Approximately $5.5x10^7$ adherent cells (HEK 293T or HeLa) were transferred to a 2125cm$^2$ expanded surface roller bottle (Greiner) with 250mL of DMEM media supplemented with 10% FBS, 1%$^{v}/_{v}$ MEM non-essential amino acids, and 1%$^{v}/_{v}$ L-Glutamine, and cells were allowed to expand to confluence for 72h before transfecting according to Aricescu *et al.* 2005 [117]. Briefly, 0.5mg of plasmid DNA was mixed with 1mg of Polyethylenimine at 1mg/mL and pH7.0 (25kDa Branched, Sigma-Aldrich) in 50mL of serum free DMEM media. After incubating at room temperature for 20min, the media in the confluent roller bottles was replaced with 200mL of low serum media (DMEM media supplemented with 2% FBS, 1%$^{v}/_{v}$ MEM non-essential amino acids, and 1%$^{v}/_{v}$ L-Glutamine) and the DNA:PEI complex added gently. After 48h the cells were harvested with the aid of a dissociating agent (0.05% Trypsin-EDTA, Invitrogen), pelleted by centrifugation at 1000x*g* for 5min and 4°C, and immediately used for protein extraction and purification.

2.2.4.2 Small-scale expression in 6-well plates or 100mm dishes

Approximately $7.5x10^5$ cells/well or $4.5x10^6$ cells were transferred to each well in a 6-well plate or a 100mm dish (Grenier), respectively, in the recommended vessel volume for each of DMEM media supplemented with 10% FBS, 1%$^{v}/_{v}$ MEM non-essential amino acids, and 1%$^{v}/_{v}$ L-Glutamine. After 24h and having reached approximately 50% confluence, the cells were transfected with FUGENE HD (Promega), according to the

manufacturer's recommended protocol. Briefly, plasmid DNA (either 3µg per well in a 6-well plate or 10µg per 100mm dish) was mixed with a threefold volume of transfection reagent in Opti-MEM® and after incubating at room temperature for 20min, the media in the vessels was replaced with low serum media (DMEM media supplemented with 5% FBS, 1%$^{v}/_{v}$ MEM non-essential amino acids, and 1%$^{v}/_{v}$ L-Glutamine) and the DNA-lipid complex added gently, drop wise. The cells were allowed to recover and expand for 24h, then split in two with the aid of a dissociating agent (0.05% Trypsin-EDTA, Invitrogen), transferred to new vessels with high serum DMEM and allowed to grow for an additional 24h before harvesting.

2.2.5 Purification of bacterial His-tagged recombinant proteins

Bacterial cell pellets stored at -20°C were thawed on ice and resuspended to ~0.1g/mL in Low Salt HisTrap Binding Buffer (50mM Tris HCl pH8.0, 150mM NaCl, 20mM Imidazole pH8.0) supplemented with Complete EDTA-free Protease Inhibitor Cocktail (Roche), 80U/mL DNAse I (Sigma) and 0.25mg/mL Lysozyme (Sigma). After homogenizing the solution while stirring at 4°C, cells were lysed via cell disruption (2 passes, 26kpsi) and cleared through centrifugation at 4°C and 48000x*g* for 1 h; the supernatant was then filtered through a 0.45µm syringe filter (Millipore) and loaded on a pre-equilibrated 5 mL HisTrap HP column (GE Healthcare) using an ÄKTA Purifier system (GE Healthcare). The column was washed with 5CV of binding buffer and all bound proteins were eluted over a 10CV linear gradient to 500mM Imidazole pH8.0. All fractions were monitored by their absorption at 280nm/260nm, collected and their composition assessed by SDS-PAGE. Fractions containing the protein of interest were pooled and diluted 1:3 with 50mM Tris HCl pH8.0 and loaded on to a pre-equilibrated 5mL HiTrap Q HP ion-exchange column (GE Healthcare) using an ÄKTA Purifier system. Unbound material was washed off with 5CV of Low Salt IEX-Q Binding Buffer (50mM Tris HCl pH8.0, 50mM NaCl), which was followed by a linear 10CV gradient to 1M NaCl for elution of bound proteins. All fractions were collected and their composition determined by SDS-PAGE. Fractions containing the protein of interest were pooled and concentrated using an Amicon centrifugation filter device (Millipore) with molecular weight cut-off between 3kDa and 50kDa, at 4500x*g* and 4°C. The concentrated protein solution was applied at 1mL/min to a HiLoad 16/600 Superdex 75 or Superdex 200

size-exclusion column (GE Healthcare) equilibrated with SEC Phosphate Buffer (20mM $NaH_2PO_4$ pH7.0 and 100mM NaCl). Appropriate fractions, identified by SDS-PAGE analysis, were pooled, concentrated with an Amicon centrifugation filter and stored at -80°C until further use.

### 2.2.6 Purification of bacterial untagged recombinant Ubiquitin

Untagged Ubiquitin was purified as described previously [118] with some modifications. Bacterial cell pellets stored at -20°C were thawed on ice and resuspended to ~0.1g/mL in Tris Buffer (50mM Tris HCl pH7.5, 10mM $MgCl_2$, 0.02% Triton X-100) supplemented with Complete EDTA-free Protease Inhibitor Cocktail (Roche), 80U/mL DNAse I (Sigma) and 0.25mg/mL Lysozyme (Sigma). Cells were lysed via sonication and cleared through centrifugation at 4°C and 48000x*g* for 1 h; the supernatant was carefully transferred to a chilled beaker and stirred while on ice. Slowly, the pH of the solution was lowered by adding 1%$^{v}/_{v}$ Perchloric Acid (Sigma-Aldrich), monitored with pH test paper until reaching pH1.0, and then was stirred at 4°C for 30min. The solution was clarified by centrifugation at 4°C and 48000x*g* for 1h, followed by raising the supernatant pH to approximately pH7.0 by adding NaOH 5M, drop wise, while stirring at 4°C and monitoring with pH test paper. The protein was then transferred to a 3.5kDa MWCO SnakeSkin™ dialysis membrane (Thermo) and equilibrated against 4L of 50mM Tris HCl pH7.5 at 4°C overnight. After dialysis, the sample was cleared by centrifugation at 4000x*g* and 4°C for 15min, filtered through a 0.45µm syringe filter (Millipore) and injected in HiLoad 26/600 Superdex 75 size-exclusion column (GE Healthcare) equilibrated with SEC Phosphate Buffer (20mM $NaH_2PO_4$ pH7.0 and 100mM NaCl). All fractions were monitored by their absorption at 215nm, collected and their composition assessed by SDS-PAGE. Fractions containing the protein of interest were pooled and concentrated using a 3kDa MWCO Amicon filter (Millipore), at 4500x*g* and 4°C. The protein was stored at -80°C until further use.

### 2.2.7 Purification of KAP1 RBCC expressed in insect cells

Baculovirus-infected Sf9 cell pellets stored at -80°C were thawed on ice and resuspended to ~0.1g/mL in High Salt HisTrap Binding Buffer (50mM Tris HCl pH8.0, 500mM NaCl, 10% Glycerol, 1mM TCEP, 20mM Imidazole pH8.0) supplemented with

Protease Inhibitor Cocktail P8849 (Sigma-Aldrich) and 10U/mL Benzonase (Millipore). After homogenizing the solution while stirring at 4°C, cells were lysed via cell disruption (1 pass, 18kpsi) and cleared through centrifugation at 4°C and 48000x*g* for 1h; the supernatant was then filtered through a 0.45µm syringe filter (Millipore) and loaded on a pre-equilibrated 5mL HisTrap HP column (GE Healthcare) using an ÄKTA Purifier system (GE Healthcare). The column was washed with 5CV of binding buffer and all bound proteins were eluted in a step-wise gradient profile to 500mM Imidazole (100%B): 2CV linear gradient from 0% to 10% B, followed by a 5CV at 10%, 3CV linear gradient from 10% to 25% B, followed by a 5CV at 25%, 2CV linear gradient from 25% to 35% B, followed by a 5CV at 35%, finishing with 5CV at 100%B. All fractions were monitored by their absorption at 280nm/260nm, collected and their composition assessed by SDS-PAGE. Fractions containing the protein of interest (at 35%B) were pooled, immediately concentrated using a 50kDa MWCO Amicon filter (Millipore) at 4500x*g* and 4°C, and injected in HiLoad 16/600 Superdex 200 size-exclusion column (GE Healthcare) equilibrated with RBCC SEC Buffer (50mM Tris HCl pH8.0, 500mM NaCl, 10% Glycerol, 1mM TCEP). Fractions were analysed with SDS-PAGE and appropriate samples pooled, concentrated and stored at -80°C until further use.

2.2.8 Purification of SMARCAD1 from mammalian cells for mass spectrometry analysis

Nt-GFP SMARCAD1 expressing cells, alternatively exposed to chemically-induced DNA damage, were processed immediately after harvesting for a GFP pull down using magnetic-agarose GFP-Trap® beads (Chromotek) following the manufacturer's recommended protocol with minor modifications. Pellets were resuspended to ~0.15g/mL with 25mL of ice-cold RIPA Buffer (10mM Tris HCl pH 7.5, 150mM NaCl, 0.5mM EDTA, 0.1% SDS, 1% Triton X-100, 1% Deoxycholate) supplemented with 2.5mM $MgCl_2$, 20mM N-Ethylmethylamine (Sigma-Aldrich), 25U/mL Benzonase (Millipore) and 1x MSSafe Protease and Phosphatase inhibitors (Sigma-Aldrich). After 30min incubation in ice, with periodic pipetting, the solution was cleared by centrifugation at 75000x*g* and 4°C for 20min. The supernatant was diluted 1:5 with ice-cold Dilution Buffer (10mM Tris HCl pH7.5, 150mM NaCl, 0.5mM EDTA, 20mM N-Ethylmethylamine) and combined with 125µL of magnetic-agarose GFP-Trap® slurry, pre-equilibrated with Dilution Buffer. The mixture was incubated overnight (~12h) at

4°C while gently rotating, then the beads were magnetically separated, washed three times with ice-cold Dilution Buffer, and boiled in 2x SDS Denaturing Buffer (100mM Tris HCl pH6.8, 20% Glycerol, 4% SDS, 2% β-Mercaptoethanol).

Similarly, endogenous protein was used to verify the phosphorylation profile of SMARCAD1 when cells were alternatively exposed to chemically induced DNA damage. For this, Protein A magnetic Dynabeads® (Invitrogen) were coupled to Prestige® rabbit anti-SMARCAD1 antibody (Sigma-Aldrich) in place of GFP-Trap® beads to immunoprecipitate endogenous SMARCAD1. Lysate preparation was as described for GFP pull-downs, followed by an overnight (~12h) incubation with 1µg of antibody per 25mL of dilute lysate at 4°C while gently rotating. Next, 10µL of Protein A Dynabeads® per 1µg of antibody were added to each sample and incubated for 2h at 4°C with rotation, then the beads were magnetically separated, washed three times with ice-cold Dilution Buffer, and boiled in 2x SDS Denaturing Buffer.

Immediately following SDS denaturation, the Mass Spectrometry Laboratory at the University of Oxford Target Discovery Institute processed all pull-down samples for proteomic profiling analysis. Briefly, the samples were reduced and alkylated using 5mM DTT and 20mM Iodoacetamide, respectively, and precipitated with two consecutive Methanol/Choloroform washes to remove all SDS [119]. The pellets were then resuspended under denaturing conditions (100mM Tris HCl pH7.8, 6M Urea) and diluted six-fold before adding Trypsin enzyme (Promega) in 1:50 volume ratio. The sample was digested overnight (~12h) at 37°C and 300rpm, after which the enzyme is inactivated by adding Trifluoroacetic Acid (TFA) to a final concentration of 1%$^{v}/_{v}$. Tryptic peptides were then desalted using a Sep-Pak C18 Plus cartridge (Waters) according to the manufacturer's instructions and eluted with 2x 600µL of 0.1%$^{v}/_{v}$ TFA, 65% Acetonitrile. Finally, the samples were lyophilised and resuspended in 20µL of 0.1%$^{v}/_{v}$ TFA, 2% Acetonitrile.

Peptides samples were analysed by using ultraHPLC (Thermo, Dionex Ultimate 3000 UHPLC) coupled to a Thermo Q Exactive hybrid quadrupole-Orbitrap tandem mass spectrometer (70000 full-width half maximum at m/z 400 resolution, Top 15 precursor-ion selection, and fragmentation by higher-energy C-trap dissociation with

normalised collision energy of 28) and eluted through an Acclaim® PepMap™ column (Dionex, nanoRSLC, C18) over 0.1%$^{v}/_{v}$ Formic Acid/5%$^{v}/_{v}$ DMSO and three steps with increasing concentrations of Acetonitrile (isocratic at 2% MeCN for 3min, linear 2-5% MeCN gradient over 6min and linear 5-35% MeCN gradient over 63min) at a flow rate of 250nL/min.

Post-translational modification analysis of raw data was accomplished using the database and *de novo* search program PEAKS version 7 (Bioinformatics Solutions) using 10ppm precursor, 0.5Da fragment mass accuracy and peptide FDR set to 1%. The search allowed for phosphorylation (Ser/Thr/Tyr), deamidation (Asn/Gln) and oxidation (Met) as variable modifications, and carbamidomethylation (Cys) as fixed modification. Novel peptide tags were manually verified by MS/MS.

## 2.3 Solution State NMR

Nuclear magnetic resonance spectroscopy (NMR) is the primary method for structural characterisation used throughout this thesis. This section describes in detail the sample preparation for structural determination data collection, along with the measurement of Residual Dipolar Couplings and various titration experiments used to probe specific protein interactions with amino acid specificity.

### 2.3.1 Sample Preparation

Isotopically-labelled protein samples (Section 2.2.2) were used at concentrations ranging 1.0mM to 2.5mM in SEC Phosphate Buffer (20mM $NaH_2PO_4$ pH7.0 and 100mM NaCl) supplemented with 5%$^{v}/_{v}$ $D_2O$ for deuterium locking. All NMR experiments were performed at 20°C in $D_2O$-matched symmetrical Shigemi microtubes (Shigemi Inc) with a minimum volume of 270µL.

### 2.3.2 Data Collection Parameters

NMR Spectrometers located in the Department of Biochemistry, University of Oxford, raging in $^{1}H$ frequencies between 500MHz and 750MHz were used for data collection. Both Bruker Avance spectrometers at 500MHz, 600MHz and 750MHz fitted with

CryoProbes™ for improved sensitivity or home built 600MHz and 750MHz spectrometers with magnets supplied by Oxford Instruments, UK, and custom triple-resonance pulsed-field gradient probe-heads controlled by GE/Omega Software were used for different experiments, including backbone and sidechain assignment, backbone dynamics, residual dipolar couplings and titrations.

Acquisition parameters for triple resonance experiments used to assign backbone resonances are shown in Table 2.8. Table 2.9 presents acquisition parameters for $^{13}C$-edited experiments used for side-chain assignments. Table 2.10 details acquisition parameters for $^{15}N$-edited experiments, such as those used to determine backbone dynamics or residual dipolar couplings. NMR data was processed with NMRPipe [120], followed by analysis in CcpNmr Analysis [121].

### 2.3.3 Residual Dipolar Coupling Measurement

Residual Dipolar Coupling (RDC) determination was performed using the change in $^{1}J_{NH}$ scalar coupling values found in an IPAP (in-phase/antiphase) experiment of a partially aligned sample, relative to an isotropic sample [122]. An anisotropic sample of 3%$^{w}/_{v}$ $C_{12}E_6$/n-Hexanol was prepared using the method of Rückert and Otting [123] and was found to give good, stable splitting of the deuterium signal. Briefly, a 15%$^{v}/_{v}$ stock of $C_{12}E_6$ (Hexaethylene glycol monododecyl ether, Sigma-Aldrich) in SEC Phosphate Buffer (20mM $NaH_2PO_4$ pH7.0 and 100mM NaCl) supplemented with 10%$^{v}/_{v}$ $D_2O$ was mixed with n-Hexanol (Sigma-Aldrich) slowly and while vortexing, until the solution turned from cloudy to opalescent. This stock was diluted 1:5 with $^{13}C$,$^{15}N$ CUE1 protein (Table 2.6) to a final concentration of ~1mM in 3% $C_{12}E_6$/n-Hexanol for the aligned sample, while an equivalent dilution with SEC Phosphate Buffer was used for the isotropic sample.

### 2.3.4 NMR Titrations

Titration experiments were performed to probe the amino acid specific contribution to various protein-protein interactions, namely SMARCAD1 SIM and CUE1 binding domains to their anticipated ligands SUMO1 and Ubiquitin, respectively. For this, changes in the chemical shift and peak intensity of a uniformly $^{15}N$-labelled target

protein at a fixed concentration are recorded upon the addition of incremental quantities of an unlabelled "ligand". Titration experiments were carried out using two samples representing the initial and end points of the titration curve: one with 100µM of $^{15}$N-labelled target protein and a the second with the final, highest concentration of the unlabelled ligand in addition to 100µM of $^{15}$N-labelled target protein. The combination of these two solutions in appropriate ratios produced a full ligand concentration range while maintaining a constant labelled (observable) protein concentration. As with all NMR experiments, titrations were performed in SEC Phosphate Buffer (20mM $NaH_2PO_4$ pH7.0 and 100mM NaCl) supplemented with 5%$^{v}/_{v}$ $D_2O$, however the pH of the mixed intermediary solutions was checked, and whenever necessary corrected, to pH7.0. Details for each titration pairing, including ligand concentration range and origin, are detailed in Table 2.11.

### 2.3.5 Structure Calculation

The ARIA (Ambiguous Restraints for Iterative Assignment) procedure [124] is widely used for structure determination from unassigned or ambiguously assigned NOESY spectra. ARIA is based on an iterative protocol to derive distance constraints from NOE cross-peaks that will be applied to calculate an ensemble of structures. The newly created bundle of molecular conformations then serves to evaluate the consistency of the assignments allowing for the re-calibration of distance restraints, and the progressive reduction of the ambiguity present in the initial NOE assignments [125].

The most important idea that underlies the ARIA methodology is the concept of *Ambiguous Distance Restraints* (ADR) [126]. When using ADR, each NOESY cross peak is treated as the additive combination of the all potential signals from each of its multiple assignments. This means that the NOE intensity is assumed to depend all the individual proton-proton distances that contribute to the signal. During structure calculation, in a similar fashion as for unambiguous distance constraints, the distance in the molecular coordinates is restrained to a radius within an interval between lower and upper distance bounds rather than a single value, with the energy term for NOE-based distance restraints resulting in an harmonic potential with a "flat-bottom", that is, zero-energy contribution between the lower and upper limits. In this way, all

potential assignments for the cross peak intensity are defined to be within a flexible range of each other and this is then evaluated in conjunction with all other restraints, leading to a specific proton-proton NOE to be confirmed or violated.

For this work, manually assigned NOESY data was classified as either ambiguous or not depending on the number of assignment possibilities that match a given resonance. These restraints were treated through 11 rounds of ARIA processing, each with 8 steps of iterative distance restraint calibration, and consisted of:

- A total of 217 NOE-derived inter-residue distances from $^{13}$C-edited NOESY spectra for aliphatic and aromatic protons, most of which correspond to methyl-methyl proton NOEs
- 30 α-helical Hydrogen bonds defined based on $^{3}J_{HNHA}$ coupling constants experimentally determined with an HNHA experiment
- 82 TALOS+ predicted dihedral angles for those residues determined to be structured based on Heteronuclear NOE data
- 34 RDCs calculated based on amide proton signal slitting as described above.

For each, after the calibrated and verified NOE distance restraints were merged to avoid overrepresentation of redundant distance data, a new structure ensemble was calculated with the program CNS through a Cartesian Space molecular dynamics simulated annealing (MDSA) protocol, followed by a refinement step in a shell of explicit solvent (water molecules). This consists in a short molecular dynamics simulation with a complete force field (including electrostatics and Lennard-Jones potential), which has been shown to significantly improve structure quality [125, 127].

**Table 2.8 Acquisition Parameters for Triple Resonance NMR Experiments used in Backbone Assignment**

| Experiment | Construct | Spectrometer | $^{1}$H Frequency (MHz) | Sweep Width (Hz) | | | Number of Complex Points | | | Reference Chemical Shift (ppm) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $^{1}$H | $^{15}$N | $^{13}$C / [$^{1}$H] | $^{1}$H | $^{15}$N | $^{13}$C / [$^{1}$H] | $^{1}$H | $^{15}$N | $^{13}$C / [$^{1}$H] |
| CBCANH | pOPINE/WT eCUE1 CtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | 9025.271 | 512 | 32 | 128 | 4.80 | 118.25 | 42.0 |
| CBCA(CO)NH | pOPINE/WT eCUE1 CtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | 9025.271 | 512 | 32 | 128 | 4.80 | 118.25 | 42.0 |
| HNCO | pOPINE/WT eCUE1 CtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | 2272.727 | 512 | 32 | 128 | 4.80 | 118.25 | 177.75 |
| HBHA(CO)NH | pOPINE/WT eCUE1 CtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | [5617.978] | 512 | 32 | [128] | 4.80 | 118.25 | [4.793] |
| HBHA(CO)NH | pOPINE/WT CUE1 WCtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | [5617.978] | 512 | 32 | [128] | 4.80 | 118.25 | [4.80] |
| (H)CC(CO)NH | pOPINE/WT eCUE1 CtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | 10526.316 | 512 | 32 | 128 | 4.80 | 118.25 | 41.9 |

**Table 2.9 Acquisition Parameters for $^{13}$C-edited NMR Experiments**

| Experiment | Construct | Spectrometer | $^{1}$H Frequency (MHz) | Sweep Width (Hz) | | | Number of Complex Points | | | Reference Chemical Shift (ppm) | | | Mixing Times (ms) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $^{1}$H | $^{1}$H | $^{13}$C | $^{1}$H | $^{1}$H | $^{13}$C | $^{1}$H | $^{1}$H | $^{13}$C | |
| $^{13}$C HSQC CT | pOPINE/WT eCUE1 WCtHis | Bruker 600 | 600.133 | 5617.978 | - | 10570.825 | 1024 | - | 256 | 4.8 | - | 43.792 | - |
| $^{13}$C HSQC CT | pOPINE/WT CUE1 WCtHis | Bruker 600 | 600.133 | 5617.978 | - | 10570.825 | 1024 | - | 256 | 4.8 | - | 40.3 | - |
| $^{13}$C NOESY | pOPINE/WT CUE1 WCtHis | Bruker 600 | 600.133 | 5617.978 | 5617.978 | 9803.922 | 512 | 128 | 64 | 4.8 | 4.8 | 40.3 | 75 |

**Table 2.10 Acquisition Parameters for $^{15}$N-edited NMR Experiments**

| Experiment | Construct | Spectrometer | $^{1}$H Frequency (MHz) | Sweep Width (Hz) | | | Number of Complex Points | | | Reference Chemical Shift (ppm) | | | Mixing Times (ms) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $^{1}$H | $^{15}$N | $^{1}$H | $^{1}$H | $^{15}$N | $^{1}$H | $^{1}$H | $^{15}$N | $^{1}$H | |
| $^{15}$N HetNOE | pOPINE/WT eCUE1 CtHis | Bruker 500 | 500.012 | 5882.354 | 1241.311 | - | 1024 | 256 | - | 4.80 | 118.25 | - | - |
| $^{15}$N HSQC | pOPINE/WT eCUE1 CtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | - | 1024 | 256 | - | 4.80 | 118.25 | - | - |
| $^{15}$N HSQC | pOPINE/WT eCUE1 CtHis | Omega 600 | 600.151 | 6666.666 | 1501.500 | - | 1024 | 256 | - | 4.80 | 118.05 | - | - |
| $^{15}$N TOCSY | pOPINE/WT eCUE1 CtHis | Omega 600 | 600.151 | 6666.666 | 1501.500 | 6666.666 | 512 | 32 | 128 | 4.80 | 118.098 | 4.8 | 50 |
| $^{15}$N NOESY | pOPINE/WT CUE1 WCtHis | Omega 600 | 600.151 | 6666.666 | 1501.500 | 6666.666 | 512 | 32 | 128 | 4.80 | 118.098 | 4.6 | 150 |
| $^{15}$N HSQC | pOPINE/WT CUE1 WCtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | - | 1024 | 256 | - | 4.80 | 118.25 | - | - |
| $^{15}$N NOESY | pOPINE/WT CUE1 WCtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | 5617.978 | 1024 | 256 | 128 | 4.80 | 118.25 | 4.8 | 75 |
| $^{15}$N HSQC | pOPINE/WT CUE1 WCtHis | Bruker 750 | 749.914 | 8620.690 | 1901.141 | - | 1024 | 256 | - | 4.80 | 118.25 | - | - |
| $^{15}$N HSQC IPAP | pOPINE/WT CUE1 WCtHis | Bruker 750 | 749.914 | 6502.250 | 1865.670 | - | 1024 | 1024 | - | 4.80 | 118.25 | - | - |
| $^{15}$N JHAHN | pOPINE/WT eCUE1 CtHis | Bruker 750 | 749.914 | 8620.690 | 1901.141 | 8620.690 | 512 | 32 | 128 | 4.80 | 118.25 | 4.8 | - |
| H(CCCO)NH | pOPINE/WT eCUE1 CtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | 5617.978 | 512 | 32 | 128 | 4.80 | 118.25 | 4.8 | - |
| $^{15}$N HSQC | pOPINE/WT eCUE1 WCtHis | Bruker 750 | 749.914 | 8620.690 | 1901.141 | - | 1024 | 256 | - | 4.80 | 118.25 | - | - |
| $^{15}$N HSQC | pOPINE/F169A eCUE1 WCtHis | Bruker 600 | 600.133 | 5617.978 | 1490.313 | - | 1024 | 256 | - | 4.80 | 118.25 | - | - |
| $^{15}$N HSQC | pOPINE/S3E eCUE1 WCtHis | Bruker 750 | 749.914 | 8620.690 | 1901.141 | - | 1024 | 256 | - | 4.80 | 118.25 | - | - |
| $^{15}$N HSQC | pOPINE/WT edCUE CtHis | Bruker 750 | 749.914 | 10000.000 | 2127.660 | - | 1024 | 256 | - | 4.80 | 118.25 | - | - |

**Table 2.11 Protein/Ligand pairings of NMR titration experiments**

| $^{15}$N-labelled protein | $^{15}$N-Protein | Unlabelled ligand | Ligand, µM | Notes |
|---|---|---|---|---|
| pOPINE/WT eCUE1 CtHis | 58 µM | NtHis SUMO-1 | 0, 202, 406, 608, 811 | Ligand from Enzo Life Sciences |
| pET15b/SUMO1-GG NtHis | 300 µM | SIM WT (Genscript) | 0, 100.0, 356, 501, 729, 1017, 1309, 1746, 2133, 2414, 2591, 2900, 3000 | Ligand from Genscript, WT SMARCAD1 SIM region peptide: EKTFNKDTVIIVSEDSEDEESQGLPY |
| pET15b/SUMO1-GG NtHis | 300 µM | SIM pS124 (Genscript) | 0, 103, 200, 304, 402, 508, 757, 1006, 1552, 2046, 2588, 2889, 3000 | Ligand from Genscript, pS124 SMARCAD1 SIM region peptide: EKTFNKDTVIIVpSEDSEDEESQGLPY |
| pET15b/SUMO1-GG NtHis | 300 µM | SIM pS127 (Genscript) | 0, 103, 302, 600, 845 1174, 1500, 2000, 2500, 3000 | Ligand from Genscript, pS127 SMARCAD1 SIM region peptide: EKTFNKDTVIIVSEDpSEDEESQGLPY |
| pET15b/SUMO1-GG NtHis | 300 µM | SIM pS132 (Genscript) | 0, 104, 353, 600, 851, 1102, 1452, 1803, 2081, 2353, 2623, 2886, 3000 | Ligand from Genscript, pS132 SMARCAD1 SIM region peptide: EKTFNKDTVIIVSEDSEDEEpSQGLPY |
| pOPINE/WT eCUE1 WCtHis | 100 µM | pET15b/SUMO1-GG NtHis | 0, 100, 306, 505, 750, 1008, 1190, 1474, 1749, 1965, 2300 | Ligand expressed and purified in-house |
| pOPINE/WT eCUE1 WCtHis | 100 µM | pET15b/SUMO1-GG NtHis | 0, 101, 316, 533, 806, 1008, 1241, 1506, 2155, 2730, 3399, 3977, 4483, 5024 | Ligand expressed and purified in-house |

**Table 2.11 (Cont.) Protein/Ligand pairings of NMR titration experiments**

| $^{15}$N-labelled protein | $^{15}$N-Protein | Unlabelled ligand | Ligand, µM | Notes |
|---|---|---|---|---|
| pOPINE/WT eCUE1 CtHis | 101 µM | pET3a/Ubiquitin | 0, 108, 502, 1009, 1500, 1256, 1794, 2251, 2869, 3897, 4404, 4906 | Ligand expressed and purified in-house |
| pOPINE/WT eCUE1 WCtHis | 100 µM | pET3a/Ubiquitin | 0, 101, 307, 519, 782, 1008, 1255, 1512, 2010, 2513, 3016, 3513, 4018, 4507, 5026 | Ligand expressed and purified in-house |
| pOPINE/WT eCUE1 WCtHis | 102 µM | pET3a/Ubiquitin, pET15b/SUMO1-GG NtHis | 0, 110, 332, 564, 1007, 1524, 2004, 2506, 2984, 3955, 4365, 4906 | Ligand expressed and purified in-house. Constant [pET15b/SUMO1-GG NtHis]: 1mM |
| pOPINE/WT eCUE1 WCtHis | 100 µM | pET15b/SUMO1-GG NtHis, pET3a/Ubiquitin | 0, 104, 302, 505, 805, 1100, 1504, 1950, 2400, 2850, 3299, 3702, 4003, 4507 | Ligand expressed and purified in-house. Constant [pET3a/Ubiquitin]: 2mM |
| pOPINE/WT edCUE CtHis | 100 µM | pOPINF/Ubiquitin Nt MAHx6G | 0, 305, 620, 937, 1176, 1626, 2066, 2360, 2866, 3371, 3978, 4614 | Ligand expressed and purified in-house |
| pOPINE/S3E eCUE1 CtHis | 100 µM | pET15b/SUMO1-GG NtHis | 0.0, 200, 467, 802, 1112, 1512, 1952, 2471, 2989, 3435, 3845, 4515, 4988 | Ligand expressed and purified in-house |

## 2.4 Protein X-ray Crystallography

Sitting drop, vapour diffusion crystallisation experiments were set up in 96-well plates (Greiner) using a nanoliter Robot (Cartesian Technologies) to screen conditions using commercially available solution in a matrix format [128, 129]. Crystallisation screens arranged for automated 96-well experiments were dispensed in CrystalQuick 96 Well crystallisation plates (Greiner Bio-one) using a Hydra96 micro-dispenser liquid handling robot (Matrix Technologies Ltd, UK). 100nL of reservoir solution was mixed with 100nL of purified 80µM KAP1 NHis/RBCC combined with SMARCAD1 CUE1/CHis in equimolar proportions, each aliquoted using a Cartesian Technologies MicrosysMIC400 nanodrop crystallization robot (Genomic Solutions). Finally, all reservoirs were adjusted to ~500mM NaCl using a 5.5M stock, and plates sealed using VIEWseal adhesive covers (Greiner Bio-one). Plates were stored either at 4°C or 20°C, in a Rock Imager 1000 (Formulatrix) or a HomeBase automated storage vault (The Automation Partnership, Ltd, UK), respectively, with automatic and periodic image acquisition.

Crystals appeared within 48h at 20°C in three conditions: 50mM MES pH6.5, 80mM Magnesium Acetate, 30%$^{v}/_{v}$ PEG4000 (JBScreen Nuc-Pro solution 2/A6, Jena Bioscience GmbH); 100mM Sodium Citrate pH5.6, 20%$^{v}/_{v}$ 2-Propanol, 20%$^{v}/_{v}$ PEG4000 (ProPlex™ solution 31, Molecular Dimensions); and 100mM HEPES pH7.5, 20m Magnesium Chloride, 22%$^{v}/_{v}$ Polyacrylic acid 5100 (Index™ solution 59, Hampton Research). Crystals grew to full size within 5 days and were afterwards promptly cryoprotected in mother liquor supplemented with 25% Ethlyleneglycol or 30% Glycerol, and directly frozen in liquid nitrogen prior to data collection. Diffraction data was collected at 100K in beamline I24 (100% transmission, exposure time of 0.25s, oscillation rate of 0.5 degrees and wavelength of 0.9686Å) at Diamond Light Source (Didcot, UK).

Initial hits were optimized using an additive screen (with and without crystal seeding) or a three-row step dilution screen. For the former, the Hampton Research Additive Screen formatted for 96-well plate experiments, was dispensed in 100nL

aliquots using Cartesian Technologies MicrosysMIC400 nanodrop crystallization robot after mixing 100nL of purified 80µM KAP1 NHis/RBCC combined with SMARCAD1 CUE1/CHis in equimolar proportions and 100nL of the selected crystallization condition reservoir. In optimisation trials where crystal seeding was employed, crystals previously used for data collection were recycled and processed Seed Beads™ (Hampton Research) in the appropriate crystallisation buffer adjusted to 500mM NaCl; each crystallisation drop was supplemented with 50nL of crystal seeds [130]. For three-row dilution step optimisation, a serial dilution of the selected crystallization condition in 3% steps is prepared along one row of a 96-well plate creating a reservoir range spanning 100% to 67%. Additionally, the protein-reservoir volume ratio was modified to include 1:1, 1:2 and 2:1 ratios, thus creating a 3x12 matrix of unique crystallisation conditions, otherwise prepared as described previously. Optimisation plates were stored at 20°C, in a HomeBase automated storage vault (The Automation Partnership, Ltd, UK), with automatic and periodic image acquisition.

## 2.5 Protein-Protein Interaction Characterisation

Throughout this work, a diverse combination of biochemical and biophysical methods were used to probe and define protein-protein interaction and post-translational modifications mediating the functional role of SMARCAD1 with the goal of identifying and describing the molecular mechanism behind it.

### 2.5.1 Co-Immunoprecipitation of GFP-SMARCAD1 mutants upon DNA damage

Plasmids encoding an Nt-GFP fusion protein of wild type SMARCAD1 or functional mutants thereof (F169A or SIM8A, Table 2.6) were co-transfected into HEK293T cells with a $SUMO1^{Q92R}$ encoding plasmid (pOPINF/SUMO1 Q92R Nt MAHx6G), as described above (Section 2.2.4.2). Cells having been exposed to chemically induced DNA damage were harvested (along with untreated controls) for co-immuniprecipitation after 48h expression. Lysates were prepared under mild conditions with NP40/Benzonase coIP Buffer (20mM Tris pH7.5, 40mM NaCl, 2mM $MgCl_2$, 0.5% NP40/Igepal) supplemented with 150U/mL Benzonase (Millipore),

20mM N-Ethylmethylamine (Sigma-Aldrich), and 1x MSSafe Protease and Phosphatase inhibitors (Sigma-Aldrich). A total of 500µL of coIP Buffer per 100mm plate was used to resuspend cells after scraping them off, followed by 10min incubation in ice, with periodic pipetting, and then 30min incubation at 4°C while gently rotating after adjusting the salt concentration to 450mM NaCl. Cell debris was cleared by centrifugation at 18000x*g* for 15min at 4°C, and the supernatant diluted 1:3 with coIP Dilution Buffer (25mM Tris HCl pH7.5, 15% Glycerol, 1.5mM TCEP, 20mM N-Ethylmethylamine) before mixing with 25µL of magnetic-agarose GFP-Trap® slurry (Chromotek) per sample. After incubating overnight (~12h) at 4°C while gently rotating, the beads were magnetically separated, washed three times with ice-cold coIP Dilution Buffer, and boiled in 2x SDS Denaturing Buffer (100mM Tris HCl pH6.8, 20% Glycerol, 4% SDS, 2% β-Mercaptoethanol).

Immediately following SDS denaturation, the Mass Spectrometry Laboratory at the University of Oxford Target Discovery Institute processed all pull-down samples for proteomic profiling analysis. Briefly, the samples were reduced and alkylated using 5mM DTT and 20mM Iodoacetamide, respectively, and precipitated with two consecutive Methanol/Choloroform washes to remove all SDS [119]. The pellets were then resuspended under denaturing conditions (100mM Tris HCl pH7.8, 6M Urea) and diluted six-fold before adding Trypsin enzyme (Promega) in 1:50 volume ratio. The sample was digested overnight (~12h) at 37°C and 300rpm, after which the enzyme is inactivated by adding Trifluoroacetic Acid (TFA) to a final concentration of 1%$^{v}/_{v}$. Tryptic peptides were then desalted using a Sep-Pak C18 Plus cartridge (Waters) according to the manufacturer's instructions and eluted with 2x 600µL of 0.1%$^{v}/_{v}$ TFA, 65% Acetonitrile. Finally, the samples were lyophilised and resuspended in 20µL of 0.1%$^{v}/_{v}$ TFA, 2% Acetonitrile. Peptides samples were split and analysed as technical duplicates using nanoLC-MS/MS (Waters, nAcquity, 75µm × 250 mm, 1.7µm particle size) coupled to a Thermo Orbitrap Elite tandem mass spectrometer (120000 full-width half maximum at m/z 400 resolution, Top 20 precursor-ion selection, and fragmentation by collision-induced dissociation) and eluted over a linear 8-35% Acetonitrile gradient over 60min at a flow rate of 250nL/min.

Data analysis was carried out with Progenesis LC-MS software (NonLinear), using default settings and normalized against SMARCAD1 based on high abundance peptides only. Relative abundance profiles between mutant and wild-type pull-downs, grouped based on neocarzinostatin treatment, were analysed as volcano plots using Microsoft Excel. Outliers with *t*-test *p*-value p≤0.05 were cross referenced with post-translational modification analysis of raw data using the database and *de novo* search program PEAKS version 7 (Bioinformatics Solutions) using 10ppm precursor, 0.5Da fragment mass accuracy and peptide FDR set to 1%. For PTM analysis, in addition with standard modifications, both fixed (cysteine carbamidomethylation) and variable (Met oxidation, Asn/Gln deamidation, Ser/Thr/Tyr phosphorylation), a series of distinct mass additions to lysine residues left by tryptic digestion of ubiquitylation moieties (+114.0429 m/z for ($N_\zeta$)-GlyGly, +383.2281 m/z for ($N_\zeta$)-GlyGlyArgLeu) and $SUMO1^{Q92R}$ sumoylation (+472.1918 m/z for ($N_\zeta$)-GlyGlyThrGlnGlu) are included as variable modifications, and novel modification peptide tags are manually verified by MS/MS.

2.5.2 Denaturing pull-down of ubiquitinated proteins using IMAC

Plasmid DNA encoding an N-terminal hexahisitidine Ubiquitin fusion protein (pOPINF/Ubiquitin Nt MAHx6G) was transfected into HEK293T cells as detailed above (2.2.4.2). Cells were harvested for lysis after 48h of expression and resuspended under denaturing conditions following previously published protocols ([131, 132] with minor modifications. Four equivalent 100mm plates (~$1.8x10^7$ cells each) were lysed in 1mL Urea Lysis Buffer (50mM Tris HCl pH8.0, 50mM $Na_2HPO_4$ pH7.0, 8M Urea pH8.0, 300mM NaCl, 0.5% NP40/Igepal) supplemented with 1x MSSafe Protease and Phosphatase inhibitors (Sigma-Aldrich) by scraping the cell monolayer off and combined into a single sample volume. Brief sonication pulses in ice ensured lysis and DNA shearing, which was followed by centrifugation at 16000x*g* and 4°C for 15min to clear cell debris. The soluble supernatant was mixed with pre-conditioned TALON® magnetic beads (Clontech, 150µL/100mm plate), and incubated for ~2h at 4°C while gently rotating. Afterwards, the beads were magnetically separated and washed once with Urea Lysis Buffer and twice with Urea Wash Buffer (Urea Lysis Buffer supplemented with 10mM Imidazole pH8.0). In a clean vial, the

beads were eluted with 100µL of Urea Elution Buffer (Urea Lysis Buffer supplemented with 200mM Imidazole pH8.0). Lysate and TALON® beads eluate protein concentration was determined using a Bradford assay kit, with BSA as standard in Urea Lysis Buffer. Normalized SDS-PAGE samples were prepared in denaturing 1x SDS Buffer (50mM Tris HCl pH6.8, 10% Glycerol, 2% SDS, 1% β-Mercaptoethanol), and analysed by western blot using rabbit polyclonal anti-KAP1 (ab10484, abcam), mouse monoclonal anti-Ubiquitin Ubi-1 (ab7254, abcam), and mouse monoclonal anti-PentaHis (33460, QIAgen) as primary antibodies. Enhanced chemiluminescence detection using goat anti-rabbit IgG-HRP (sc2004, Santa Cruz Biotechnology), goat anti-mouse IgG-HRP (sc2005, Santa Cruz Biotechnology) and Clarity™ ECL western substrate (BioRad) was carried out in a ChemiDoc™ MP imaging system (BioRad).

### 2.5.3 Size Exclusion Chromatography / Multiple Angle Laser Light Scattering

SEC-MALLS experiments were performed on a Shimadzu HPLC system coupled to a size exclusion Superdex S200 10/300 column (GE Healthcare) equilibrated in MALLS buffer (50mM Tris pH8.0, 500mM NaCl, 5% Glycerol and 1mM TCEP) at 0.4 ml/min. The column was followed in-line by a Dawn Heleos 8 light-scattering detector (Wyatt Technologies, Germany) and an Optilab rEX refractive index monitor (Wyatt Technologies) with laser light scattering detected at 662 nm wavelength at eight scattering angles between 20.6° and 149.1°. Molecular mass calculations were performed using ASTRA 6.1 (Wyatt Technologies), using the formalism developed by B.H. Zimm of the Rayleigh-Debye-Gans model [133], assuming a specific refractive index increment (dn/dc**)** of 0.185 ml/g.

### 2.5.4 Isothermal Calorimetry

Isothermal titration calorimetry was carried out on a MicroCal PEAQ-ITC isothermal titration calorimeter (Malvern) using KAP1 RBCC NtHis as the analyte and SMARCAD1 WT CUE1 WCtHis as titrant. Titrations were carried out at 25°C using a reference power of 2cal/s at a stirrer speed of 500rpm. Isotherms were analysed and fitted in MicroCal PEAQ-ITC Analysis Software using a “one site” binding model, with N, ΔH and Kd as fitting parameters. Concentrations of both proteins were measured

by absorption at 280 nm in Phosphate Buffer (20mM $NaH_2PO_4$ pH7.0 and 100mM NaCl) supplemented with 1mM β-Mercaptoethanol.

2.5.5 Co-Immunoprecipitation of GFP-KAP1 truncation mutants

Plasmid DNA encoding C-terminal GFP KAP1 truncation mutant fusion proteins (Table 2.6) were transfected into HEK293T cells as described above (2.2.4.2). Cell were harvested for protein co-immunoprecipitation after 48h of expression under mild lysis conditions with NP40/Benzonase coIP Buffer (20mM Tris pH7.5, 40mM NaCl, 2mM $MgCl_2$, 0.5% NP40/Igepal) supplemented with 150U/mL Benzonase (Millipore), 20mM N-Ethylmethylamine (Sigma-Aldrich), and 1x MSSafe Protease and Phosphatase inhibitors (Sigma-Aldrich). Two equivalent 100mm plates per construct (~$1.8x10^7$ cells each) were scraped off and resuspended with 500µL of coIP Buffer each. Once combined, the lysate volume was incubated for 10min on ice with periodic pipetting, and then for 30min at 4°C while gently rotating once the salt concentration had been adjusted to 450mM NaCl. Cell debris was cleared by centrifugation at 16000x*g* for 15min at 4°C, and the supernatant diluted 1:3 with coIP Dilution Buffer (25mM Tris HCl pH7.5, 15% Glycerol, 1.5mM TCEP, 20mM N-Ethylmethylamine) before mixing with 50µL of magnetic-agarose GFP-Trap® slurry (Chromotek) per sample. After incubating overnight (~12h) at 4°C while gently rotating, the beads were magnetically separated, washed three times with ice-cold coIP Dilution Buffer, and boiled in 2x SDS Denaturing Buffer (100mM Tris HCl pH6.8, 20% Glycerol, 4% SDS, 2% β-Mercaptoethanol). Soluble whole cell lysate protein concentration was determined using a Bradford assay kit, with BSA as standard in NP40/Benzonase coIP Buffer. Normalized SDS-PAGE samples were prepared in denaturing 1x SDS Buffer (50mM Tris HCl pH6.8, 10% Glycerol, 2% SDS, 1% β-Mercaptoethanol), and along with the eluted samples denatured off the GFP-Trap® magnetic agarose beads, were analysed by western blot using goat polyclonal anti-GFP (ab5450, abcam), rabbit polyclonal anti-SMARCAD1 (HPA016737, Sigma-Aldrich), rabbit polyclonal anti-PCNA (07-2162, Millipore) and rabbit polyclonal anti-Histone3 (ab1791, abcam) as primary antibodies. Enhanced chemiluminescence detection using donkey anti-goat IgG-HRP (sc2020, Santa Cruz Biotechnology), goat anti-rabbit IgG-HRP (sc2004, Santa Cruz Biotechnology), goat anti-mouse IgG-HRP

(sc2005, Santa Cruz Biotechnology) and Clarity™ ECL western substrate (BioRad) was carried out in a ChemiDoc™ MP imaging system (BioRad).

### 2.5.6 Proximity ligation assay coupled immunolocalisation

In order to determine the functional relevance of each ancillary domain in SMARCAD1 and to better characterise the cellular role it might exert upon DNA damage, immunofluorescence microscopy was used coupled to a proximity ligation assay (Duolink® *In Situ* Mouse/Rabbit Red PLA, Sigma-Aldrich) so as to distinguish from the total nuclear population the fraction of SMARCAD1 coupled to KAP1. The assay was carried out as previously described [134], closely following the manufacturer's recommended protocol. Minor modifications were introduced to include additional antibodies against markers of consequential cellular processes such as DNA damage and heterochromatin mediated transcriptional repression.

#### 2.5.6.1 Mammalian cell culture

For PLA analysis of un-transfected samples, approximately $7.5x10^5$ adherent cells per well (HeLa and U2OS, Section 2.2.4) were seeded on 6-well plates over previously autoclaved 13mm glass coverslips (VWR) for overnight 50% confluency in complete DMEM10 (high glucose DMEM media supplemented with 10% FBS, 1%$^v/_v$ MEM non-essential amino acids, and 1%$^v/_v$ L-Glutamine). Growth media was discarded and cells were gently washed once with PBS (GIBCO, Invitrogen), followed by a 20min fixation at room temperature with 2%$^w/_v$ formaldehyde. After fixation, cells were washed once with PBS, permeabilized for 5min at room temperature with 0.1%$^v/_v$ Triton X-100 and blocked with fresh and 0.45µm filtered 3%$^w/_v$ BSA in PBS/0.1%$^v/_v$ Tween 20. After blocking, cells were washed twice and were then ready for traditional immunostaining or PLA.

Alternatively, cells transfected with plasmid DNA encoding NtGFP fusion proteins of full length SMARCAD1 or functional mutants thereof (F169A or SIM8A) as described above (section 2.2.4.2) were prepared in the same fashion but only after 48h of expression at 37°C and 5% $CO_2$ in a humidity chamber.

2.5.6.2 Immunofluorescence staining colocalisation

Primary antibodies were prepared in blocking solution (3%$^{w}/_{v}$ BSA in PBS/0.1%$^{v}/_{v}$ Tween 20) at appropriate dilutions following manufacturer's recommendations or experimentally optimised concentrations (Table 2.13). For multiple epitope staining, all primary antibodies were combined and incubated simultaneously. Thus, blocked and permeabilized cells were incubated with primary antibodies for 2h at room temperature, washed twice with PBS/0.1%$^{v}/_{v}$ Tween 20, followed by Alexa Fluor™ coupled secondary antibodies (Molecular Probes) at appropriate dilutions for 1h at room temperature, protected from light. Finally, after washing twice, cells were counter-stained with 0.2µg/mL DAPI in PBS for 30-60min at room temperature and protected from light. Coverslips were mounted using ProLong® Gold antifade reagent (Molecular Probes) and stored in the dark at 4°C.

2.5.6.3 Proximity Ligation Assay colocalisation

Primary rabbit polyclonal anti-SMARCAD1 (HPA016737, Sigma-Aldrich) and mouse monoclonal anti-KAP1 20C1 (ab22553, abcam) were mixed at appropriate dilutions in blocking solution (3%$^{w}/_{v}$ BSA in PBS/0.1%$^{v}/_{v}$ Tween 20) and incubated 2h at room temperature, followed by two washes with PBS/0.1%$^{v}/_{v}$ Tween 20. PLA secondary antibody MINUS and PLUS probes were diluted fivefold in the provided Diluent Solution and incubated with the coverslips for 1h at 37°C in a pre-heated humidity chamber. PLA detection was performed as recommended by the manufacturer. Briefly, coverslips are incubated with Ligation Buffer for 30min, followed by Amplification Buffer for 100min; both incubations were at 37°C in a pre-heated humidity chamber and all wash steps at room temperature while protecting the samples from light.

Profiting from the single fluorophore detection of a protein complex conferred by the PLA, it is possible to detect additional epitopes and analyse, relative to the protein pair of interest, various colocalisation events that may address its functional role. To circumvent the limitation established by the original host species in the SMARCAD1/KAP1 PLA complex, direct labelling of primary antibodies was carried out using Zenon® technology, which employs anti-rabbit Fc monovalent goat Fab

fragments conjugated to an Alexa Fluor® dye. Rabbit monoclonal anti-γH2A.X 20E3 (9718, Cell Signalling) and rabbit polyclonal anti-H3K9me3 (07-442, Millipore) were labelled with one of three labelled Fab fragments available in the Zenon® Tricolor Rabbit IgG Labelling Kit 1 (Molecular Probes), following the manufacturer's recommendations. After the final washes with PLA recommended buffers A (10mM Tris HCl pH 7.4, 150mM NaCl 0.05%$^{w}/_{v}$ Tween 20) and B (200mM Tris HCl pH 7.5, 100mM NaCl), coverslips are incubated for 30min with directly labelled primary antibodies against γH2A.X and H3K9me3 at appropriate dilutions, followed by a brief 5min fixation step with 2%$^{w}/_{v}$ formaldehyde. Finally, after washing twice with PBS, cells were counter-stained with 0.2µg/mL DAPI in PBS for 30-60min at room temperature and protected from light. Coverslips were mounted using ProLong® Gold antifade reagent (Molecular Probes) and stored in the dark at 4°C.

2.5.6.4 siRNA mediated knockdown of SMARCAD1 and KAP1

Designed as negative controls for the proximity ligation assay, HeLa and U2OS cells were transfected with siGENOME siRNA SMARTpool mixtures against SMARCAD1 (M-013801-01, Dharmacon) and KAP1 (M-005046-01, Dharmacon) at 25nM and 50nM final concentration, respectively, and using DharmaFECT1 transfection reagent following the manufacturer's recommendations. Briefly, $4x10^5$ cells per well were seeded on 6-well plates for ~70% confluency after 96h in 2mL each of DMEM media supplemented with 10% FBS, 1%$^{v}/_{v}$ MEM non-essential amino acids, and 1%$^{v}/_{v}$ L-Glutamine, and transfected in consecutive days in order to create an optimized incubation period prior to harvesting for PLA analysis. Thus, siGENOME siRNA SMARTpool for either SMARCAD1 or KAP1 (50pmol or 100pmol per well, respectively) was mixed with 4µL of DharmaFECT1 transfection reagent in Opti-MEM® and after incubating at room temperature for 20min, gently added to freshly replaced media in the well. The cells were allowed to recover and expand for 24h, 48h or 72h before harvesting. Coverslips were treated and stained as described above.

2.5.6.5 Antibodies used for immunofluorescence

Antibodies used for immunofluorescence staining and PLA are listed in Table 2.12.

**Table 2.12 Antibodies used for immunofluorescence and PLA analysis**

| Antibody | Host | Target Epitope | Dilution | Source |
|---|---|---|---|---|
| IgG polyclonal | Rabbit | SMARCAD1 | 1:600 | Sigma-Aldrich |
| IgG monoclonal, 20C1 | Mouse | KAP1 | 1:500 | Abcam |
| IgG monoclonal, | Rabbit | gH2A.X | 1:250 | Cell Signalling |
| IgG polyclonal | Rabbit | H3K9me3 | 1:250 | Millipore |
| IgG polyclonal | Rabbit | GFP | 1:1000 | Invitrogen |
| IgG1 polyclonal Alexa Fluor 488 | Goat | Mouse IgG(H+L) | 1:500 | Molecular Probes |
| IgG polyclonal Alexa Fluor 546 | Goat | Rabbit IgG(H+L) | 1:500 | Molecular Probes |
| IgG polyclonal Alexa Fluor 555 | Goat | Rabbit IgG(H+L) | 1:500 | Molecular Probes |
| F(ab')2 polyclonal Alexa Fluor 633 | Goat | Mouse IgG(H+L) | 1:500 | Molecular Probes |
| Zenon Fab Alexa Fluor 488 | Goat | Rabbit Fc | - | Molecular Probes |
| Zenon Fab Alexa Fluor 555 | Goat | Rabbit Fc | - | Molecular Probes |
| Zenon Fab Alexa Fluor 647 | Goat | Rabbit Fc | - | Molecular Probes |
| IgG PLA Probe PLUS | Donkey | Rabbit IgG(H+L) | 1:5 | Sigma-Aldrich |
| IgG PLA Probe MINUS | Donkey | Mouse IgG(H+L) | 1:5 | Sigma-Aldrich |

Chapter 3

# Structural elucidation of SMARCAD1 binding domains

Chromatin remodeler SMARCAD1 is one of the few known members of the SNF2/SWI2 family of remodeling ATPases to be targeted and adopt a functional role in the heterochromatic regions of the genome. Specifically, SMARCAD1 is a member of a multi-subunit complex (along with KAP1, HP1, HDAC1/2 and GP9a) involved in re-establishing heterochromatic organisation and propagating epigenetic post-translational modifications (PTMs) on histone proteins in nucleosome-packed, newly replicated DNA [50, 51]. Furthermore, recent evidence suggesting a potential role in Homologous Recombination during the DNA damage response [59, 61, 62] has made this non-canonical chromatin remodeler a new, attractive target to improve our understanding of vital processes such as replication, transcription and repair (all of which require extensive remodeling of the chromatin landscape) in the context of densely packed, transcriptionally repressive heterochromatin. SMARCAD1 is unique in its possession of tandem ubiquitin binding CUE domains as the sole known protein-protein interaction domains, and potentially responsible for targeting it to its various functional contexts. It would therefore be useful to fully understand its structure as a gateway to probe its biological role. Previous work in the Mancini lab has derived a solution structure for SMARCAD1 CUE2 domain, but equivalent work with CUE1 failed to produce an unambiguous structural model, likely due to multiple technical drawbacks in the construct used [135]. This work resulted in potentially equivalent families of structures with ambiguously defined structural elements that might be the result of legitimate dynamics or simply to poorly resolved data, which called for further analysis and finally the unequivocal elucidation of its structure.

This chapter describes the design of a new set of constructs only encoding the first CUE domain in the tandem arrangement, and a novel SUMO-interacting motif (SIM), newly identified through bioinformatics sequence alignment (Section 3.1). These were expressed and purified for structural studies using NMR (Section 3.2). Following the confirmation of an active SIM by chemical shift perturbation studies with recombinant SUMO1 (Section 3.3), a combination of 3D double- and triple-resonance experiments resulted in 99% of the construct being assigned (Section 3.4). Backbone dynamics indicated that each binding domain is structurally independent (Section 3.5), with CUE1 as a structured helical domain and the SIM as part of a highly flexible N-terminal extension. Structure calculation with further restraints, such as Residual Dipolar Couplings and long-range NOEs (Section 3.6), produced a well-defined triple helix bundle (Section 3.7) with none of the ambiguity in secondary structure or helix packing previously described and consistent with other members of the CUE domain family (Section 3.8), such as yeast CUE2 or human TOLL-interacting protein (TOLLIP).

## 3.1 Sequence analysis and construct design

The SMARCAD1 sequence has been previously reported to possess two ubiquitin-binding CUE domains near its N-terminus [136], each a triple helix bundle according to secondary structure prediction software (Figure 3.1). The reported boundary domains are in agreement with disorder prediction software and indicate a sharp transition at the end of CUE1 helix 3, followed by a highly flexible linker connecting it to CUE2.

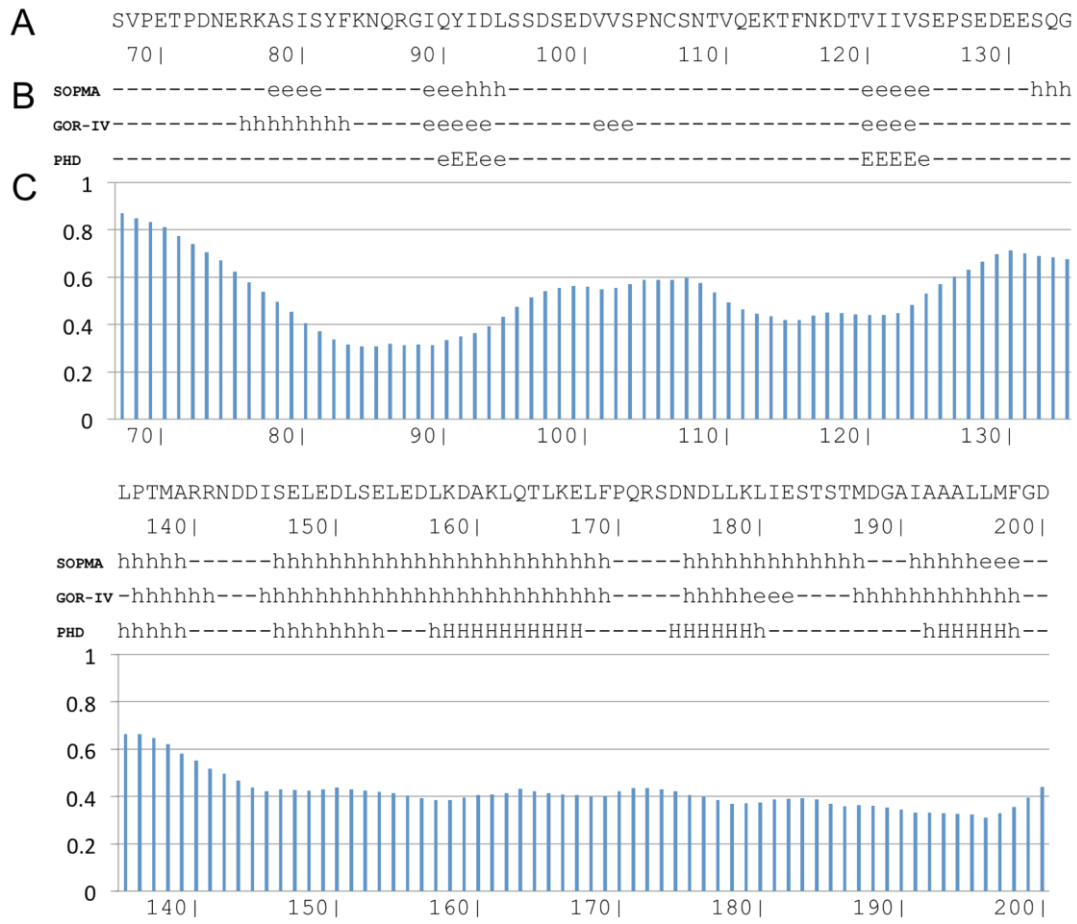



**Figure 3.1 Secondary structure and disorder prediction plot.** Bioinformatics analysis of SMARCAD1 primary sequence to assess its propensity for disorder and defined secondary structure elements. A) SMARCAD1 primary sequence between amino acids 67 and 207 (Uniprot Q9H4L7-2), B) Secondary structure prediction from individual algorithms (SOPMA, GOR-IV, and PHD) in the meta-server ProteinCCD (NKI, The Netherlands) [109], C) Probability of disorder as predicted by RONN (Regional Order Neural Network, University of Oxford, UK) [110], using ≥0.5 as a threshold for disorder.

Sequence alignment of these regions does indicate a poor degree of sequence conservation among orthologs, even in the key amino acid residues in the consensus domain sequence well established to directly mediate the interaction with ubiquitin through the hydrophobic patch centred on Ile44, as presented in Figure 3.2. Two motifs are of particular importance: an (M/L)FP sequence at the C-terminal end of the first helix, and a di-leucine motif in the third helix [92, 94].

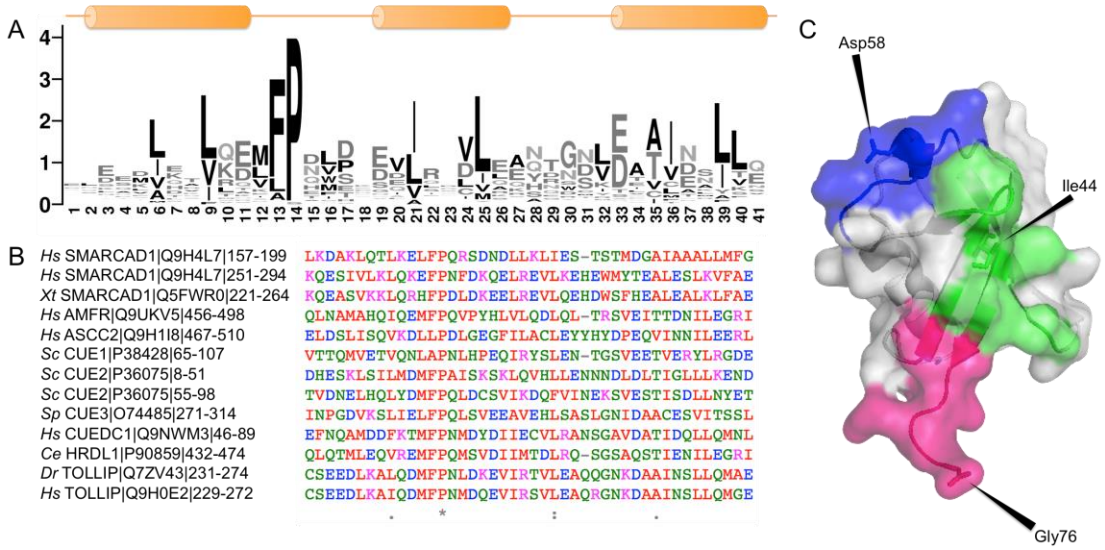



**Figure 3.2 Sequence homology of SMARCAD1 CUE domains.** A) Consensus sequence for CUE domains, with the secondary structure of the N-terminal CUE domain (CUE1) from *Saccharomyces cerevisiae* protein CUE2 [95] shown above it as reference, B) Sequence alignment of a selection of CUE domains identified using the PROSITE database profile (PDOC51140 [137]) to highlight the broad range of proteins bearing a CUE domain, with different functions and across species. Protein names and species are listed along with the sequence position of the CUE domain: Sc (budding yeast, *Saccharomyces cerevisiae*), Sp (fission yeast, *Schizosaccharomyces pombe*), Dr (zebrafish, *Danio* rerio), Ce (roundworm, *Caenorhabditis elegans*), Xt (western clawed frog, *Xenopus tropicalis*), and Hs (human, *Homo sapiens*), C) Major recognition patches on ubiquitin (PDB: 1UBQ), with the hydrophobic patch centred on Ile44 in green, the polar patch centred on Asp58 in blue, and the di-glycine patch near the C-terminal Gly76 in pink [94, 138].

Furthermore, bioinformatics sequence mining led to the identification of a small sequence stretch consistent with SUMO Interacting Motif (SIM): a hydrophobic centre, rich in non-polar amino acids (valine, leucine and isoleucine), with neighbouring acidic patches on either side [74]. This contrasting sequence is meant to interact with the cleft surface between α-helix 1 and β-strand 2 in SUMO, inducing an elongated strand secondary structure upon binding and extending the β-sheet motif in SUMO [72, 102]. As presented in Figure 3.3, this potential SIM appears to

only be present in vertebrates, with a slightly modified sequence in *Gallus gallus*, and the human sequence conserved throughout mammals.

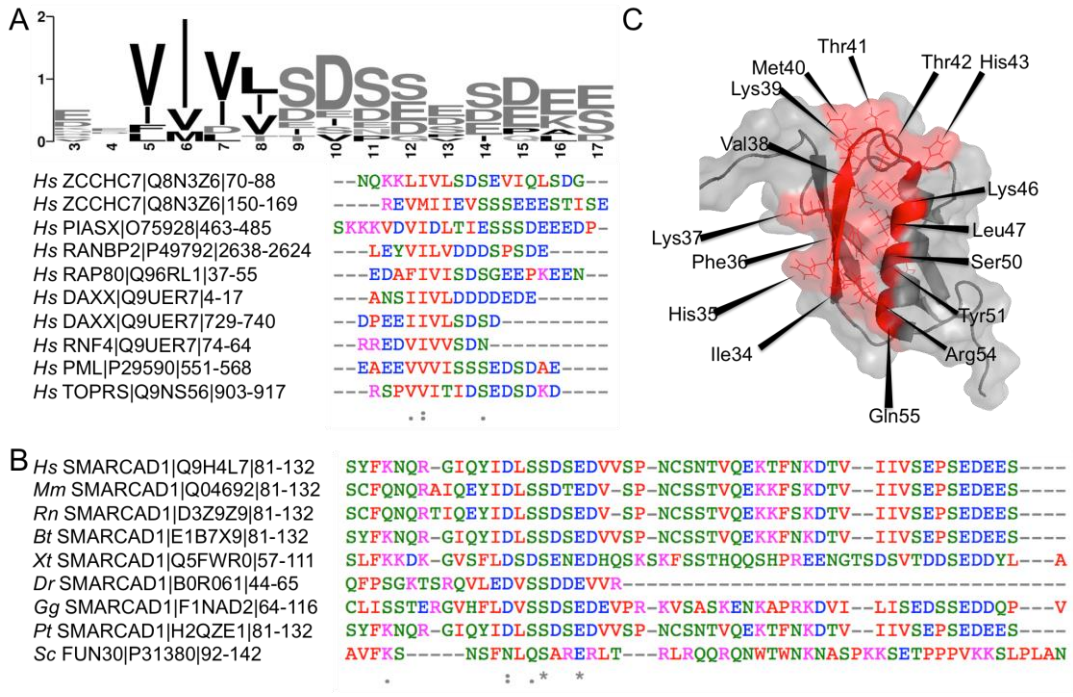



**Figure 3.3 Sequence homology of SMARCAD1 SIM domains.** A) Sequence alignment of various SUMO Interacting Motif (SIM) containing proteins, with a generalised consensus sequenced generated with WebLOGO [139], highlighting the hydrophobic centre and acidic neighbouring patches that characterise this binding motif [74], B) Sequence alignment of SMARCAD1 orthologs, selected from the curated Swiss-Prot/UniprotKB database. Protein names, Uniprot identifiers and species are listed along with the relative sequence position of the SIM: Sc (budding yeast, *Saccharomyces cerevisiae*), Dr (zebrafish, *Danio* rerio), Xt (clawed frog, *Xenopus tropicalis*), Gg (domesticated chicken, *Gallus gallus*), Mm (house mouse, *Mus musculus*), Rn (brown rat, *Rattus norvegicus*), Bt (domesticated cattle, *Bos taurus*), Pt (common chimpanzee, *Pan troglodytes*), and Hs (human, *Homo sapiens*), C) Major recognition surface in SUMO1 (PDB: 2KQS), with main residues involved in SIM contacts along the cleft between α-helix 1 and β-strand 2.

Based on this combined data, considering secondary structure and disorder predictions along with consensus sequence motifs, a series of domain boundaries were selected for recombinant constructs to be heterologously expressed in *E. coli* as N-terminal or C-terminal hexahistidine fusion proteins. Ligation-independent cloning was performed using vectors from the OPPF pOPIN suite [106], as described in Chapter 2. Initial expression screening using recombinant constructs with alternatively longer N-terminal extensions beyond the CUE1 domain in SMARCAD1 were tested in *E. coli* Rosetta2 (DE3) pLacI cells and assessed by SDS-PAGE. As evidenced in Figure 3.4, the C-terminal hexahistidine SIM/CUE1 construct spanning the SMARCAD1 sequence between amino acids 109-206 offered the best expression yields while limiting the length of the flexible N-terminal extension. This is an important factor for consideration, given that unstructured sequences would inevitably overcrowd any NMR spectrum derived from the construct, limiting the amount and quality of potential information drawn from it. Because the native sequence of this construct does not include any tryptophan residues, which would limit the ease and precision of UV-Visible spectrometry mediated concentration determination, the C-terminal tag was modified to include one and this alternative quantifiable construct was exclusively used for all NMR titration experiments.

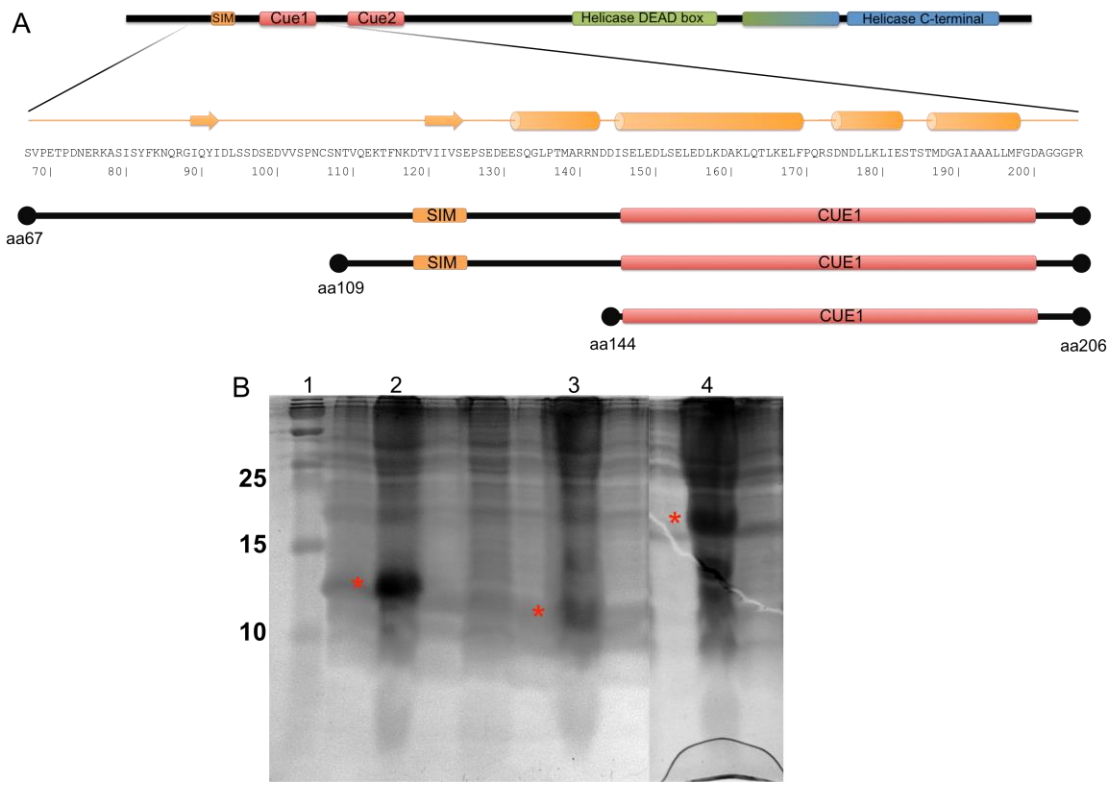



**Figure 3.4 Design and initial expression screening for recombinant SMARCAD1 SIM/CUE1 constructs.** A) Schematic representation of SIM/CUE1 constructs with alternative N-terminal domain boundaries resulting in flexible sequence extensions of varying length attached to the CUE1 domain. The averaged secondary structure predicted in Figure 3.1 is included, along with the primary sequence, as reference for the relative composition of each construct. B) SDS-PAGE analysis of whole cell lysates prepared by lysing bacterial pellets, after 16h of induced expression of each construct, in denaturing SDS loading buffer. Lane 1: Molecular weight ladder, Lane 2: SIM/CUE1 CtHis aa109-206, Lane 3: CUE1 CtHis aa 144-206, Lane 4: SIM/CUE1 CtHis aa67-206.

## 3.2 Protein Expression and Purification

Constructs were expressed as uniformly labelled $^{15}N$- or $^{15}N$, $^{13}C_6$-labelled proteins in *E. coli*, as described in detail in Chapter 2. These proteins were purified by a sequential affinity and anion exchange chromatography strategy, in order to remove all non-specific contaminants and minor degradation products. The construct is very stable, does not seem affected by variations in salt or imidazole concentration, and is

very soluble, even to relatively high concentrations (between 3-5mM). Following ion exchange chromatography, fractions containing the protein were identified by SDS-PAGE analysis, pooled, and further purified through Size Exclusion Chromatography using a Superdex S75 16/600 SEC column (GE Healthcare), pre-equilibrated with Phosphate Buffer (20mM $NaH_2PO_4$ pH7.0, 100mM NaCl). PAGE analyses of each chromatographic step, along with UV-Visible traces of eluate absorption at 280nm are presented in Figure 3.5. Given the high level of purity, fractions containing the recombinant SMARCAD1 SIM/CUE1 CtHis construct were combined and concentrated to approximately 1.0-2.5mM, with aliquots stored at -80°C. Yields ranged between 8mg-12mg of uniformly labelled protein per litre of cultured cells harvested after expression.

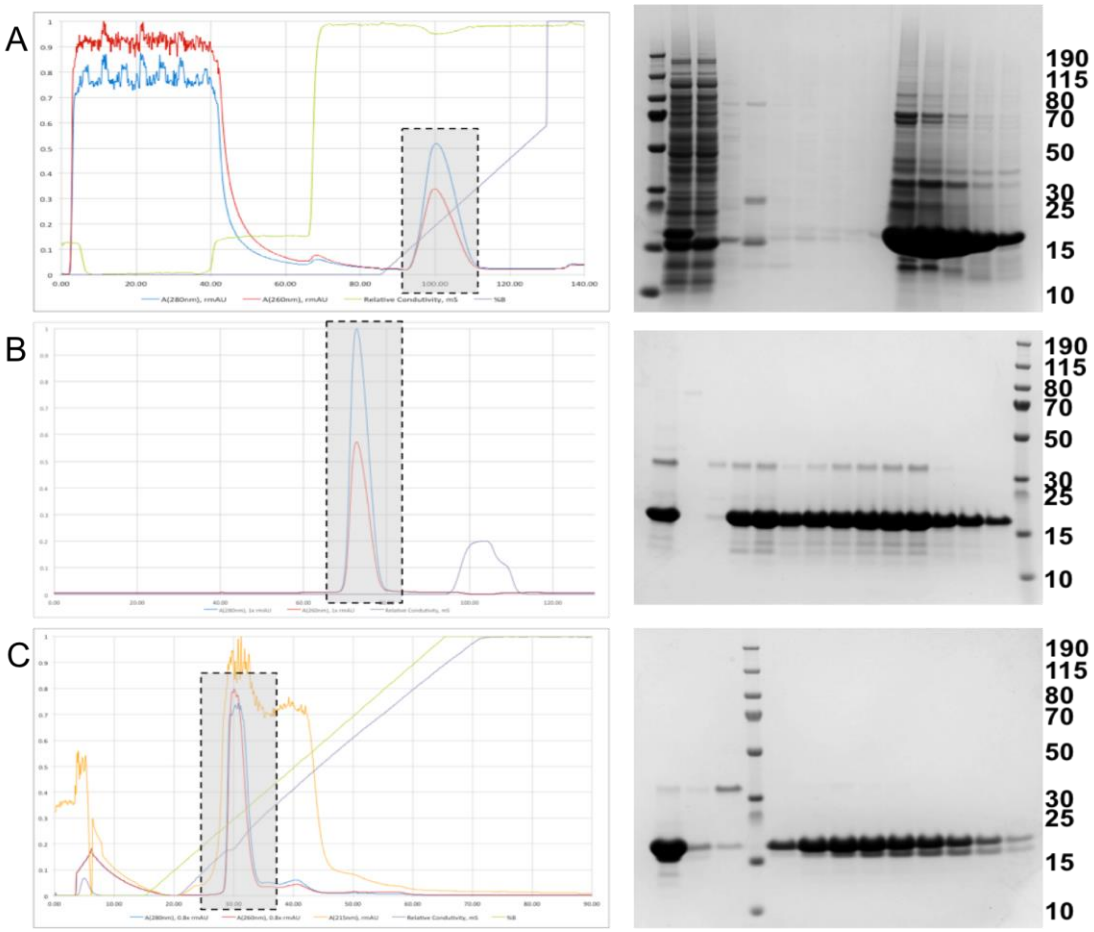



**Figure 3.5 Strategy for recombinant SMARCAD1 SIM/CUE1 CtHis purification (Legend on next page).**

**Figure 3.5 Strategy for recombinant SMARCAD1 SIM/CUE1 CtHis purification (previous page).** SDS-PAGE analysis and chromatographic traces of elute absorption at 280nm and 260nm. Samples analysed by PAGE are marked on the chromatogram: A) Nickel affinity purification, B) Anion exchange chromatography, and C) Size exclusion chromatography. SDS-PAGE gels were stained with Coomassie Brilliant Blue; molecular weight ladder is Page Ruler (Thermo Fisher). Chromatogram traces are: absorption at 280nm in blue, absorption at 260nm in red, ionic strength in brown, elution buffer composition in green.

## 3.3 Confirmation of SMARCAD1 SUMO interacting motif

In order to experimentally verify the bioinformatics identification of a potential novel SIM in SMARCAD1, an NMR titration with unlabelled SUMO1 as ligand was performed. The range of ligand concentrations was relatively small, limited by the availability of commercial SUMO1 (Enzo, human recombinant hexahistidine tagged). The procedure of a titration is described in detail in Chapter 2; briefly two solutions of uniformly labelled SMARCAD1 SIM/CUE1 CtHis ($^{15}$N eCUE1) were prepared with one of them containing the highest concentration point of unlabelled SUMO1 as ligand, and mixed in fixed ratios maintaining the labelled protein concentration constant throughout the experiment. 2D heteronuclear single quantum correlation ($^{1}$H,$^{15}$N HSQC) spectra were collected for each titration point and are presented in Figure 3.6. It becomes immediately evident that, indeed, a limited number of resonances corresponding to amide protons in the structure are selectively significantly perturbed with increasing SUMO1 concentration.

Furthermore, it was necessary to identify these perturbed peaks to confirm that they in fact correspond to the SIM sequence. Because the construct is fairly small at approximately 12kDa and the resonances in question were well dispersed, with little overlap in the $^{1}$H,$^{15}$N HSQC spectrum, it was possible to employ a sequential assignment strategy that relies only on $^{15}$N-edited 3D NMR methods ($^{15}$N-edited NOESY and TOCSY) and establishes connections between neighbouring residues with through-space NOE data [140]. Use of chemical shift reference tables for

unstructured amino acids relative to through-bond TOCSY data, allowed the identification of individual spin systems, that is, the set of resonances belonging to a single residue in the construct [141, 142]. In combination with through-space NOESY data, it was possible to arrange these spin systems in the correct sequence [141, 143], confirming the perturbed peaks in the SUMO1 NMR titrations as the predicted SIM between amino acids 119-125 in SMARCAD1. Figure 3.7 presents a diagram for the magnetisation transfer in each experiment and the overlaid spectra in strips corresponding to the SIM sequence.

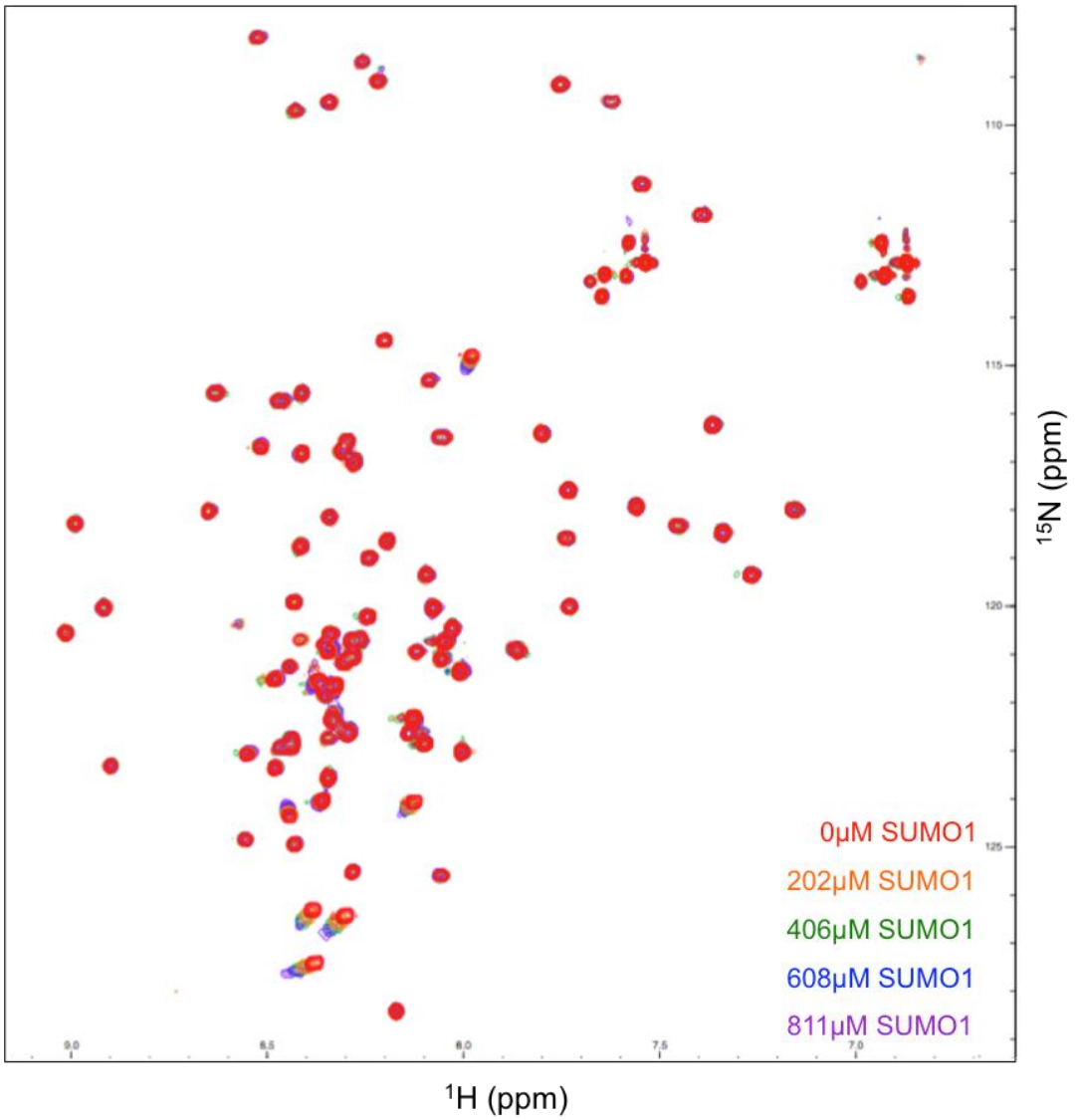



**Figure 3.6 Confirmation of novel SIM in SMARCAD1.** Overlay of five $^{1}$H,$^{15}$N HSQC spectra for $^{15}$N-labelled SMARCAD1 SIM/CUE1 CtHis (eCUE1, 58μM, 11.83kDa), each with different colour-coded SUMO1 concentrations.

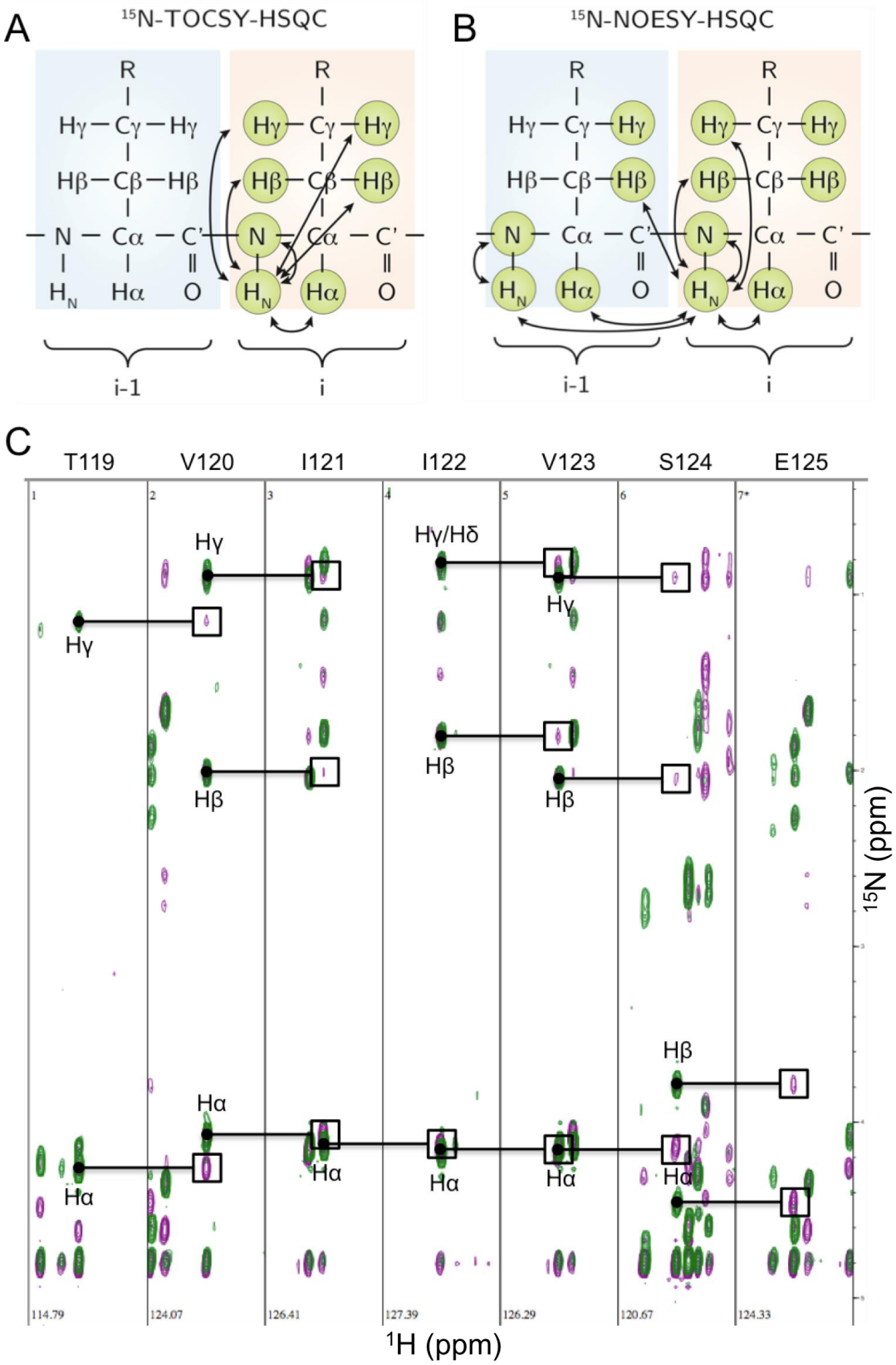



**Figure 3.7 Identification of a novel SIM in SMARCAD1. (Legend on next page).**

**Figure 3.7 (Previous page) Identification of a novel SIM in SMARCAD1.** A) Coherence transfer diagram for $^{15}$N-TOCSY-HSQC: magnetisation is transferred between $^{1}$H and then filtered via the HSQC (allowing $^{15}$N labelling) and finally the $^{1}H^{N}$, where it is detected, B) Coherence transfer diagram for $^{15}$N-NOESY-HSQC: magnetisation is exchanged between all $^{1}$H in close proximity using the NOE before being filtered via the HSQC to $^{1}H^{N}$ for detection. C) Sequential backbone assignment through the simultaneous use of $^{15}$N-edited TOCSY (purple) and $^{15}$N-edited NOESY (green) spectra, where the former indicates sidechain proton resonances for a given residue, and the latter indicates all proton resonances in close proximity to the amide proton of a given residue, both its own and from its neighbours. Circles (•) indicate $^{1}$H peaks from residue *i*-1, and open squares are peaks from residues *i*, while numbers at the bottom of each strip indicate the $^{15}N^{H}$ chemical shift (ppm).

## 3.4 Backbone and sidechain sequence assignment

Complete sequence assignment of the SIM/CUE1 construct was necessary for subsequent structural and functional characterisation using solution NMR. In this work, a triple resonance assignment strategy was employed to relate the amide backbone to aliphatic backbone and sidechain resonances of $^{13}$C,$^{15}$N double-labelled samples, with through-bond scalar coupling resulting in the selective identification of all resonances in a given residue as well as the preceding one [140]. Thus, both identification of spin systems and sequential connectivity may be determined. As with $^{15}$N-edited TOCSY data, spin systems are identified based on comparison with chemical shifts tables for unstructured amino acids.

Partial backbone and sidechain assignments were derived from a pair of experiments in which aliphatic carbon chemical shift information for Cα and Cβ is collected either simultaneously for a given residue *i* and its preceding neighbour *i*-1 (CBCANH), or alternatively only for residue *i*-1 (CBCA[CO]NH), selectively gating the magnetisation through the carbonyl carbon with the $^{15}N^{H}$-$^{13}C_{CO}$ scalar coupling [141, 143-146]. Combining both spectra allowed the differentiation between resonances for each residue and was used to arrange these spin systems in the correct sequence.

Moreover, these experiments were crucial in the unequivocal identification of spin systems, particularly for overcrowded regions of the $^{1}H,^{15}N$ HSQC spectrum where overlapping residues could be differentiated based on their unique neighbouring amino acids. Because carbon chemical shifts can vary significantly depending on their chemical environment, and with characteristic shift patterns, the use of Cα and Cβ for pairs of consecutive residues in the sequence reduced the range of possible spin system types and eventually led to unambiguous assignments. An important exception was the case of highly degenerate regions in the primary sequence where additional experiments were required to ensure connections and determine residue type. Figure 3.8 presents a diagram of the magnetisation transfer for each experiment and an example of Cα and Cβ assignments for a continuous stretch of the sequence.

**Figure 3.8 (Next page) Assignment and sequence connection of Cα and Cβ in SMARCAD1 SIM/CUE1.** A) Coherence transfer diagram for CBCANH: magnetisation is transferred between $^{1}H\alpha$ and $^{1}H\beta$ to $^{13}C\alpha$ and $^{13}C\beta$, respectively; then through the backbone onto the $^{15}N^{H}$ and to $^{1}H^{N}$ for detection. Because magnetisation can be transferred to $^{15}N^{H}$ *i* from $^{13}C\alpha$ *i* and $^{13}C\alpha$ *i*-1, for each amide group there are two Cα and Cβ peaks visible, B) Coherence transfer diagram for CBCA[CO]NH: as with CBCANH, magnetisation is transferred between $^{1}H\alpha$ and $^{1}H\beta$ to $^{13}C\alpha$ and $^{13}C\beta$, respectively, however it is then transferred through the backbone first to the $^{13}C_{CO}$ *i*-1 then selectively to $^{15}N^{H}$ *i* and to $^{1}H^{N}$ *i* for detection, C) Overlay of CBCANH (orange) and CBCA[CO]NH (blue/green) experiments for a continuous stretch of residues in the primary sequence of SMARCAD1 SIM/CUE1. Circles (•) indicate $^{13}C$ peaks from residue *i*, and open squares are peaks from residues *i*-1, while numbers at the bottom of each strip indicate the $^{15}N^{H}$ chemical shift (ppm).

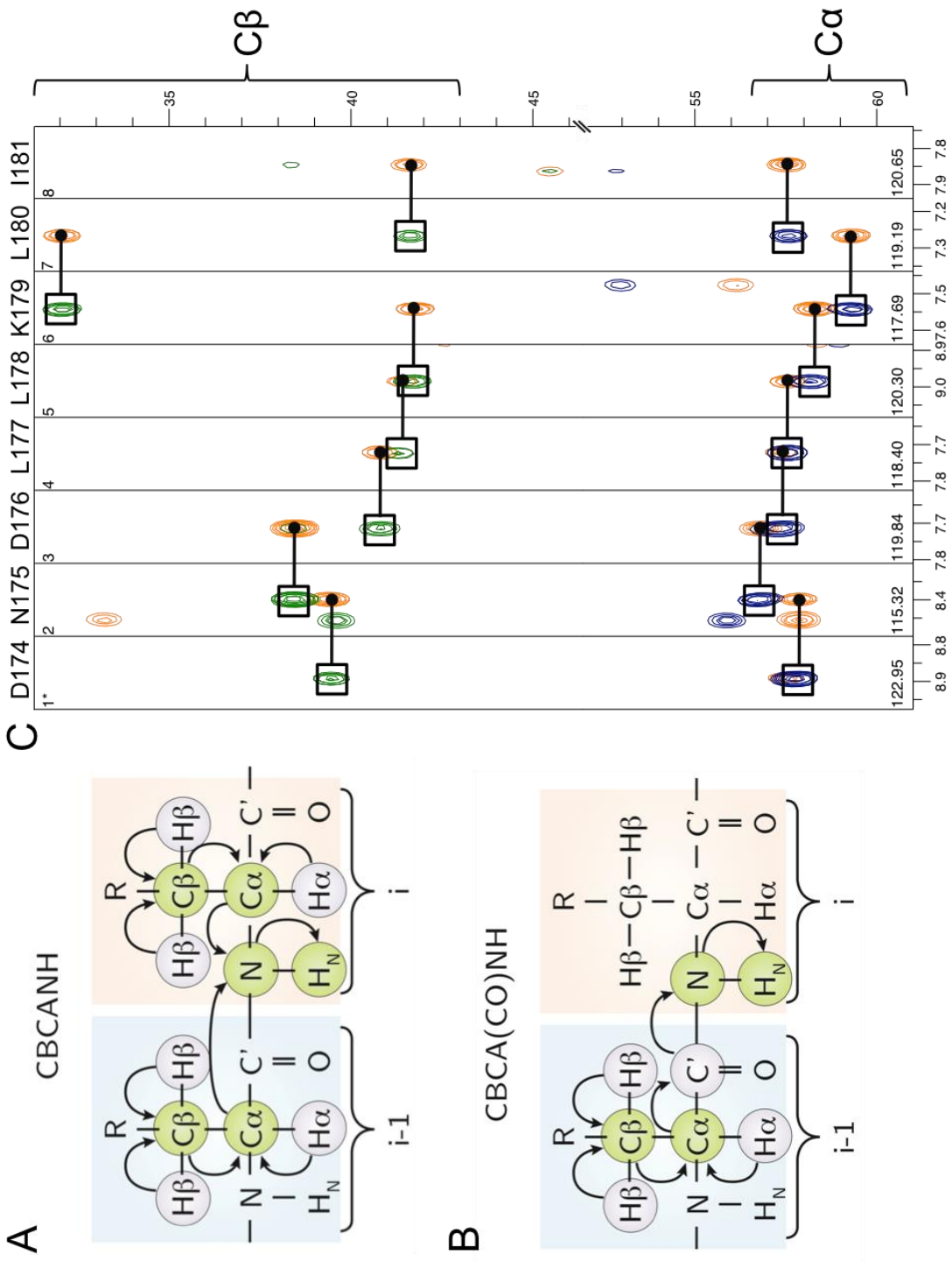



**Figure 3.8 Assignment and sequence connection of Cα and Cβ in SMARCAD1 SIM/CUE1 (Legend on previous page).**

Assignment of aliphatic protons was achieved with the HBHA[CO]NH experiment, in which magnetisation is transferred from the aliphatic protons from the preceding residue *i*-1 through the backbone to the carbonyl carbon and from there to the amide proton of residue *i* without evolving chemical shift on any of the carbon nuclei, but rather only on the aliphatic protons and amide nitrogen and proton nuclei, creating a three dimensional spectrum with two $^{1}H$ dimensions ($H^{C}$ and $H^{N}$) and one $^{15}N$ dimension ($^{15}N^{H}$) [141, 144]. As with CBCA[CO]NH, this experiment produces a spectrum where only resonances corresponding to the previous residue are associated with a given amide proton, with Hα and Hβ chemical shift information recorded for all residues except those preceding a residue without an amide resonance in the $^{1}H$,$^{15}N$ HSQC spectrum, such as prolines. Figure 3.9 presents a diagram of the magnetisation transfer for this experiment and an example of Hα and Hβ assignments for a continuous stretch of the sequence.

**Figure 3.9 (Next page) Assignment and sequence connection of Hα and Hβ in SMARCAD1 SIM/CUE1.** A) Coherence transfer diagram for HBHA[CBCACO]NH: magnetisation is transferred between $^{1}H\alpha$ and $^{1}H\beta$ through the backbone ($^{13}C\alpha$ and $^{13}C\beta$) first to the $^{13}C_{CO}$ then selectively to $^{15}N^{H}$ and to $^{1}H^{N}$ for detection, B) Strips of HBHA[CBCACO]NH spectrum (purple) for a continuous stretch of residues in the primary sequence of SMARCAD1 SIM/CUE1. Open squares are peaks from residues *i*-1, while numbers at the bottom of each strip indicate the $^{15}N^{H}$ chemical shift (ppm).

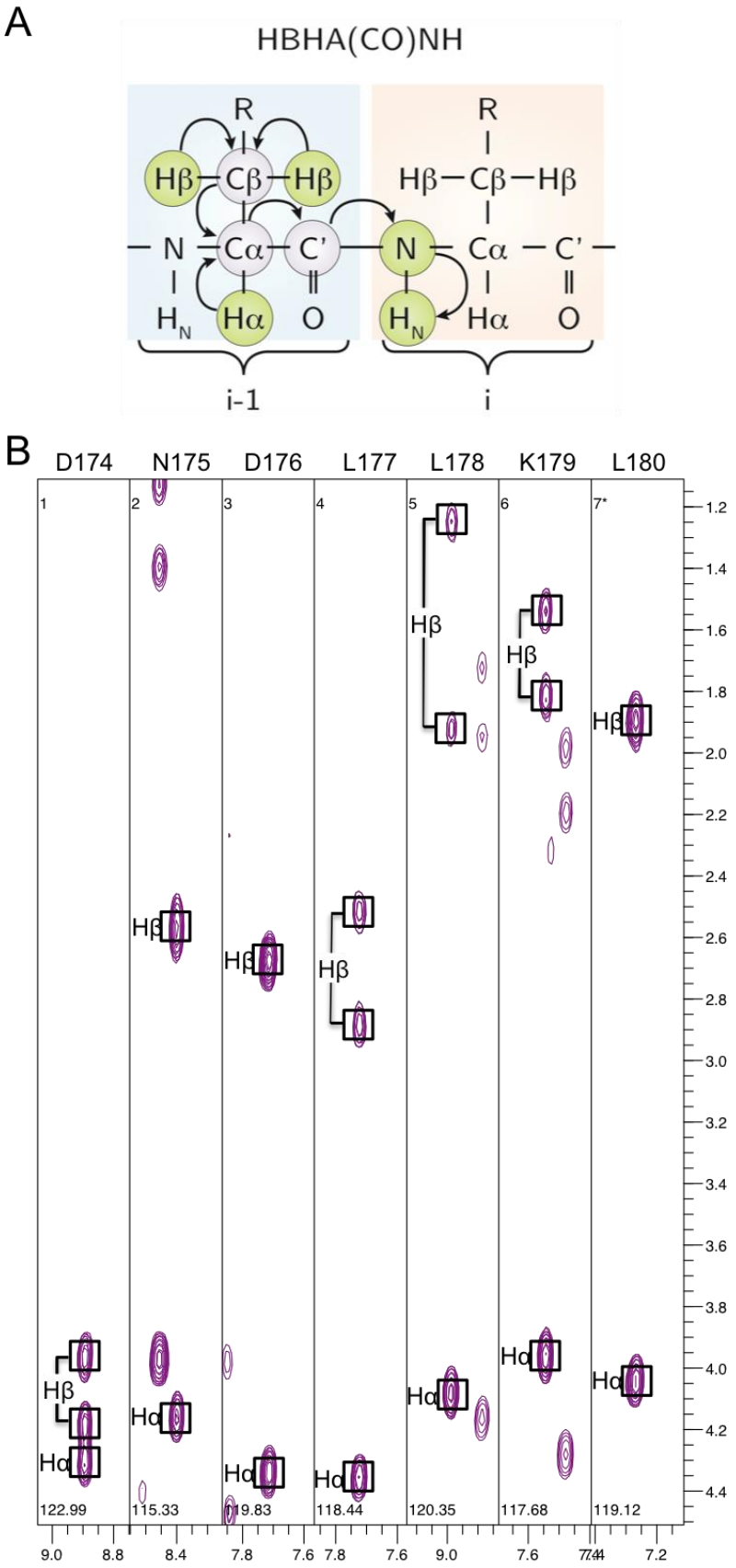



**Figure 3.9 Assignment and sequence connection of Hα and Hβ in SMARCAD1 SIM/CUE1 (Legend on previous page).**

Furthermore, it was possible to assign the chemical shift of carbonyl carbon atoms with the HNCO experiment, whose coherence transfer diagram is represented in Figure 3.10. In this experiment, the magnetisation is transferred from the amide proton selectively to the carbonyl carbon of the preceding residue through the amide nitrogen nucleus, and back to the amide proton for detection, resulting in a 3D spectrum with a single carbonyl resonance from residue *i*-1 for a given amino acid in the sequence. This is the most sensitive triple-resonance experiment available for double-labelled samples [143, 145, 147], and despite the limited chemical shift dispersion it was possible to unequivocally assign nearly all peaks, with the noted exception of residues preceding proline. Figure 3.10 shows a selection of continuous resonances in the SMARCAD1 SIM/CUE1 construct.

**Figure 3.10 (Next page) Assignment and sequence connection of $^{13}C_{CO}$ in SMARCAD1 SIM/CUE1.** A) Coherence transfer diagram for HNCO: magnetisation is transferred from $^{1}H^{N}$ to $^{15}N^{H}$ and then selectively to $^{13}C_{CO}$ through the $^{15}N^{H}$-$^{13}C_{CO}$ scalar coupling. Magnetisation is then transferred back to $^{1}H^{N}$ for detection. The chemical shift is evolved on all three nuclei resulting in a three-dimensional spectrum B) Strips of HNCO spectrum (red) for a continuous stretch of residues in the primary sequence of SMARCAD1 SIM/CUE1. Numbers at the bottom of each strip indicate the $^{15}N^{H}$ chemical shift (ppm).

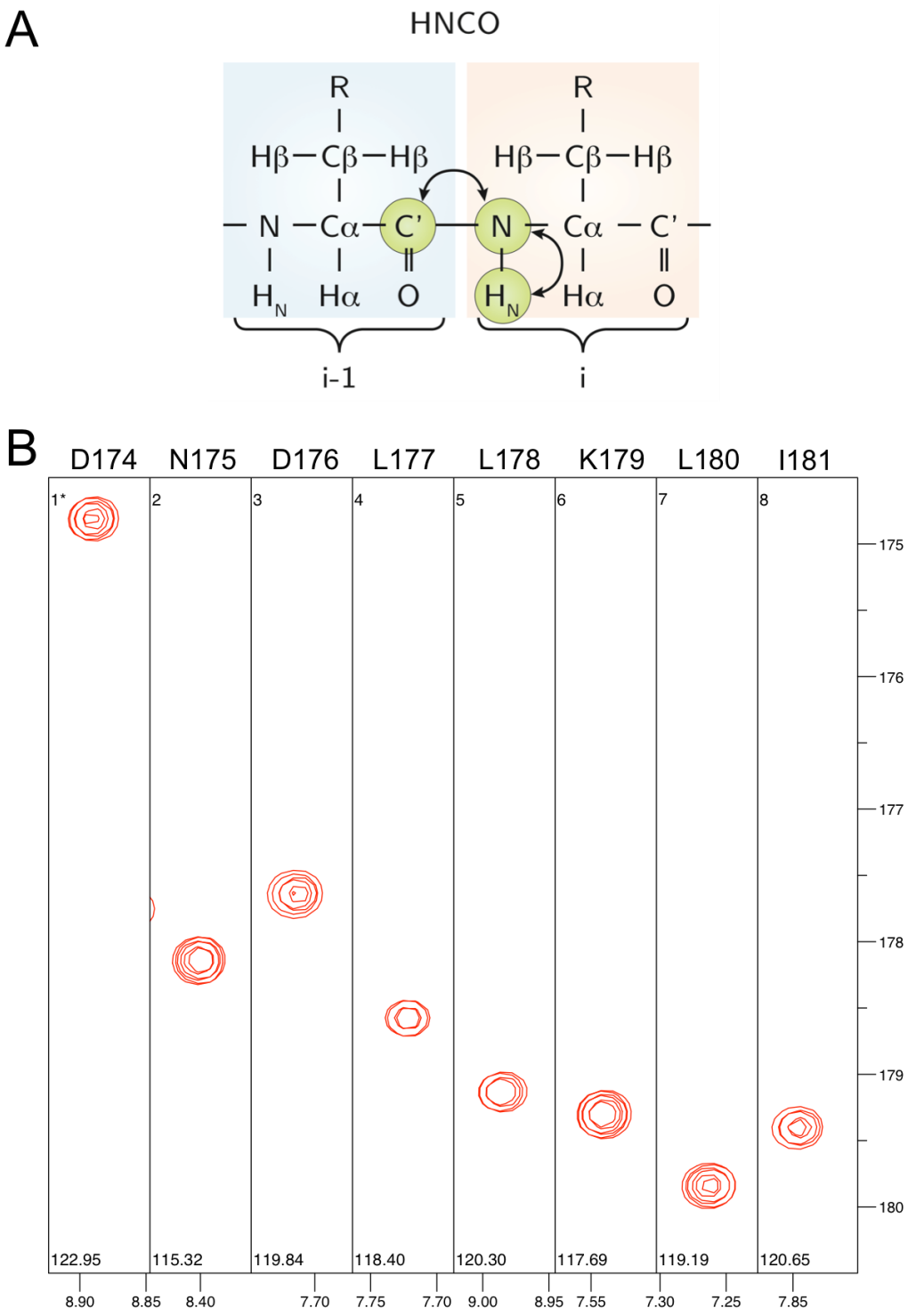



**Figure 3.10 Assignment and sequence connection of $^{13}C_{CO}$ in SMARCAD1 SIM/CUE1 (Legend on previous page).**

Finally, sidechain assignments beyond Cβ were determined through the use of [H]CC[CO]NH and related H[CCCO]NH experiments. In both, magnetisation is transferred from all aliphatic sidechain protons to the aliphatic carbons and through the backbone to the amide proton, selecting specifically for the magnetisation originating from *i*-1 via the $^{15}N^{H}$-$^{13}C_{CO}$ scalar coupling. The difference between the experiment arises from the nuclei for which chemical shift information is evolved, equivalent to the CBCA[CO]NH and HBHA[CO]NH experiments described above. In the case of [H]CC[CO]NH chemical shift is evolved on aliphatic carbons along with amide nitrogen and proton nuclei, resulting in a 3D triple-resonance experiment ($^{1}H^{N}$, $^{15}N^{H}$, $^{13}C^{H}$). On the other hand, for H[CCCO]NH chemical shift is evolved on aliphatic protons, and along with the amide nuclei, result in a 3D double-resonance experiment ($^{1}H^{N}$, $^{15}N^{H}$, $^{1}H^{C}$). Figure 3.11 shows the diagram representation for the coherence energy transfer of each experiment, along with the corresponding example of chemical shift assignment of a small stretch of sequential amino acids in the SMARCAD1 SIM/CUE1 construct.

**Figure 3.11 (Next page) Assignment of sidechain resonances beyond Cβ in SMARCAD1 SIM/CUE1.** A) Coherence transfer diagram for H[CCCO]NH: magnetisation is transferred between all aliphatic protons $^{1}H^{C}$ to its corresponding aliphatic carbon $^{13}C^{H}$ and then through the backbone first to the $^{13}C_{CO}$ then selectively to $^{15}N^{H}$ and to $^{1}H^{N}$ for detection. Chemical shift is not evolved for aliphatic carbons, but rather only for $^{1}H^{C}$, $^{15}N^{H}$ and $^{1}H^{N}$, B) Coherence transfer diagram for [H]CC[CO]NH: as with H[CCCO]NH, magnetisation is transferred between all aliphatic protons $^{1}H^{C}$ to its corresponding aliphatic carbon $^{13}C^{H}$ and then through the backbone to $^{1}H^{N}$ for detection, via the $^{13}C_{CO}$ and $^{15}N^{H}$. Chemical shift is not evolved for aliphatic protons, but rather only for $^{13}C^{H}$, $^{15}N^{H}$ and $^{1}H^{N}$, C) Strips of H[CCCO]NH spectrum (orange) for a continuous stretch of residues in the primary sequence of SMARCAD1 SIM/CUE1. D) Strips of [H]CC[CO]NH spectrum (purple) for a continuous stretch of residues in the primary sequence of SMARCAD1 SIM/CUE1. Open squares are peaks from residues *i*-1, while numbers at the bottom of each strip indicate the $^{15}N^{H}$ chemical shift (ppm).

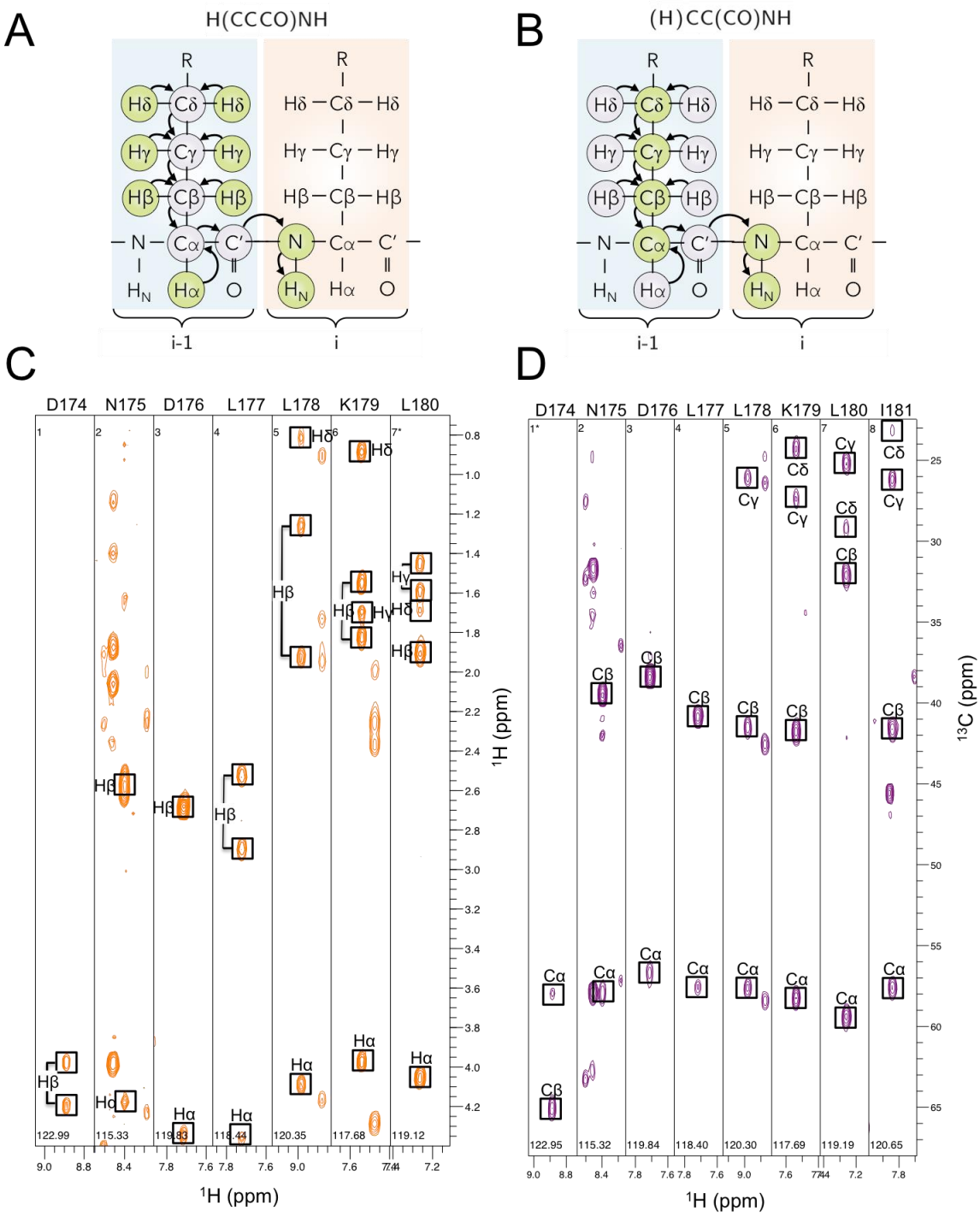



**Figure 3.11 Assignment of sidechain resonances beyond Cβ in SMARCAD1 SIM/CUE1 (Legend on previous page).**

Residues preceding prolines (Glu125, Leu135, Phe169, Gly204) or those followed by amino acids without an amide resonance in the $^{1}H,^{15}N$ HSQC spectrum, where carbonyl carbon selective experiments could not provide chemical shift information for residue *i*-1, were complemented with additional experiments such as $^{15}N$-TOCSY-HSQC and HCCH-TOCSY in which magnetisation of all sidechain resonances is coupled directly to the amide proton or aliphatic carbons of the residue *i* itself. These experiment were generally used to supplement and verify all assignments. The overall strategy resulted in 91% of all residues between amino acids 109-206 to be fully assigned, and 8% partially assigned. Of the latter subset, two residues (Val110 and Met178) do not present an amide resonance, and its preceding residues, along with those of the four prolines in the sequence, lack carbonyl carbon chemical shift information. Only one residue, Thr109 is fully unassigned. Table 3.1 presents detailed statistics of the backbone and sidechain assignments.

**Table 3.1 Extent of backbone and sidechain assignment for SMARCAD1 SIM/CUE1 CtHis construct spanning amino acids 109-206.**

| Backbone Assignments | |
|---|---|
| $^{15}N^{H}$ | 92.9% (91/98) |
| $^{1}H^{N}$ | 96.8% (91/94) |
| $^{1}H\alpha$ | 99.0% (103/104) |
| $^{13}C\alpha$ | 99.0% (97/98) |
| $^{13}C_{CO}$ | 91.8% (90/98) |
| Backbone $^{1}H$ | 95.1% (369/388) |
| Backbone non-$^{1}H$ | 94.6% (278/294) |
| **Sidechain Assignments** | |
| $^{1}H\beta$ | 99.4% (161/162) |
| $^{13}C\beta$ | 98.9% (91/92) |
| $^{1}H\gamma$ | 98.1% (103/105) |
| $^{13}C\gamma$ | 78.8% (67/85) |
| $^{1}H\delta$ | 98.5% (67/68) |
| $^{13}C\delta$ | 69.8% (44/63) |
| $^{1}H\varepsilon$ | 76.7% (23/30) |
| $^{13}C\varepsilon$ | 66.67% (8/12) |
| Sidechain $^{1}H$ | 97.0% (458/472) |
| Sidechain non-$^{1}H$ | 76.4% (217/284) |

## 3.5 Backbone dynamics

Having assigned all resonances for the SMARCAD1 SIM/CUE1 working construct it was now possible to assess and interpret additional experiments to structurally characterize the protein. Among these, the analysis of fast time scale backbone dynamics is particularly important and NMR is ideally suited to provide detailed information concerning dynamically active regions of the system [148, 149]. In the case of a $\{^{1}H\}$-$^{15}N$ heteronuclear NOE (HetNOE) experiment it is possible to identify mobile residues within the construct sequence, of special importance for the N-terminal extension beyond the CUE1 domain, which is expected to be largely flexible.

Figure 3.12 presents the steady state $\{^{1}H\}$-$^{15}N$ HetNOE ratio, determined from the intensities of each amide proton resonance when the $^{1}H$ frequency is alternatively saturated or not [150]. This plot reflects a decrease in amide proton intensity proportionate to the flexibility propensity of a given residue, which is a consequence of faster relaxation rates of the saturated nuclei for individual $N^{H}$-$H^{N}$ bond vectors that undergo faster motion in solution than the overall tumbling of the molecule. From this it may be deduced that the CUE1 domain is a fairly rigid structure, resulting in high heteronuclear NOE ratios, whereas the N-terminal extension is considerably flexible.

**Figure 3.12 (Next page) $\{^{1}H\}$-$^{15}N$ Heteronuclear NOE of SMARCAD1 SIM/CUE1.** Histogram of heteronuclear NOE ratio and amino acid residue number, derived from individual resonance intensity between two spectra used in the experiment, one with proton saturation and one without. The relative decrease in peak intensity for each resonance is correlated with residue mobility; an NOE ratio of ca. 0.7-0.8 represents a rigid backbone while a ratio of < 0.6 indicates backbone mobility.

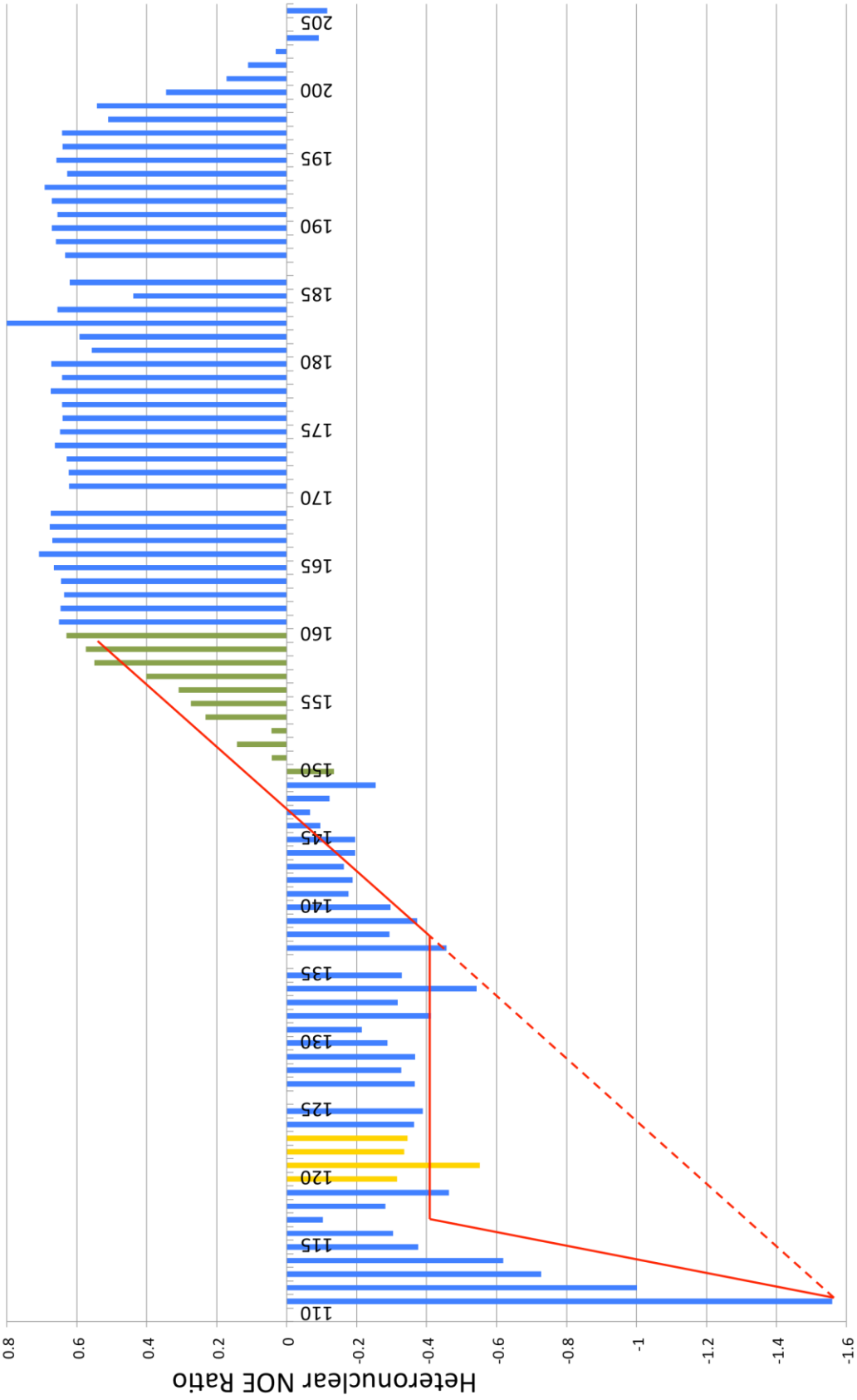



**Figure 3.12 {$^{1}$H}-$^{15}$N Heteronuclear NOE of SMARCAD1 SIM/CUE1 (Legend on previous page).**

Figure 3.12 displays two regions of particular note: first, the region spanning Asp150-Ala160 (green), just prior the CUE1 domain. In it the heteronuclear NOE ratio decreases less drastically than other transition points in the histogram marking changes from a structured to a freely flexible segment (the C-terminus region between Phe198 and Arg206, for example). This could suggest a reduced range of conformational space sampled for various reasons: partial inter-domain interaction and steric hindrance with CUE1, an additional helical element separated from the globular core by a flexible hinge, or a dynamic, transitioning section of the protein, actively switching between flexible and structured conformational states. This same region is assigned as helical by TALOS+, the torsion angle prediction software [151], commonly used to derive structural restraints, which ascribes three helical segments spanning the CUE1 domain connected by short flexible loops, as shown in Figure 3.13 (a detailed description of TALOS+ is presented in Section 3.6). Interestingly, this is also the region that introduced the structural ambiguity in the CUE1 models previously obtained in our group. The unusual behaviour in the Heteronuclear NOE, indicating neither a fully flexible nor entirely rigid conformation, coupled with helical torsion angles predicted by TALOS+ might be enough to constrain structure calculations and erroneously assign this section of the protein as an α-helix without sufficient evidence.

The second region of interest is that of the SUMO interacting motif, between amino acids Thr119 and Glu125 (yellow in Figure 3.12). Though flexible, with a HetNOE ratio < 0, this section of the histogram, along with the N-terminal acidic patch neighbouring it, appears to plateau before swiftly decreasing when approaching the N-terminus of the protein. A set of red lines in Figure 3.12 describes the sharp change in the HetNOE ratio (continuous line) and the expected behaviour had the structure sequence segment truly behaved as a random coil in solution (dashed line). This behaviour might reflect a tendency towards conformational rearrangement or indicate a more rigid patch connected to a more flexible linker.

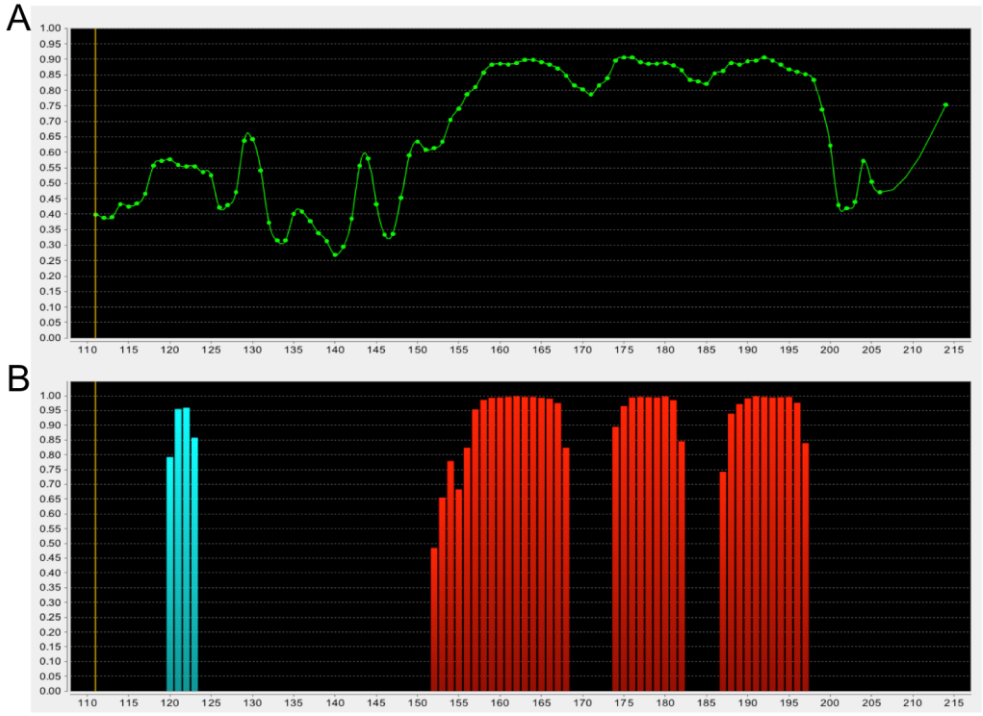



**Figure 3.13 Torsion angles predicted by TALOS+ for SMARCAD1 SIM/CUE1 domains.** A) Green plot shows the random coil index (RCI) derived order parameter $S^2$, a measure of rigidity based on the difference between the individual torsion angles derived from experimental and random coil chemical shift values, B) Artificial neural network (ANN) predicted secondary structure propensity. TALOS+ results suggest a triple helical structured region in the CUE1 domain (red bars), and an extended β-strand conformation for the short hydrophobic SIM patch (blue bars).

To further evaluate these regions of relative flexibility ambiguity, experimental determination of coupling constants was performed through the use of an HNHA experiment. Three-bond $H^N$-$N^H$-Hα coupling constants may be used to determine dihedral angles ϕ of the peptide backbone through a Karplus curve [152, 153], which allows assessment of the secondary structure propensity of the protein.

The HNHA experiment is a double-resonance 3D quantitative experiment where the coupling constant is related to the ratio of the diagonal ($H^N$-$H^N$) and cross peak ($H^N$-Hα) corresponding to each residue [154, 155]. The intensity of the $H^N$-Hα cross peak is related to the three-bond $H^N$-$N^H$-Hα coupling constants; amide resonances in α-helical secondary structure element the $^3J_{HNHA}$ < 4Hz give in weak cross peaks,

whereas β-strands with $^{3}J_{HNHA}$ > 8.5Hz produce strong cross peaks. Figure 3.13 shows the histogram of HNHA coupling constants for the SMARCAD1 SIM/CUE1 primary sequence. From it, is possible to deduce that there are indeed three well-defined helices in the CUE1 domain, consistent with uniformly rigid section of the protein suggested by the heteronuclear NOE. Coupling constants also indicate that the structurally ambivalent N-terminal section of the CUE1 domain, emphasised in Figure 3.14 by a semi-transparent cylinder, is in fact fully flexible, with $^{3}J_{HNHA}$ constants between 5Hz and 8Hz, as expected for conformational averaging. Finally, the SIM does have coupling constants consistent with an elongated conformation, but much of the sequence is within the range of fully flexible averaged conformation.

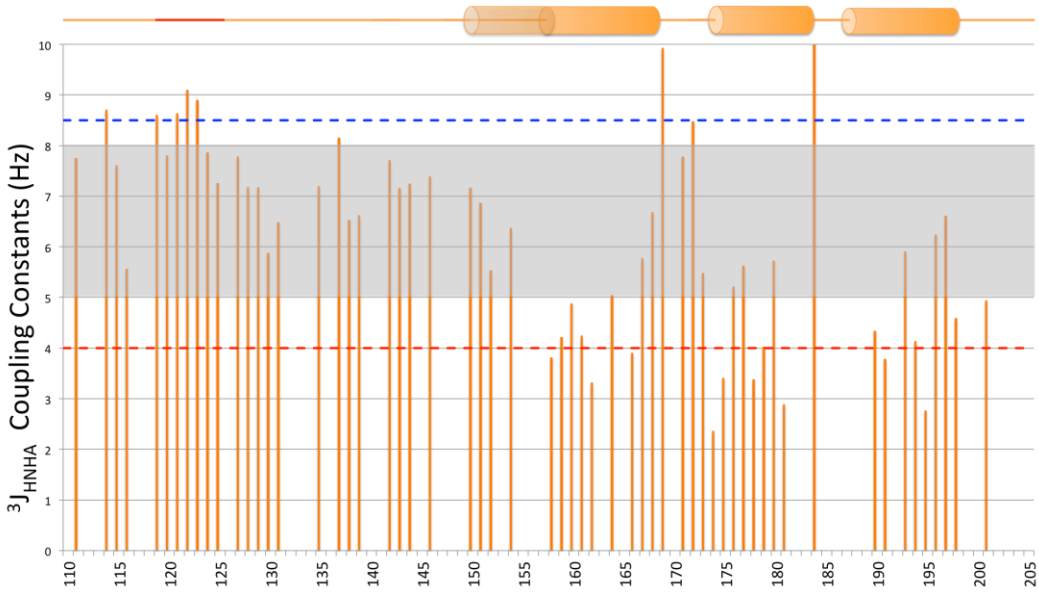



**Figure 3.14 HAHN coupling constant determination for SMARCAD1 SIM/CUE1.** Histogram of HNHA ratio and amino acid residue number, derived from individual resonance intensity between the diagonal ($H^{N}$-$H^{N}$) and cross peak ($H^{N}$-Hα) of a double-resonance 3D spectra. The relative ratio intensity for each residue is correlated with the three-bond $H^{N}$-$N^{H}$-Hα coupling constant; a $^{3}J_{HNHA}$ ≤ 4Hz (red) is consistent with a α-helix secondary structure and $^{3}J_{HNHA}$ ≥ 8.5Hz (blue) with a β-strand. Coupling constants were analysed in CcpNMR Analysis, assuming Karplus curve constants of a: 6.98, b: -1.38, and c: 1.72 [153].

Based on these results all subsequent structural calculations were restricted to the core CUE1 domain, between residues 158 and 198.

### 3.6 Collection of structural restraints

Structural determination through NMR requires geometric conformational information in order to limit and define the correct spatial arrangement for all atoms in the protein. Such restraints are derived from NMR data and interpreted as distance information between a pair of nuclei, with the most important subset being the NOE measurements between aliphatic protons in different spin systems [141, 156]. Additional information in the form of dihedral angles, hydrogen bond networks and residual dipolar couplings supplement long-range distance restraints and improve the precision of local conformational arrangements by reducing the divergence between potential models [157].

Hydrogen bond restraints refer to the known characteristic network present in secondary structure elements (both α-helices and β-strand forming a sheet), and as such were derived only for those regions in the protein where multiple data indicated the presence of secondary structure features, namely NOE patterns observed in $^{15}$N-NOESY-HSQC data for residues in the helical structured regions of CUE1 (*i* to *i*+4 links, primarily) and $^{3}J_{HNHA}$ coupling constants. Hydrogen bond restraints were designed so that all three atoms involved (the amide N of residue *i*, the amide proton, and the carbonyl oxygen of residue *i*+4) result in an energy minimum when they are co-linear. A total of 30 distance restraints mimicking H-bonds were introduced for structural calculations.

Backbone torsion angles ϕ and ψ were empirically predicted using the TALOS+ webserver [151]. As mentioned above, local protein conformation significantly influences backbone chemical shift contributions. This relationship is exploited by TALOS+ (and other similar prediction software) by mining databases of high-resolution X-ray structures, for which chemical shift information is also available, in search of peptide triplets with secondary chemical shift similarity to the queried

construct. Moreover, TALOS+ implements an artificial neural network (ANN) classification scheme to estimate the likelihood of a given residue being part of specific secondary structure element (helix, strand or loop), with about 89% prediction accuracy [151]. Chemical shift information for all assigned $^{15}N^{H}$, $^{13}C_{CO}$, $^{13}C\alpha$, $^{13}C\beta$, $^{1}H^{N}$, and $^{1}H\alpha$ nuclei were submitted, resulting in 82 dihedral angle restraints for the core globular domain for which it predicts three helical sections, in agreement with heteronuclear NOE and $^{3}J_{HNHA}$ coupling data. The predicted values of the ϕ and ψ angles were used to restrain the backbone to adopt regular secondary structures during calculations.

It is well established that the quality of a solution structure increases with the number of consistent input constraints used in the structure calculation, and as was mentioned above, $^{1}H$-$^{1}H$ NOE are particularly important in these efforts as they create a three dimensional network of distance restraints between sidechain protons across the sequence separated by less than ~5Å. Therefore, for a high-resolution structure determination, the largest possible number of NOE constraints must be collected as input for the calculation of the complete structural model [157]. As resonances were to be assigned based on pre-determined chemical shift information, a new, smaller construct, corresponding exclusively to the CUE1 domain of SMARCAD1 spanning residues 144 to 206, was expressed and purified as described above to reduce poor dispersion due to overcrowding and potential ambiguity when assigning NOE resonances. This construct, as a C-terminal tryptophan/hexahistidine fusion protein, resulted in a much simpler $^{1}H$,$^{15}N$ HSQC spectrum, with little overlapping throughout, as may be appreciated in Figure 3.15.

Long range $^{1}H$-$^{1}H$ NOEs between sidechain resonances were assigned based on a double-resonance $^{13}C$-edited NOESY, with magnetization exchanged between all protons, then transferred to neighbouring $^{13}C$ nuclei and back to $^{1}H$ for detection [141]. Resonances describe NOEs between a given aliphatic $CH_x$ and all others proton nearby, including intra-residue NOEs, local contacts (such as *i* to *i*±1 or *i* to *i*+4) and long-range connections. Additionally, $^{13}C$-$^{1}H$ HSQC was used to navigate the 3D

spectrum, much as the $^{1}H,^{15}N$ HSQC is related to a $^{15}N$-edited NOESY. Figure 3.16 presents an example of long-range distance restraint connections for Met187.

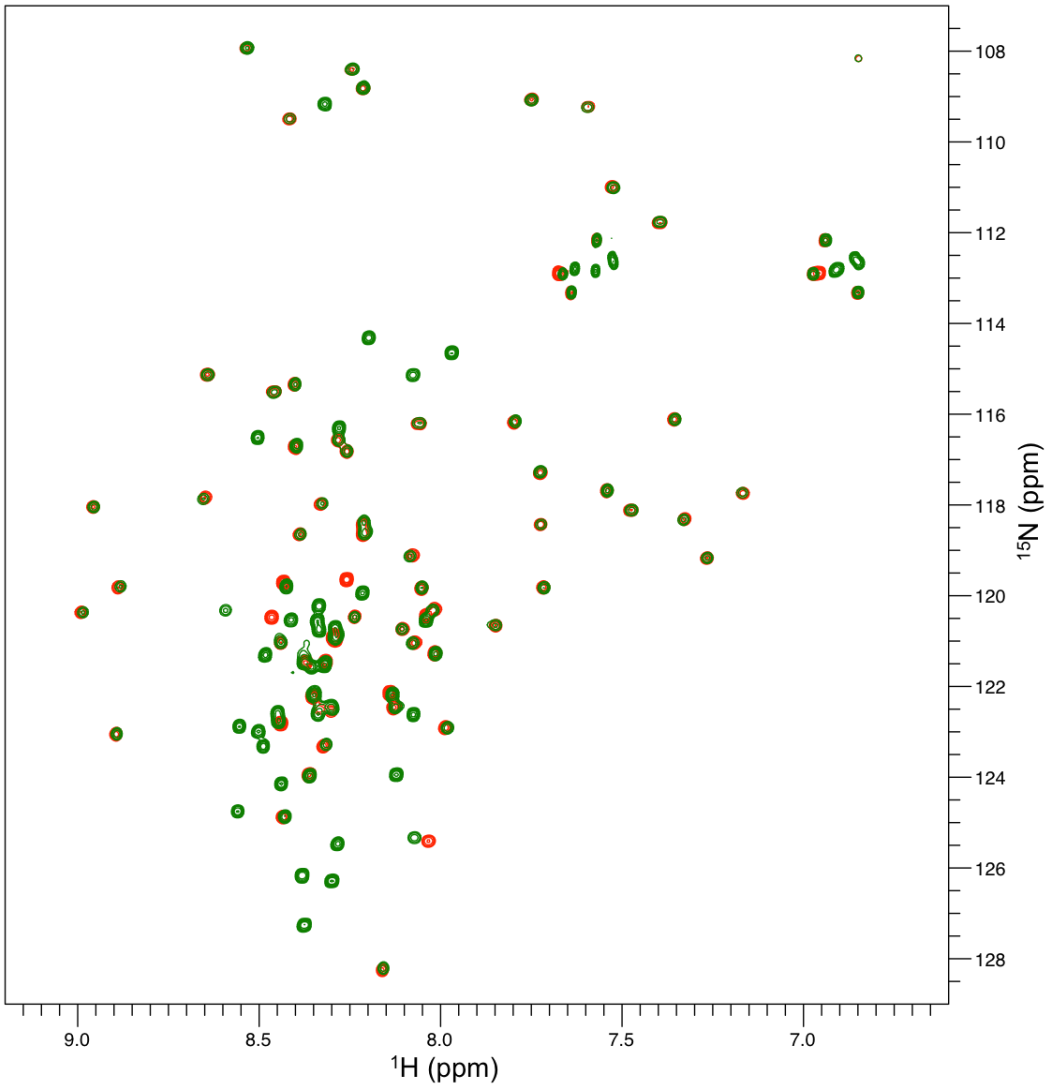



**Figure 3.15 $^{1}H,^{15}N$ HSQC spectra of SMARCAD1 constructs.** Overlay of HSQC spectra SIM/CUE1 CtHis (SMARCAD1 residues 109-206) in green and CUE1 CtHis (SMARCAD1 residues 144-206) in red.

**Figure 3.16 (Next page) Long-range distance restraints in SMARCAD1 CUE1 domain.** A $^{13}C$-edited NOESY-HSQC experiment (purple) was used to identify connections between nuclei, and in particular, those between sidechain nuclei from different residues (long and medium range). Chemical shift information for the fully assigned construct was used to tentatively assign all resonances. Circles (•) indicate the cross-peak of the strips within a given spin system and open squares mark NOE peaks on a different residue. NOE connections for Met187 are shown.

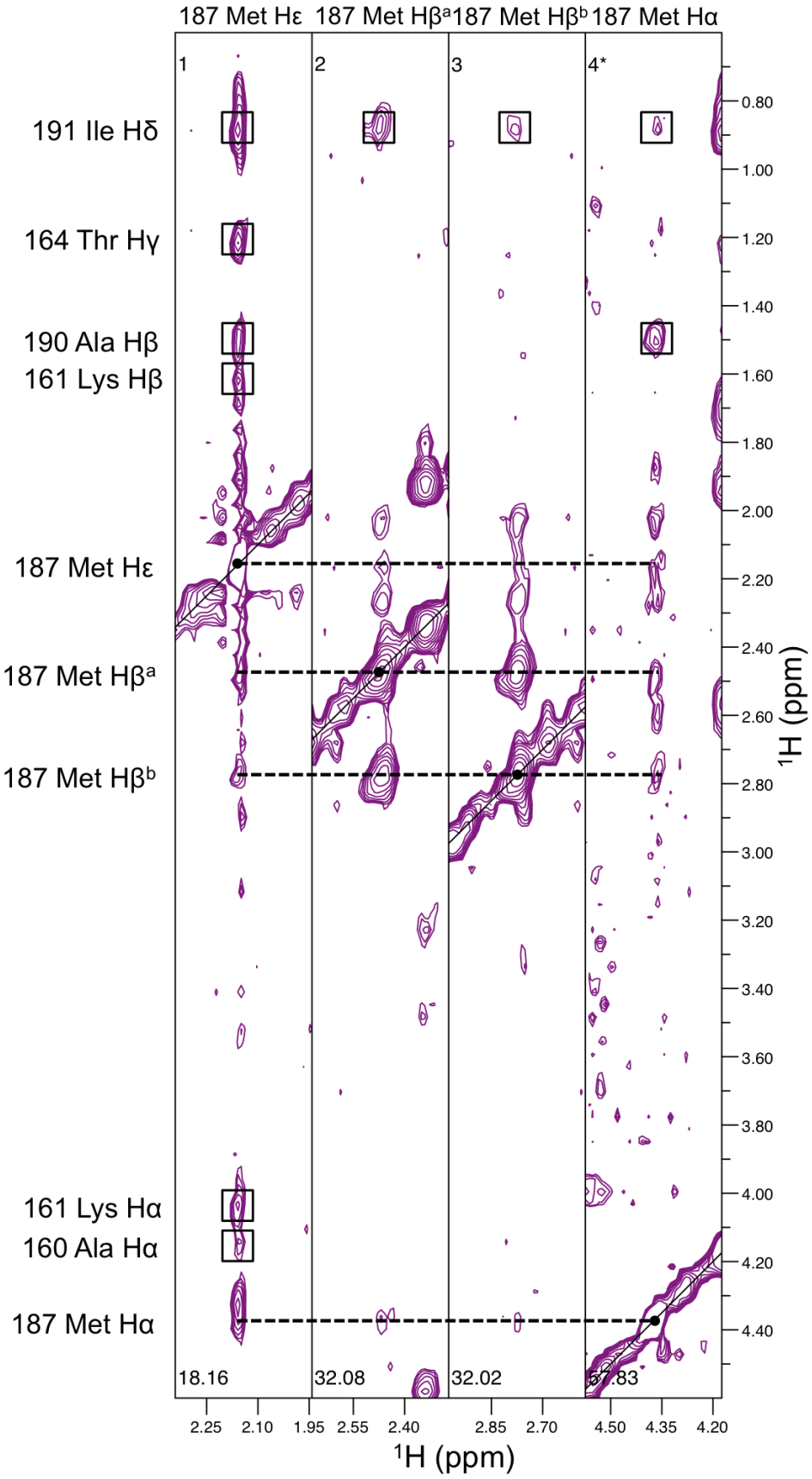



**Figure 3.16 Long-range distance restraints in SMARCAD1 CUE1 domain. (Legend on previous page).**

It should be stated, however, that NOESY spectra (both $^{15}N$- and $^{13}C$-edited) tend to be very crowded because of the large number of proton pairs with inter-nuclear distances under 5Å, spanning short-, medium-, and long-range restraints. Therefore, many ambiguities may not be resolved based on chemical shift information alone, with several possibilities existing even in a small construct such as the ≈9kDa SMARCAD1 CUE1 CtHis. In practice, iterative cycles of assignment and structural calculation lead to the discrimination of ambiguous assignment possibilities, improving the precision of 3D structures after each round [158, 159].

Finally, additional constraints were derived from residual dipolar coupling constants (RDC), which are observable in weakly aligned proteins. These couplings are related directly to the orientation of $H^N$-$N^H$ inter-nuclear vectors relative to the molecular frame (the global alignment tensor) and arise from the fact that anisotropic contributions to nuclear interactions do not average to zero when the protein has a preferred orientation [160-162]. To measure these residual contributions a frequency resolved IPAP (in-phase/anti-phase) experiment was employed, which results in $^1J_{NH}$ splitting for all amide resonances in the $^{15}N$ dimension of a $^1H$,$^{15}N$ HSQC-type spectrum [163]. Specifically, two samples of SMARCAD1 CUE1 domain were prepared, one isotropic in Phosphate Buffer (20mM $NaH_2PO_4$ pH7.0, 100mM NaCl), and a second one in Phosphate Buffer supplemented with 3%$^v/_v$ hexanol/$C_{12}E_6$ (hexaethylene glycol monododecyl ether), which creates a partially aligning dilute liquid crystal, and is thus anisotropic [164]. Both samples were measured with the IPAP method and the doublet peak separation for each residue was determined, with the difference between isotropic and partially aligned samples used to calculate the RDCs, and with them the alignment tensor magnitude and rhombicity. Figure 3.17 presents a section of the $^1H$,$^{15}N$ HSQC-IPAP spectra of both the isotropic and aligned samples and Figure 3.18 shows a set of histograms of all RDC values unambiguously measured for the construct.

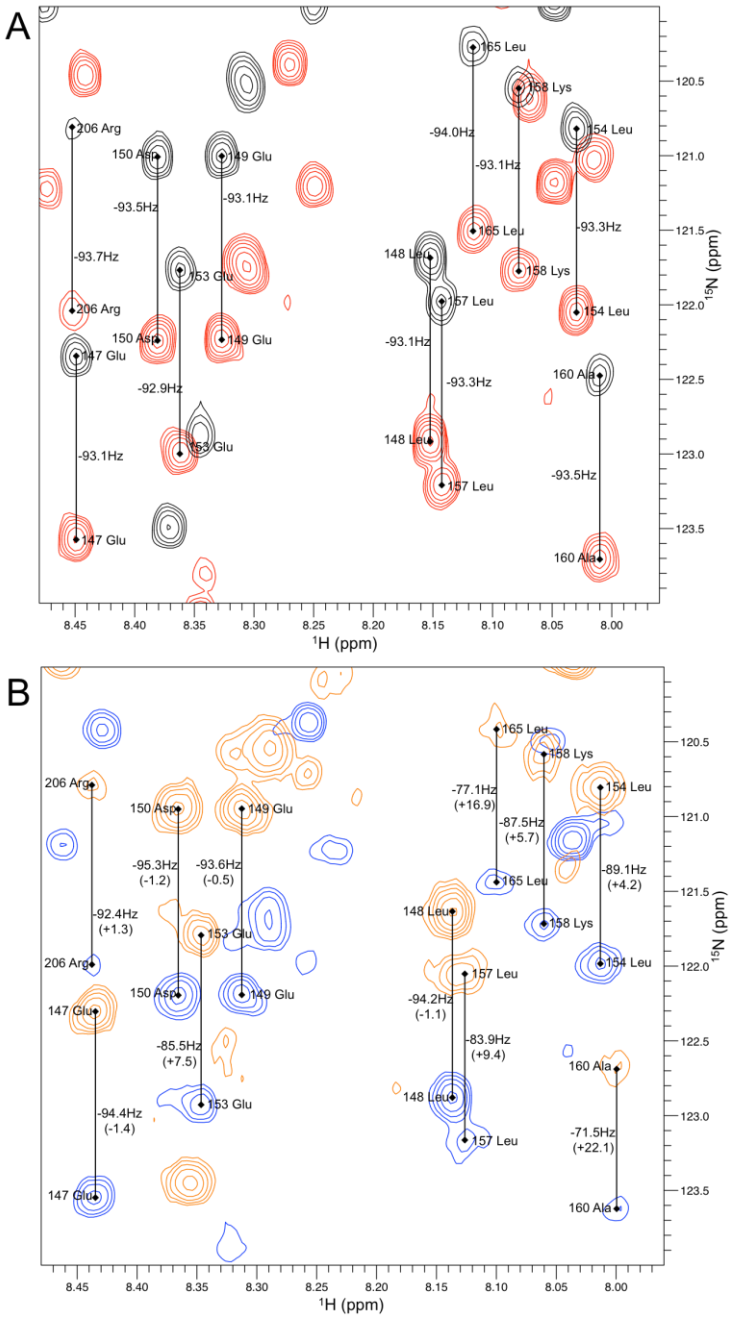



**Figure 3.17 Residual Dipolar Couplings for SMARCAD1 CUE1 domain from HSQC-IPAP Spectra.** A) Isotropic $^{1}H,^{15}N$ HSQC-IPAP spectra of $^{15}N$-CUE1 CtHis, B) Anisotropic $^{1}H,^{15}N$ HSQC-IPAP spectra of $^{15}N$-CUE1 CtHis. RDC values were determined by measuring peak separation in each sample, with the difference between the two proportional to the residue specific contribution. Splitting in the isotropic solution is uniform, without contribution from RDC, and corresponds to the $^{1}J_{NH}$ scalar coupling (approximately -93 Hz), whereas the aligned sample in the presence of 3%$^{v}/_{v}$ hexanol/$C_{12}E_6$ results in splittings of varied magnitude due to RDC contributions.

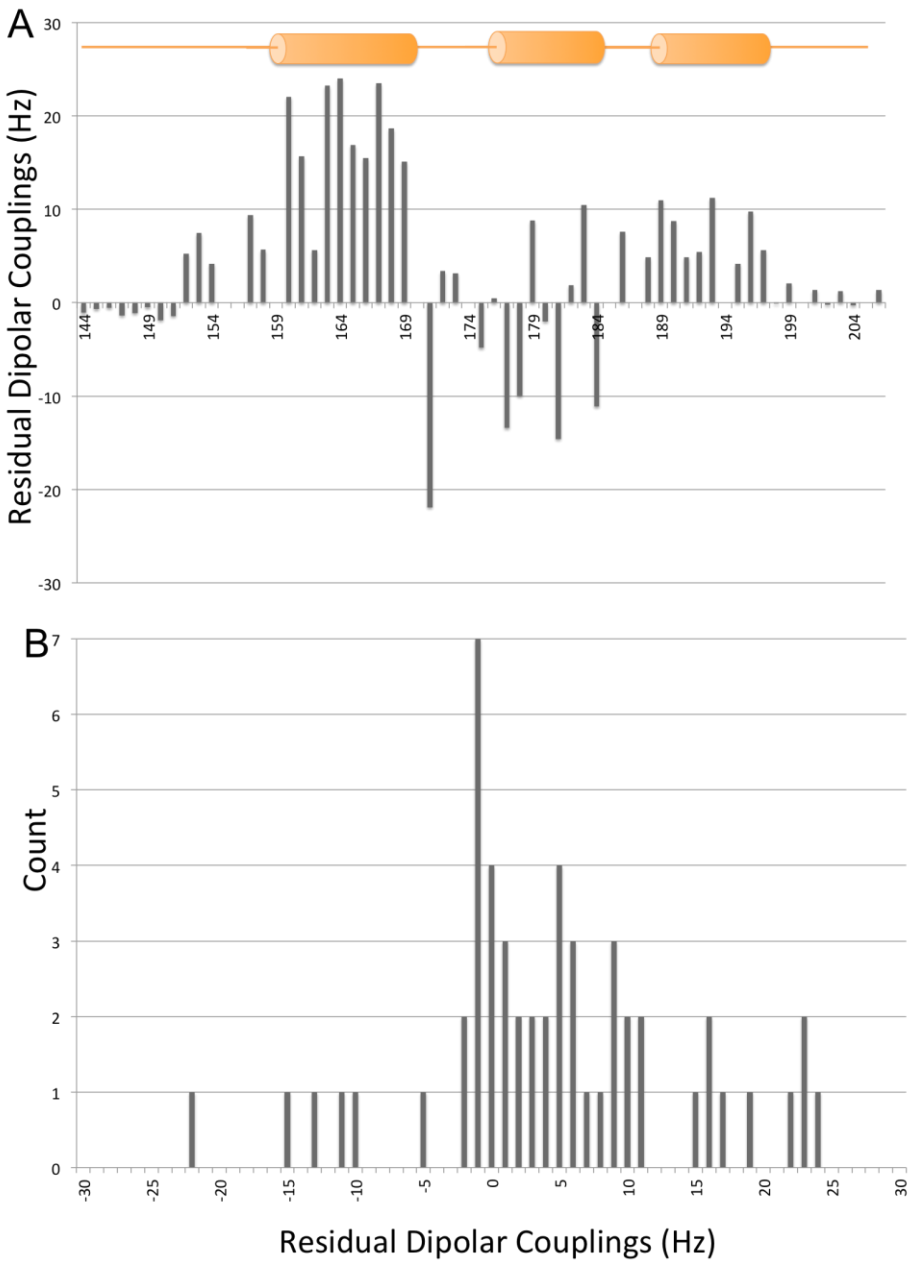



**Figure 3.18 Residual dipolar couplings for SMARCAD1 CUE1 domain.** A) Plot of $H^N$-$N^H$ RDC values relative to SMARCAD1 CUE1 domain primary sequence. Similar RDC values are observed for all residues within an individual helix as all amide bond vectors are aligned parallel to the helix axis and thus to one another. B) Histogram of RDC magnitude distribution for CUE1 domain. Due to distinct asymmetrical distribution it was not possible to use the traditional histogram method to determine the global alignment tensor constants, $D_{NH}$ and R.

Empirical determination of the magnitude ($D_{NH}$) and rhombicity (R) of the global alignment tensor for the protein was performed with an iterative homology based approach (FITALIGN, Prof Christina Redfield, University of Oxford), using a preliminary structure of SMARCAD1 CUE1 domain determined without RDC restraints. Alternative values for magnitude and rhombicity were introduced, aiming to minimize the difference between predicted and experimental RDC values over repetitive iterations. The final value for each constant was $D_{NH}$ = -12Hz and R = 0.6, and these were introduced in all subsequent structure calculations.

3.7 Structural calculations.

Calculation of NMR structures was performed using a molecular dynamics based simulated annealing protocol to minimize a global energy function, using the fully automated ARIA (Ambiguous Restraints for Iterative Assignment) program for NOE assignment validation and structure calculation [124, 158, 165, 166]. ARIA is based on an iterative protocol to derive distance constraints from cross-peaks NOE assignments and distance calibration based on chemical shift information. These restraints are then applied to calculate an ensemble of structures, which serve to screen and discard violating or unlikely assignment possibilities, and thus the ambiguity present in the initial NOE assignments is progressively reduced. It is important to note that the experimental constraints do not uniquely describe one exact 3D structure because NMR-derived constraints typically describe a range of possible values and many distances cannot be determined. By repeating the structure calculation multiple times from independent starting points (elongated polypeptide chain with randomised backbone torsion angles) it is possible to determine an ensemble of structures consistent with the input data set, which minimizes violations and samples the complete conformational space allowed by said constraints.

For this work, manually assigned NOESY data was classified as either ambiguous or not depending on the number of assignment possibilities that match a given resonance. These restraints, a total of 217 NOE-derived inter-residue distances from

$^{13}$C-edited NOESY spectra for aliphatic and aromatic protons, were combined with 30 helical H-bonds, 82 dihedral angles and 34 RDCs and treated through 11 rounds of ARIA processing. For each, after the calibrated and verified NOE distance restraints were merged to avoid overrepresentation of redundant distance data, a new structure ensemble was calculated with the program CNS [167] through a Cartesian Space molecular dynamics simulated annealing (MDSA) protocol [168, 169], followed by a refinement step in a shell of explicit solvent (water molecules). This consists in a short molecular dynamics simulation with a complete force field (including electrostatics and Lennard-Jones potential), which has been shown to significantly improve structure quality [166, 170, 171]. Figure 3.19 shows an ensemble superposition of the calculated 25 lowest energy structures.

The model consists of a three-helix bundle, consistent with previous reports of CUE domains. From Figure 3.19 it may be appreciated that helices 1 and 3 are approximately parallel to each other, with helix 2 connecting the two while oriented diagonally. Furthermore, backbone atoms, as well as most sidechains involved in helix packing, show little deviation from the mean (Figure 3.20). Table 3.3 presents the structure refinement statistics for the ensemble, determined through the validation webserver iCING [172, 173]. From it, it is possible to appreciate that the structure has only small deviations from idealized covalent geometry, with all residues in allowed regions of the Ramachandran plot and the vast majority (90.3%) in the most favoured regions.

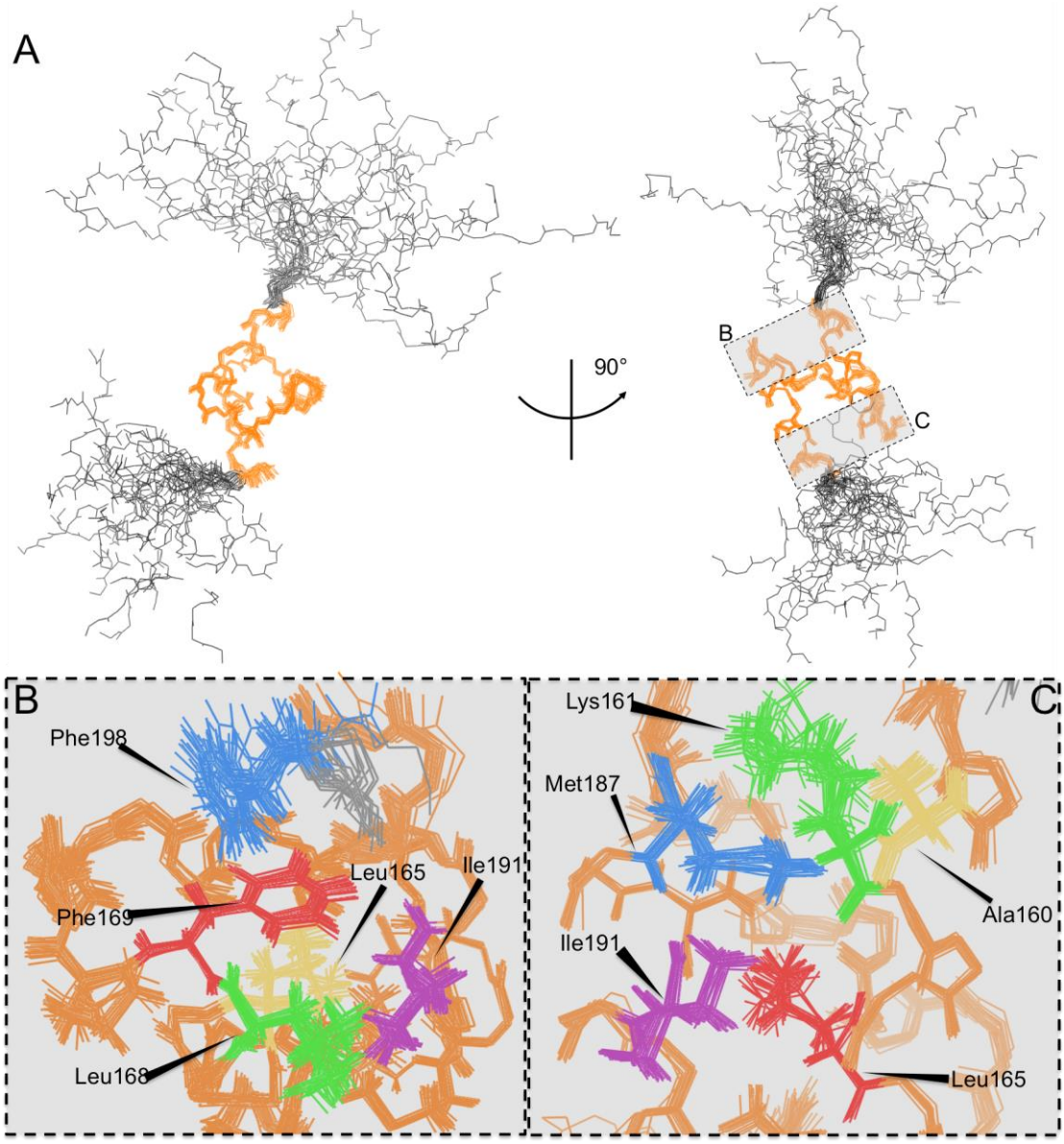



**Figure 3.19 SMARCAD1 CUE1 low-energy ensemble.** Backbone atom representation of the SMARCAD1 CUE1 domain, with 25 structures shown as an ensemble with ordered residues (158-198) in orange and more flexible, non-helical residues in grey. B) Sidechain atom representation for selected residues involved in helix packing near the C-terminus: L165 (yellow), L168 (green), F169 (red), I191 (purple), F198 (blue). C) Sidechain atom representation for selected residues involved in helix packing near the N-terminus: A160 (yellow), K161 (green), L165 (red), M187 (blue), I191 (purple).

**Table 3.2 NMR Structure refinement statistics for SMARCAD1 CUE1 low-energy ensemble.** [1]Based on 25 lowest total energy structures. [2]Ramachandran statistics generated using the protein structure validation software suite iCING. [3]RMSD calculated for ordered residues only aa157-198.

| **Backbone Assignments** | |
|---|---|
| Total NOE | 217 |
| Hydrogen bonds | 30 |
| RDC | 34 |
| Total dihedral angle constraints | 82 |
| **Sidechain Assignments** | |
| Potential energies[1] (kcal/mol) | |
| Total | -2148±78 |
| Bonds | 9.4±0.5 |
| Angles | 68±4 |
| Improper | 31±4 |
| van der Waals | -175±13 |
| NOE | -37±2 |
| Dihedral angles | 0.6±0.2 |
| RDC (Sani) | 12±2 |
| Ramachandran plot statistics[2] | |
| Residues in most favoured regions | 90.70% |
| Residues in allowed regions | 9.30% |
| Residues in disallowed regions | 0 |
| Root-mean-square deviation[3] (RMSD) (Å) | |
| All heavy atoms | 0.92 |
| Backbone atoms | 0.31 |

A single, energy-minimised average structure was generated from the ensemble of 25 lowest-energy structures, from which average deviation of backbone and sidechain carbons was determined and indicates the variability of the models in the ensemble [171, 174]. This plot, Figure 3.20, provides an indicator of potentially problematic residues in the structure, for which a higher level of uncertainty is associated.

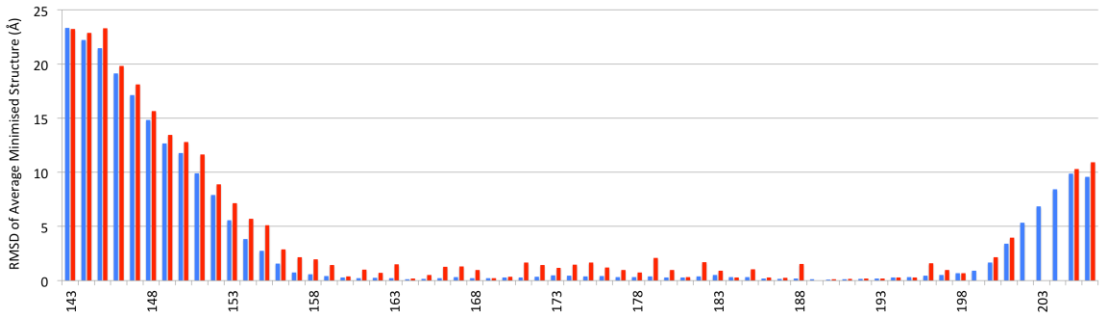



**Figure 3.20 Root-mean-square deviation (RMSD) of SMARCAD1 CUE1 low-energy ensemble relative to mean coordinates.** Histogram of average deviations of backbone (blue) and sidechain (red) atoms in the CUE1 ensemble.

## 3.8 Comparison of SMARCAD1 CUE1 with other CUE Domains

Comparison of the CUE1 energy-minimised average model with existing structures for other CUE domains confirms the high degree of structural conservation in this family fold and firmly places this domain as an unequivocal member. It is interesting to note, however, that the slight concave kink established by helices 1 and 3 relative to each other diverges from the general pattern of closely parallel helices by creating a more acute angle than that observed in *Sc*CUE2 (PDB: 1OTR) or *Hs*CUEDC1 (PDB: 2DHY), for which a superposition RMSD indicates a deviation at 3.35Å and 3.14Å, respectively. On the hand, an equivalent assessment with alternative CUE domains *Hs*TOLLIP (PDB: 1WGL) and *Hs*AMFR (PDB: 2LVO), whose helical bundles exhibit a sharper concave arrangement result in RMSD of 2.71Å and 2.95Å, respectively, suggesting that this variation in globular packing is a common feature in the fold. Figure 3.20 presents all four CUE domains previously mentioned, overlaid with the averaged model for SMARCAD1 CUE1.

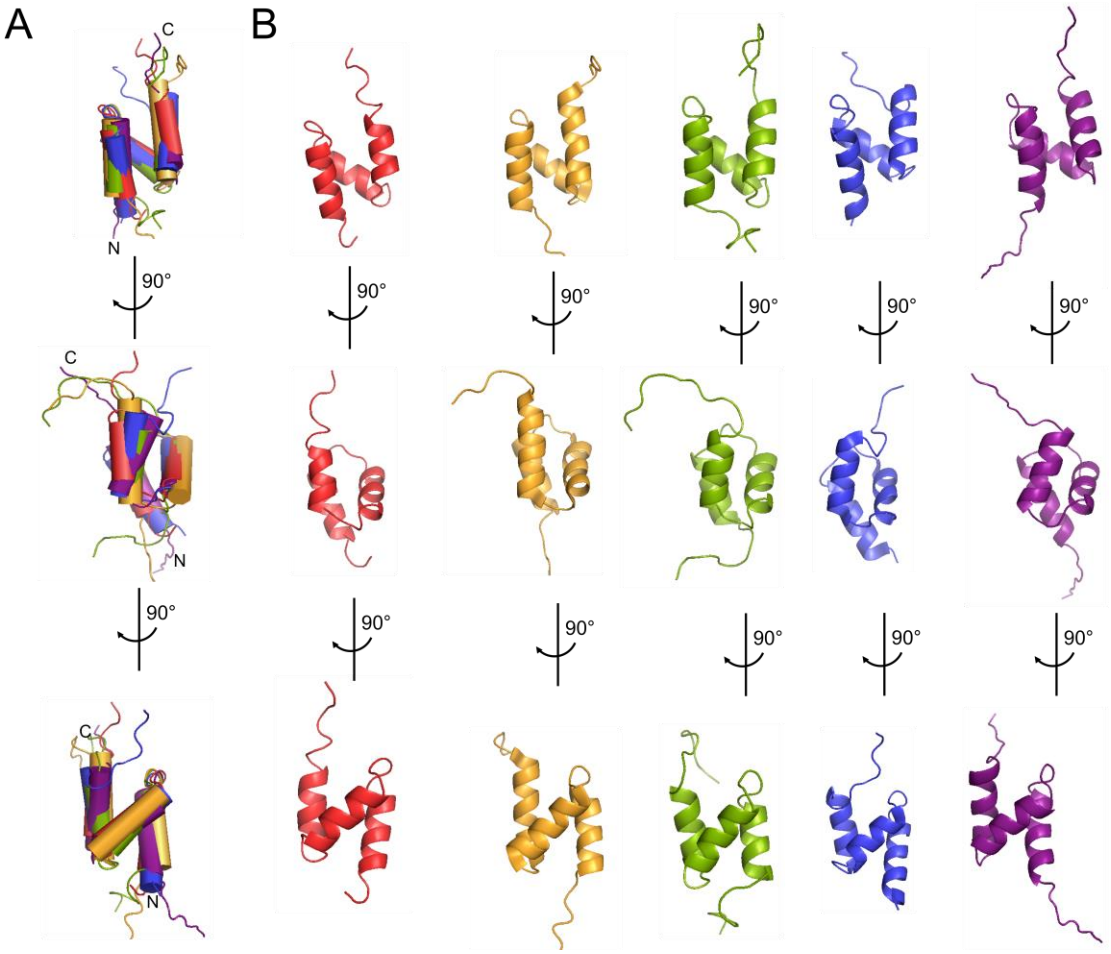



**Figure 3.20 Overlay of SMARCAD1 CUE1 structure with alternative CUE domains.** A) Energy minimised model for SMARCAD1 CUE1 (purple) was superimposed over residues aa157-198 with the deposited structures for the N-terminal CUE domain from yeast Cue2 protein (red), the CUE domain from human CUEDC1 (CUE Domain Containing Protein 1; yellow), the CUE domain form human TOLL-interacting protein (TOLLIP, green) and the CUE domain from human E3 ubiquitin-protein ligase AMFR (blue), with helices drawn as cylinders. It is easily appreciated that all structures superimpose well on two of three helices, whereas the last one exhibits a broad distribution of angles relative to the complementarily "parallel" helix1. B) Front, side and top view of all five CUE domain structures.

Chapter 4

# Characterisation of SMARCAD1 protein interactions mediated by the SMARCAD1 SIM and CUE1 domains

One of the most striking features of SMARCAD1 protein architecture is the presence of tandem ubiquitin-binding CUE domains as the only apparent protein interaction mediators. Though there is a high degree of sequence conservation between SMARCAD1 orthologs, this feature is unique among SNF2/SWI2 members of ATPase chromatin remodelers [136]. Interestingly, sequence conservation between either of the SMARCAD1 CUE domains and canonical members of this family such as yeast Cue2 or human TOLLIP is low, particularly among established key residues involved in Ubiquitin recognition (as described in Chapter 3), which is expected to negatively affect the affinity for this interaction. Having identified and confirmed a novel SUMO interacting motif (SIM) in the N-terminal vicinity of SMARCAD1 tandem CUE domains, careful characterisation of such a clustered region is required, as it is likely to mediate most, if not all, SMARCAD1 interactions. Of particular interest for the functional characterisation of SMARCAD1 in relation to DNA repair signalling cascade is the presence of potential phosphorylation sites, which are part of the SIM and have been identified as potential downstream targets of Ataxia telangiectasia mutated (ATM) kinase in a proteome wide study [66]. Thus, *in vitro* analysis of the binding events between these domains and their proposed ligands was performed in the form of NMR titrations and mass spectrometry protein profiling of co-immunoprecipitated interactors with SIM and CUE1 abrogating mutants.

This chapter describes the binary interaction between a SMARCAD1 SIM/CUE1 recombinant construct with its two canonical ligands: SUMO1 and Ubiquitin, individually assessed and quantified by NMR chemical shift perturbation assays (Sections 4.1 and 4.5, respectively). Furthermore, the possibility of phosphorylation dependent regulation of the SMARCAD1 SIM-SUMO1 interaction was evaluated through NMR titrations with phosphorylated peptides alternatively mono-phosphorylated on one of three Serine residues identified as potential targets of ATM kinase as part of the signalling cascade following DNA damage (Section 4.2). This is followed by a mass spectrometry PTM analysis to determine the endogenous phosphorylation state of SMARCAD1 upon chemically induced DNA damage (Section 4.3) and an additional SUMO1 titration experiment, this time with a recombinant phosphomimic mutant that better models the charge distribution of the protein in its native state (Section 4.4). Finally, differential protein abundance from co-immunoprecipitation assays were analysed by mass spectrometry profiling (Section 4.6). These samples, alternatively treated with or without a DNA damage inducing chemical agent, consisted of mammalian cells transiently transfected with GFP fusion proteins of full length, wild type SMARCAD1 or abrogating SIM/CUE1 mutants. Quantitation of differential protein abundance, relative to the cell state or protein sequence, facilitated the identification of significantly altered binding partners.

## 4.1 SUMO1 interaction with SMARCAD1 SIM

NMR titrations were performed as described in Chapter 2. Briefly, two solutions of uniformly labelled SMARCAD1 SIM/CUE1 ($^{15}$N eCUE1) were prepared with one of them containing the highest concentration point of unlabelled SUMO1 as ligand, and mixed in fixed ratios maintaining the labelled protein concentration constant through the experiment. 2D heteronuclear single quantum correlation (HSQC) spectra were collected for each titration point and the relative chemical shift changes were followed and plotted against ligand concentration.

Figure 4.1 shows an overlay of $^{1}$H,$^{15}$N HSQC spectra with varying concentrations of unlabelled SUMO1. It may be easily appreciated that multiple main chain amide

resonances are strongly perturbed, most of which correspond to the expected SIM sequence. Moreover, residues in the hydrophobic patch (V120, I121, I122, V123) are not easily detected beyond a ligand concentration of 1.0mM due to extensive line broadening, whereas residues adjacent to it (Asn116, Thr119, Ser124, Glu125, Ser127, Glu128, Asp129) shifted along a linear path with increasing concentrations of SUMO1 without severe line broadening, consistent with fast chemical exchange.

A chemical shift perturbation map of the interaction with SUMO1 at the highest total ligand concentration (5mM SUMO1, 50-fold excess) provides a clear illustration for the effect of ligand binding on all resonances throughout the SMARCAD1 SIM/CUE1 sequence. As appreciated in Figure 4.2, only the SIM sequence and immediate neighbouring acidic residues are perturbed significantly (Thr119 – Glu125), with a weaker effect for residues further removed (Asn116, Ser127, Glu128, Asp129). A second site, closer to the ubiquitin-binding CUE1 domain (Leu151, Ser152), shows larger than average perturbations though not enough to be considered significant. Moreover, this section of the sequence is highly degenerate and repetitive, with most of the amide resonances clustered together, and it was not possible to unambiguously assign the peaks as they shift with ligand binding. Finally, it is clear that no part of the globular CUE1 domain is interacting with SUMO1, as was expected given that despite sharing a similar tertiary structure, each protein is known to mediate interactions through distinct surface contacts.

**Figure 4.1 (Next page) NMR titration experiment of $^{15}$N eCUE1 with unlabelled SUMO1.** A) Overlay of fourteen $^{1}$H,$^{15}$N HSQC spectra for eCUE, each with different, colour coded SUMO1 concentrations. B) Selection of magnified $^{1}$H,$^{15}$N HSQC peaks that shifted considerably; the sign and magnitude of shift is consistent with increasing ligand concentration.

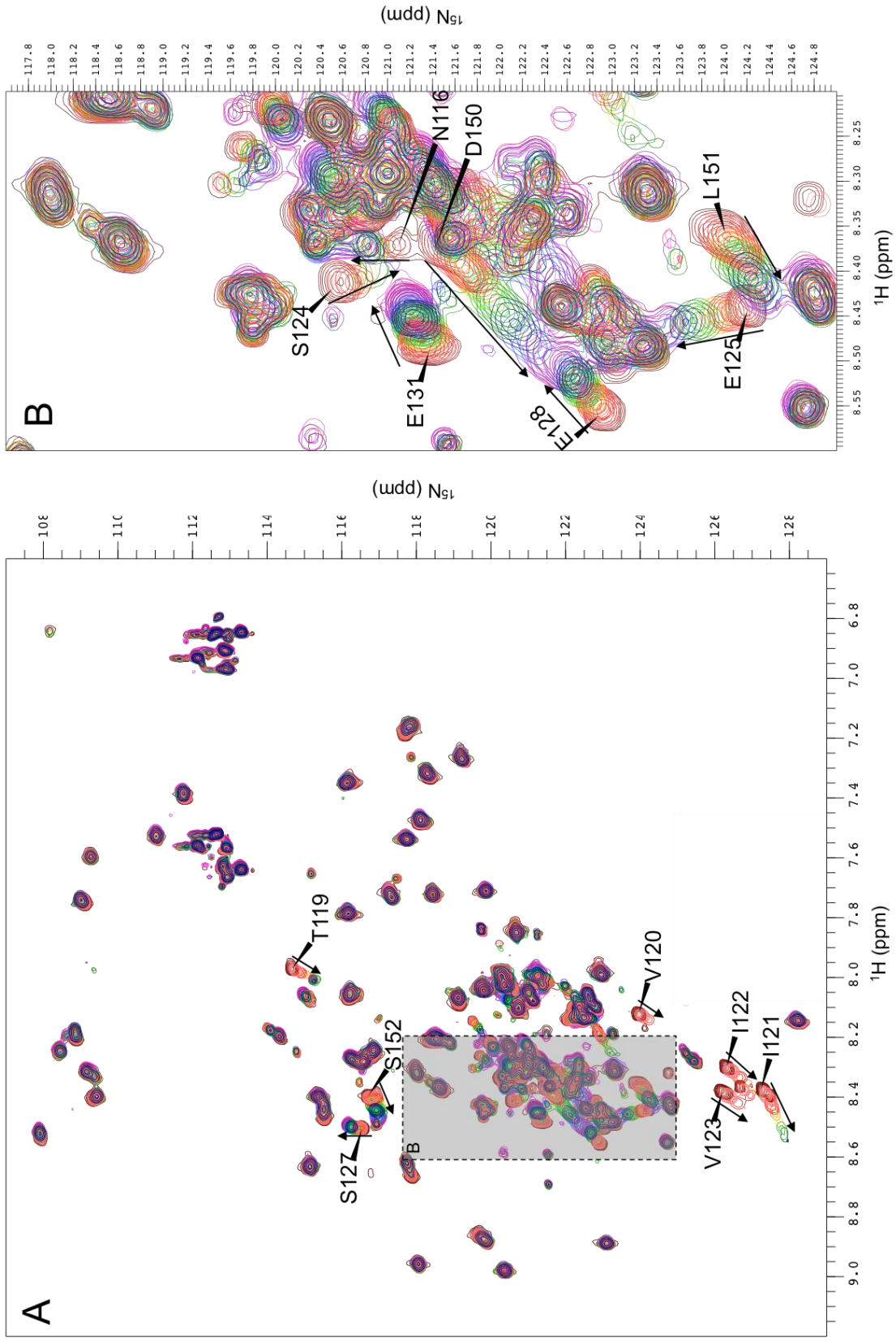



**Figure 4.1 NMR titration experiment of $^{15}$N eCUE1 with unlabelled SUMO1 (Legend on previous page)**

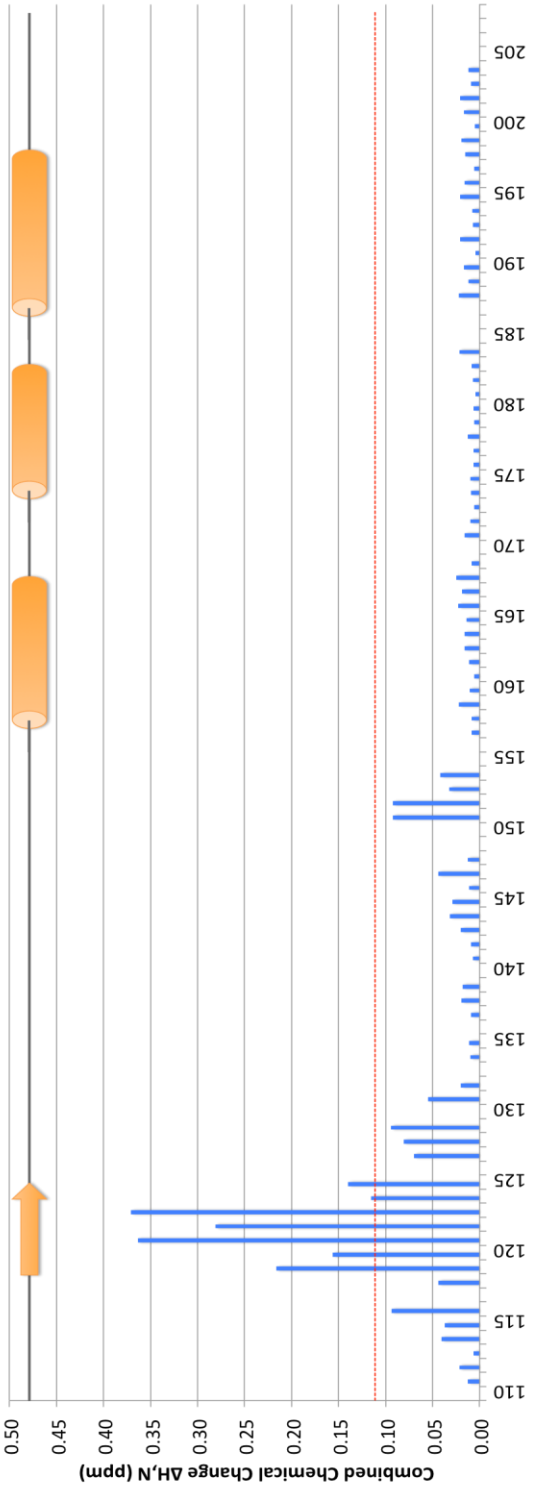



**Figure 4.2 Chemical shift perturbations for SMARCAD1 SIM/CUE ($^{15}$N eCUE1) bound to a 50-fold excess of SUMO1.** Maximum chemical shift change for all assigned amide resonances in the bound state, except Prolines (Pro126, Pro136, Pro170, Pro205). Dashed line corresponds to significance threshold of one S.D. greater than the mean.

The interaction was further characterized by determining a quantitative dissociation constant by independently fitting to 1:1 binding isotherms the chemical shifts from residues with significant perturbations that were simultaneously distinct and unambiguously assigned throughout the titration experiment (Thr119, Ile122 and Glu125), resulting in an average dissociation constant (Kd) of 1.3±0.2mM, as presented in Figure 4.3. This reveals a rather weak interaction, even among similarly characterized SIM sequences [74, 175]. One of the likely factors negatively affecting the dissociation constant is the ionic strength of the system, given the well-established role that the acidic residues neighbouring the SIM hydrophobic core have in stabilizing and orienting the elongated SIM along the SUMO1 groove acting as binding surface. In similar buffer conditions (20mM $NaH_2PO_2$ pH7.0, 100mM NaCl), and for a similarly sequenced SIM (DAXX SIM-N, Figure 3.3), an increase from no salt to 100mM KCl led to a >8-fold decrease in affinity [104].

Additionally, the SIM sequence is likely to simultaneously bind to the SUMO1 groove in a parallel and antiparallel orientation relative to SUMO1 β-sheet 2, as has been shown to be the case for another weaker binding SIM, DAXX SIM-C [104]. Though there is no direct evidence for such conformational variability (intermolecular NOE restraints, for instance), SMARCAD1 SIM/CUE1 exhibits several sequence features that suggest it is potentially binding with no orientation preference such as its symmetrical hydrophobic sequence ($V_{120}I_{121}I_{122}V_{123}$), free of bulky or charged residues present in other SIMs that limit the conformational space available to establish satisfactory hydrophobic non-covalent interactions. Furthermore, there is a symmetrical distribution of polar and acidic charges in the immediate vicinity of the SIM core ($D_{118}T_{119}$ and $S_{124}E_{125}$), which generally favours salt bridge contacts with basic residues in the SUMO1 surface.

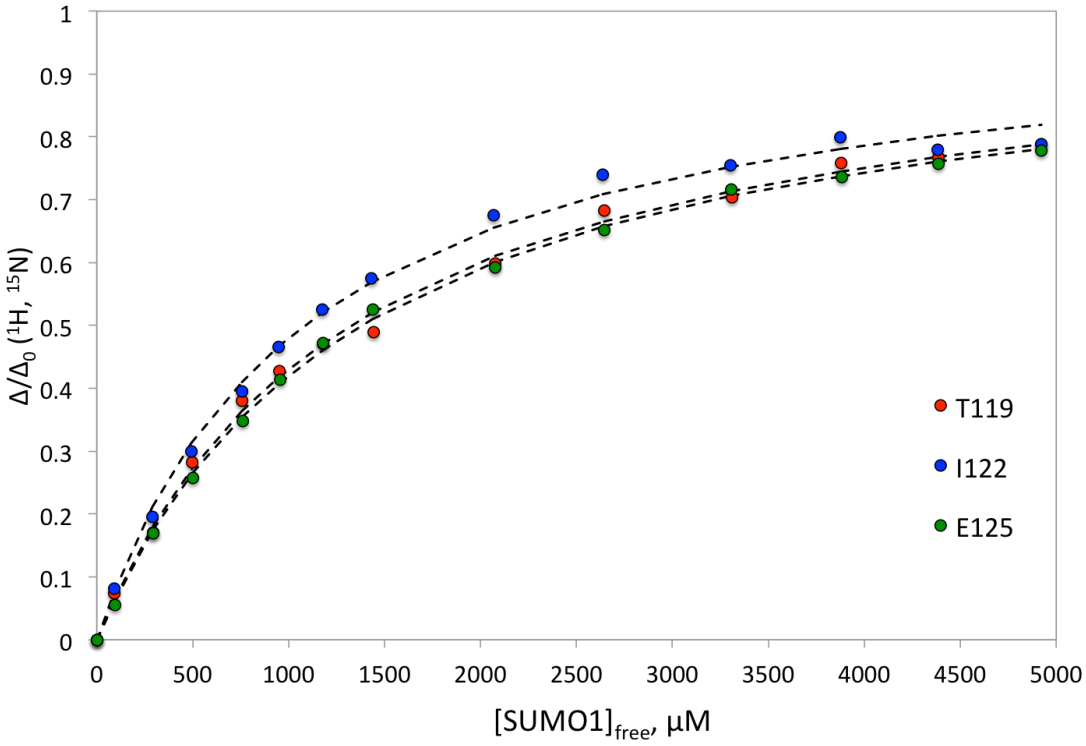



**Figure 4.3 Relative chemical shift changes from SUMO1 and SMARCAD1 SIM/CUE1 interaction.** Analysis of chemical shifts changes for three eCUE1 residues upon titration with SUMO1. In each case an individual dissociation constant (Kd) was determined using an iterative process of non-linear curve fitting (SigmaPlot from Systat Software Inc).

4.2 Potential phosphorylation mediated regulation of SIM-SUMO1 interaction.

To examine the previously reported possibility that SMARCAD1 is targeted by ATM/ATR kinases following DNA damage at Serine residues in or around the SIM (Ser124, Ser127 and S132), a series of NMR titrations were carried out with peptide sequences alternatively phosphorylated on each one of the potential phospho-sites. The argument for phosphorylation-mediated regulation of SIM interactions is well established, with a number of SIM known to be upregulated upon kinase signalling [101, 103, 176-178], notably the transcriptional regulator Daxx [103] or the BRCA1 complex component RAP80 [90]. Mechanistically, phosphorylation enhances the overall negative charge distribution in the sequences flanking the hydrophobic core,

introducing a regulatory switch that may potentially modify substrate specificity, cellular localisation, and signalling pathways.

Following findings from Shiloh, Gygi and Elledge [66] that identify SMARCAD1 as a downstream phosphorylation target from ionizing radiation induced DNA damage, a closer study of the region where all three potential phosphorylation sites are clustered reveals not only the canonical ATM/ATR recognition sequence, $S_{132}Q_{133}$ [66, 179, 180], but two CK2 phosphorylation sites following the established recognition sequence (S/T)XX(E/D) [181] with $S_{127}E_{128}D_{129}E_{130}$ as a potentially canonical target and $S_{124}E_{125}P_{126}pS_{127}$ acting as a hierarchical phosphorylation event. However, all three target Serine residues are reported to be modified downstream of DNA damage and thus an individualised approach of single phospho-sites on synthetic peptides was employed to elucidate the effect of each site on SUMO1 binding.

NMR titration experiments were performed as described in Chapter 2, maintaining a constant uniformly labelled $^{15}$N-SUMO1 concentration and mixing unlabelled peptide ligands up to a maximum concentration of ~3mM. 2D $^{1}$H,$^{15}$N-HSQC spectra were collected for each titration point and the relative chemical shift changes were followed and plotted against free ligand concentration. Figures 4.4 through 4.7 show each overlaid titration set with expanded regions for strongly perturbed resonances.

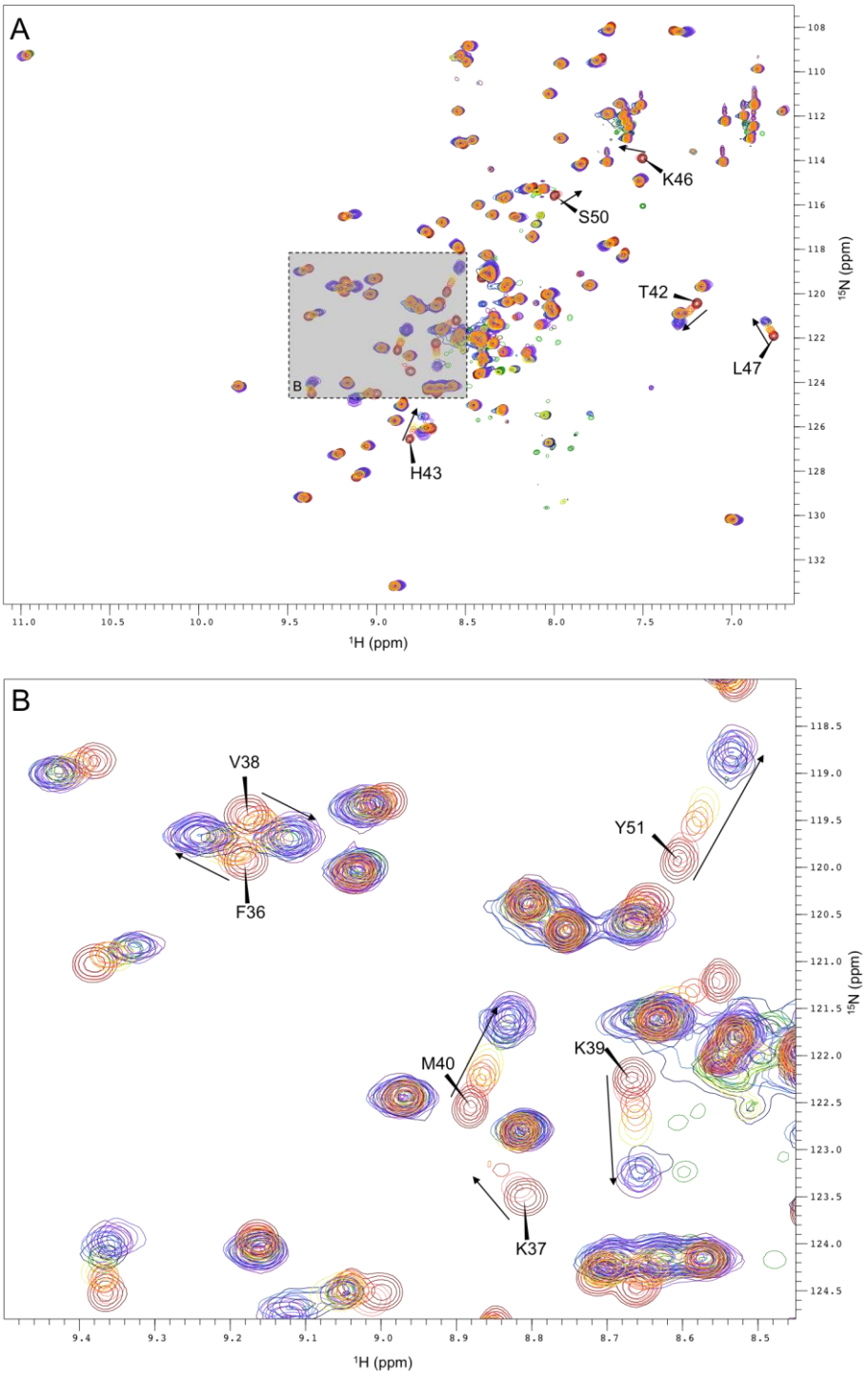



**Figure 4.4 NMR titration experiment of $^{15}$N SUMO1 with unlabelled SMARCAD1 SIM peptide without modified phosphorylation sites.** A) Overlay of thirteen $^{1}$H,$^{15}$N HSQC spectra for SUMO1, each with different, colour coded unphosphorylated peptide (EKTFNKDTVIIVSEDSEDEESQGLPY) concentrations. B) Selection of magnified $^{1}$H,$^{15}$N HSQC peaks that are significantly shifted with increasing ligand concentration.

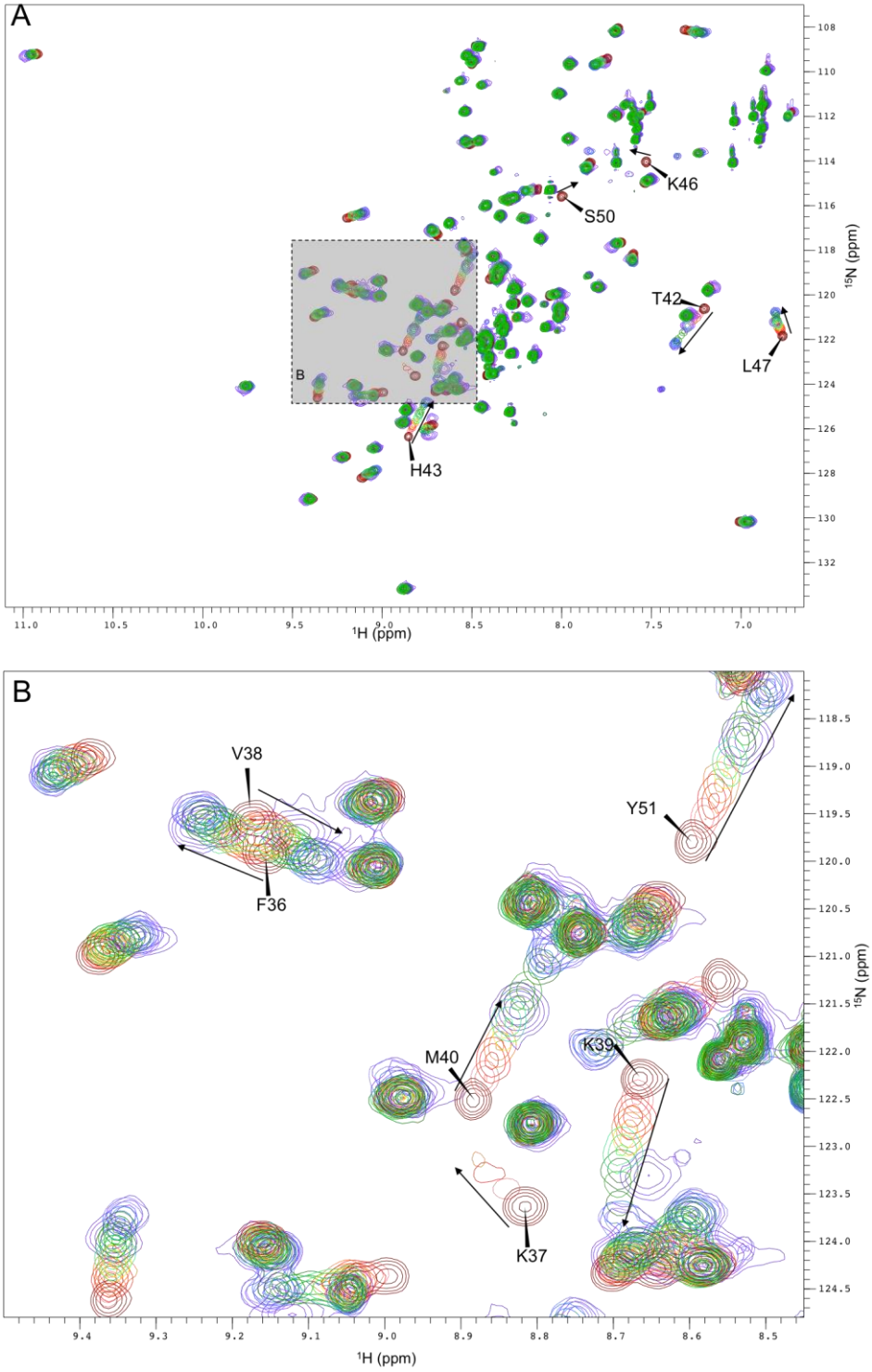



**Figure 4.5 NMR titration experiment of $^{15}$N SUMO1 with unlabelled SMARCAD1 SIM peptide phosphorylated on Ser124.** A) Overlay of twelve $^{1}$H,$^{15}$N HSQC spectra for SUMO1, each with different, colour coded pSer124 peptide (EKTFNKDTVIIVpSEDSEDEESQGLPY) concentrations. B) Selection of magnified $^{1}$H,$^{15}$N HSQC peaks that are significantly shifted with increasing ligand concentration.

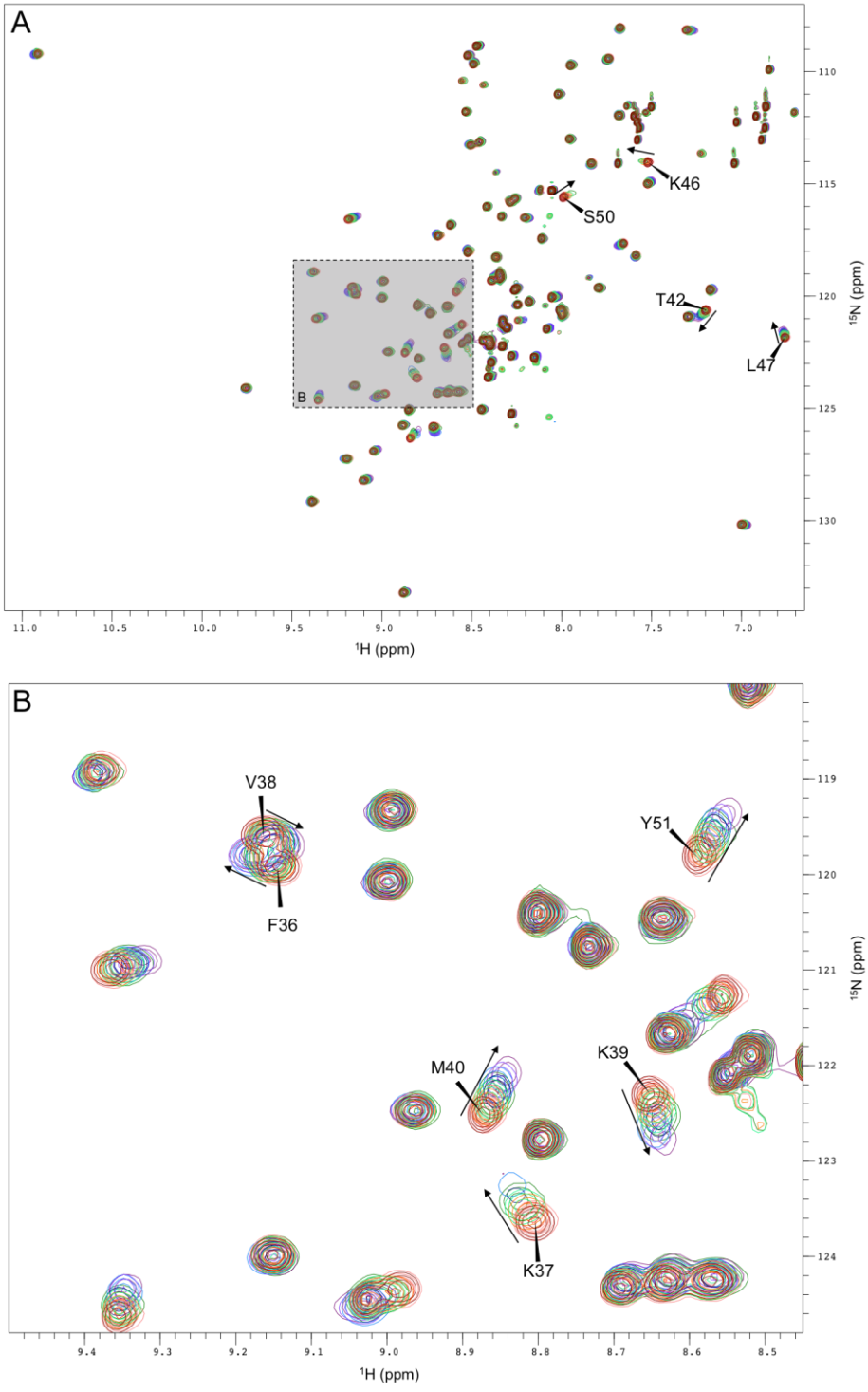



**Figure 4.6 NMR titration experiment of $^{15}$N SUMO1 with unlabelled SMARCAD1 SIM peptide phosphorylated on Ser127.** A) Overlay of ten $^{1}$H,$^{15}$N HSQC spectra for SUMO1, each with different, colour coded pSer127 peptide (EKTFNKDTVIIVSEDpSEDEESQGLPY) concentrations. B) Selection of magnified $^{1}$H,$^{15}$N HSQC peaks that are significantly shifted with increasing ligand concentration.

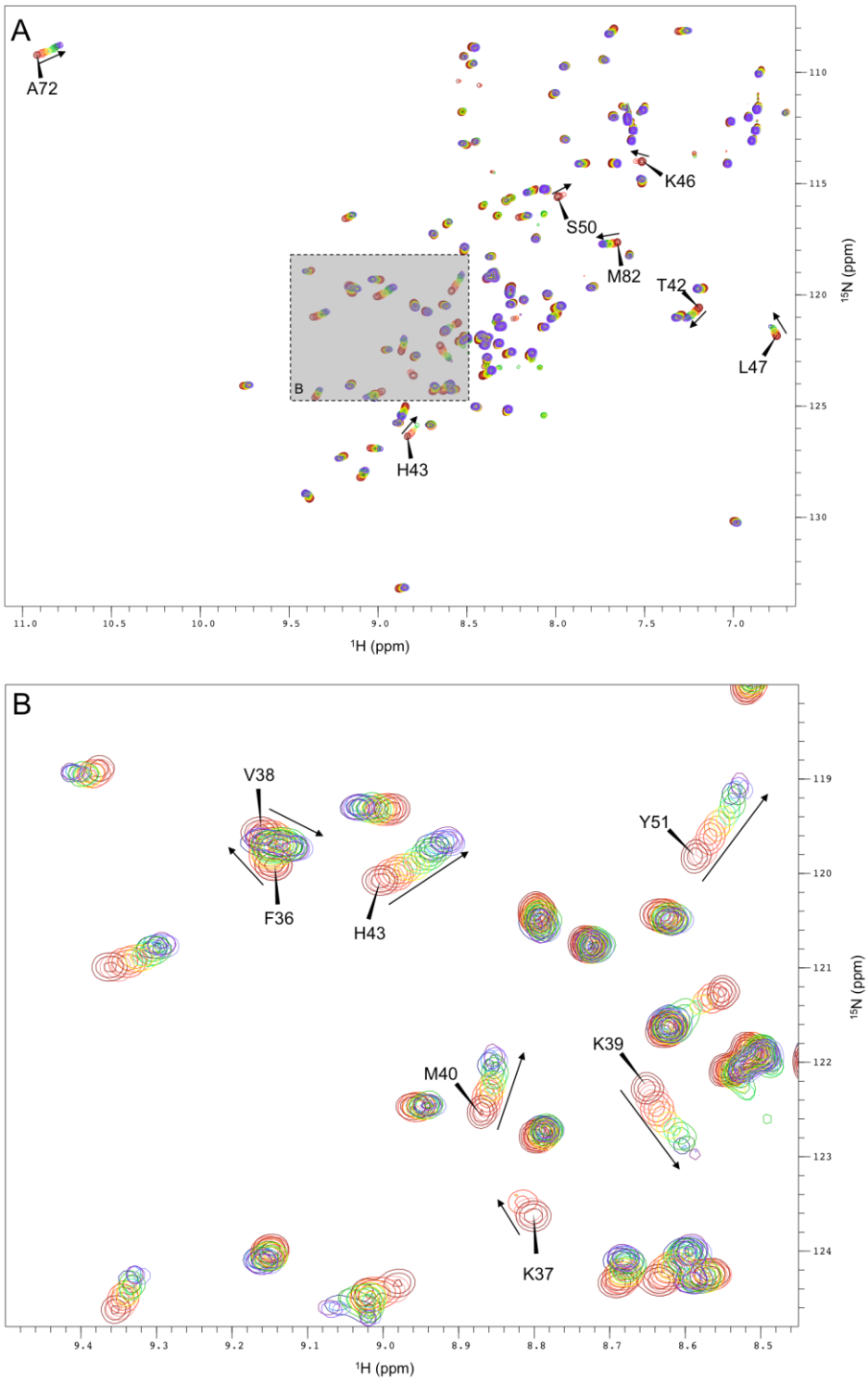



**Figure 4.7 NMR titration experiment of $^{15}$N SUMO1 with unlabelled SMARCAD1 SIM peptide phosphorylated on Ser132.** A) Overlay of thirteen $^{1}$H,$^{15}$N HSQC spectra for SUMO1, each with different, colour coded pSer132 peptide (EKTFNKDTVIIVSEDSEDEEpSQGLPY) concentrations. B) Selection of magnified $^{1}$H,$^{15}$N HSQC peaks that are significantly shifted with increasing ligand concentration.

With the sole exception of perturbed resonances upon pSer124 peptide titration, all shifts exhibit fast exchange behaviour in solution. The individual chemical shift perturbations for pSer127 are considerably smaller than those for the unphosphorylated peptide, suggesting that this modification might not favour SUMO1 as a substrate following signalling events. Similarly, perturbations are somewhat smaller upon pSer132 binding, though less so. Furthermore, distinct additional resonances are specifically perturbed in this system, including Arg63, Ala72 and Met82. Finally, binding of pSer124 presents tighter, stronger interactions, even though residues still exhibit linear, fast exchange perturbations. Intriguingly, Lys 46 and Ser50, both localised in α-helix 1 of SUMO1, become broader at lower ligand concentrations compared to the unphosphorylated peptide and, exclusively for this system, at high ligand concentration (~10-fold excess) the perturbed resonances are once again observable, though with a different chemical shift, indicating an overall tighter interaction characterised by intermediate chemical exchange.

Chemical shift perturbation maps of the interaction with alternatively phosphorylated SMARCAD1 SIM peptides at the highest total ligand concentration (~3mM peptide, 10-fold excess) provide a clear illustration of the effect of ligand binding on all resonances throughout the SUMO1 sequence for each individual phosphorylated Serine. As appreciated in Figure 4.8, the bulk of significantly perturbed peaks are clustered between residues Lys47 through Tyr51, which map to the expected groove between α-helix 1 and β-sheet 2 in SUMO1, with a weaker (though larger-than-average) perturbation in the sequences spanning residues Tyr21 through Leu24 which correspond to β-sheet 1 in SUMO1. These are all canonical perturbations, well established for SIM-SUMO1 interactions [74, 175].

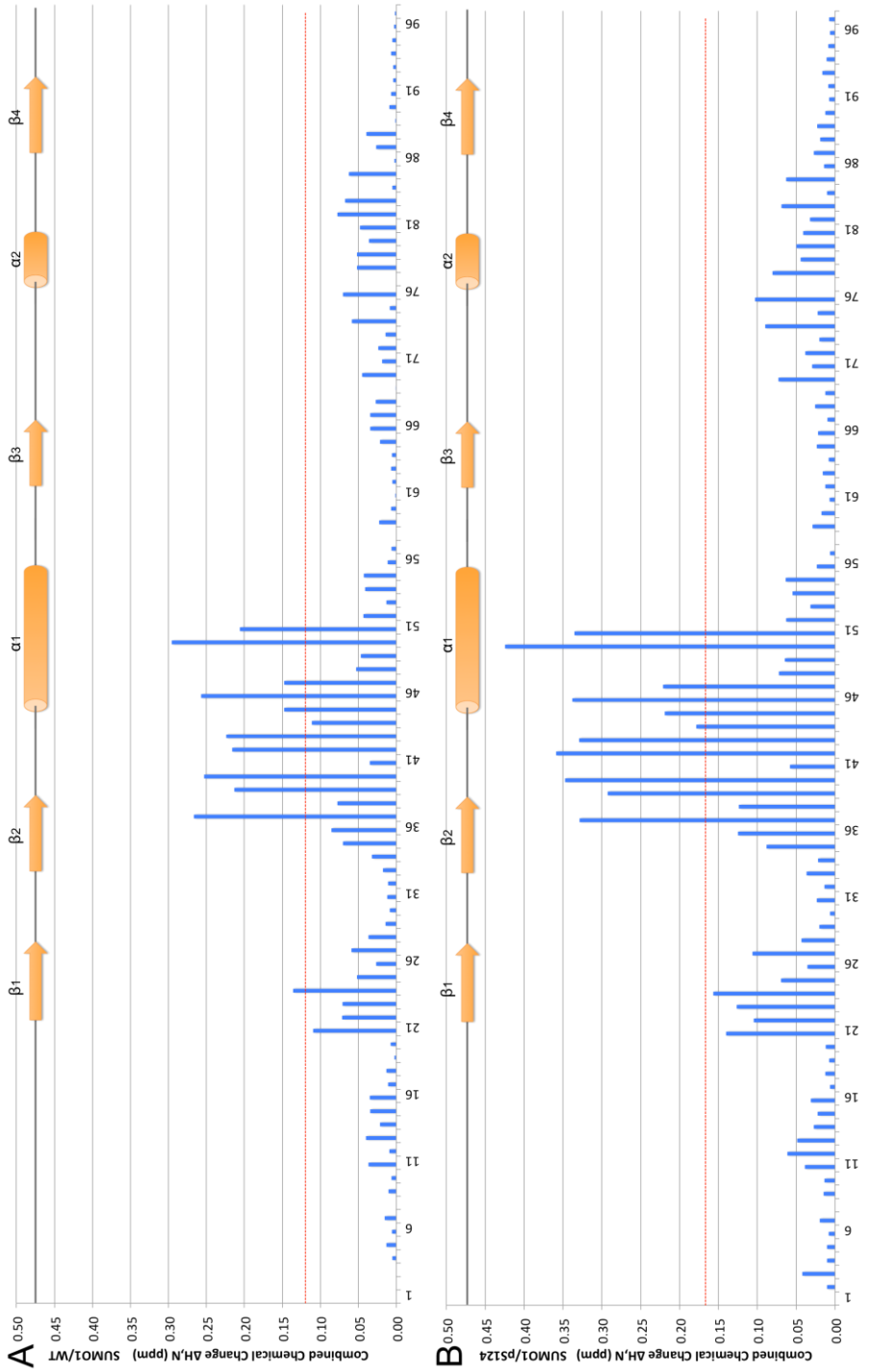



**Figure 4.8 Chemical shift perturbations for $^{15}$N SUMO1 bound to a 10-fold excess of SMARCAD1 SIM differentially phosphorylated peptides.** Maximum chemical shift change for all assigned amide resonance in the bound state, except Prolines (Pro8, Pro58, Pro77). Dashed lines correspond to individual significance threshold of one S.D. greater than the mean. A) unphosphorylated SIM peptide, B) pSer124 monophosphorylated SIM peptide.

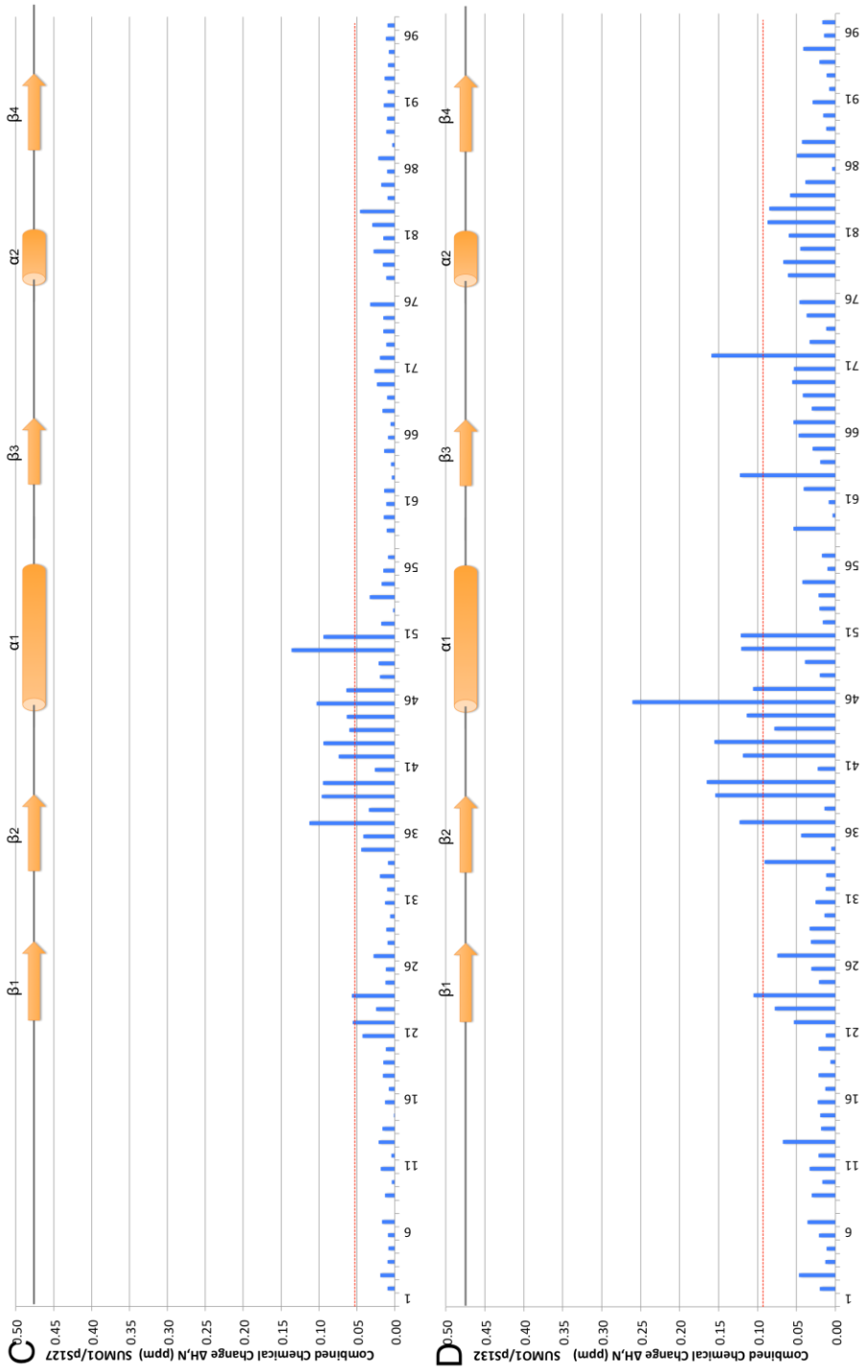



**Figure 4.8 (Continued) Chemical shift perturbations for $^{15}$N SUMO1 bound to a 10-fold excess of SMARCAD1 SIM differentially phosphorylated peptides.** Maximum chemical shift change for all assigned amide resonance in the bound state, except Prolines (Pro8, Pro58, Pro77). Dashed lines correspond to individual significance threshold of one S.D. greater than the mean. C) pSer127 monophosphorylated SIM peptide, D) pSer132 monophosphorylated SIM peptide.

It becomes immediately apparent on analysing Figure 4.8 that pSer124 SIM peptide interaction is uniformly stronger than that of other peptides in the experiment (See Figure 3.3 for reference to SUMO1 binding surface). Extrapolating the potential conformational fluctuations of unmodified SIM binding to SUMO1, it appears that this phosphorylation event reduces any such ambiguity. Indeed, the sequence now bears a closer resemblance to that of RanBP2 M-IR2 SIM whose high affinity interaction with SUMO1 has been mapped in detail [182], and it is known to establish an antiparallel extended conformation relative to SUMO1 β-sheet 2. Moreover, by fixing the orientation of the SIM peptide, it may be inferred that the extended sequence bearing the two supplementary phosphorylated Serines would contact the opposite SUMO1 surface, along the flexible linker connecting β-sheets 3 and 4 to α-helix 2. In agreement with this analysis, the additional SUMO1 resonances significantly perturbed upon pSer132 binding, Arg63 and Ala72, are close together in space along an axis parallel to the canonical SIM grove. Finally, if such a binding mode is considered, the projected binding surface for pSer127 doesn't present many basic residues to form stabilizing salt bridges, and without a fixed conformational orientation through SIM core phosphorylation of $S_{124}$, this likely leads to the distinctly poor interaction observed.

These interactions were further characterized by determining a quantitative dissociation constant (Table 4.1) by independently fitting to 1:1 binding isotherms the chemical shifts from residues with significant perturbations that were simultaneously distinct and unambiguously assigned throughout each titration experiment, as presented in Figures 4.9 though 4.12. No dissociation constant may be inferred from pSer127 binding other than to suggest it is larger than the maximum ligand concentration in these assays (>3mM).

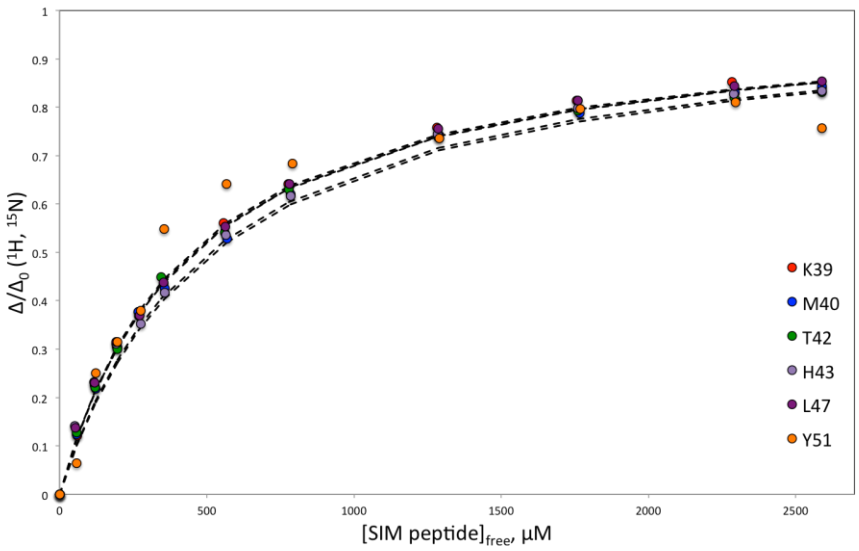



**Figure 4.9 Relative chemical shift changes from SUMO1 and unmodified SMARCAD1 SIM peptide interaction.** Analysis of chemical shifts changes for six SUMO1 residues upon titration with unphosphorylated SIM peptide. In each case an individual dissociation constant (Kd) was determined using an iterative process of non-linear curve fitting (SigmaPlot from Systat Software Inc.).

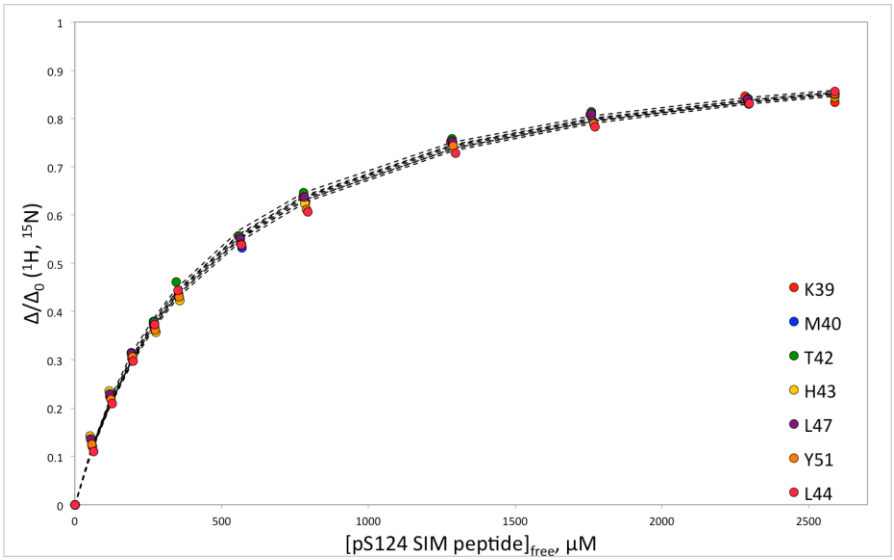



**Figure 4.10 Relative chemical shift changes from SUMO1 and pSer124 SMARCAD1 SIM peptide interaction.** Analysis of chemical shifts changes for seven SUMO1 residues upon titration with pSer124 SIM peptide. In each case an individual dissociation constants (Kd) were determined using an iterative process of non-linear curve fitting (SigmaPlot from Systat Software Inc.).

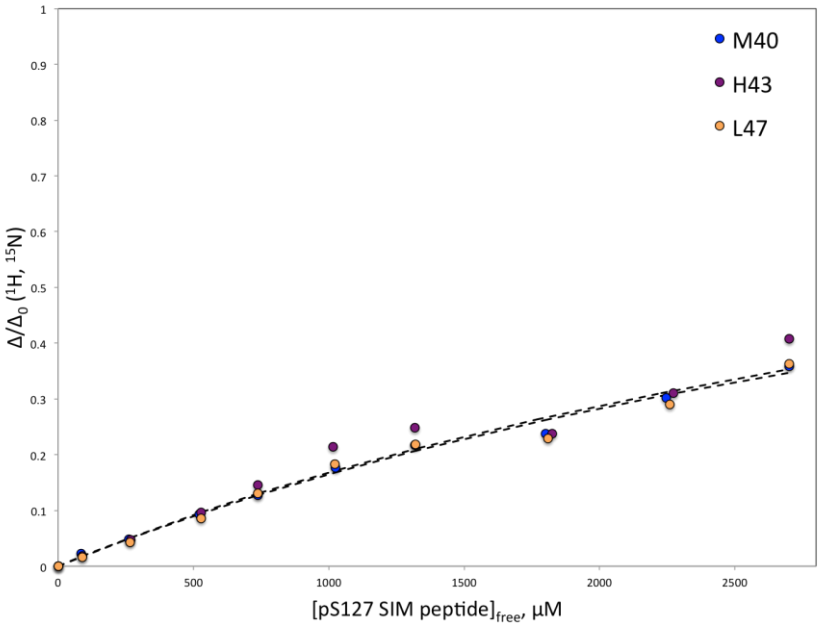



**Figure 4.11 Relative chemical shift changes from SUMO1 and pSer127 SMARCAD1 SIM peptide interaction.** Analysis of chemical shifts changes for three representative SUMO1 residues upon titration with pSer127 SIM peptide.

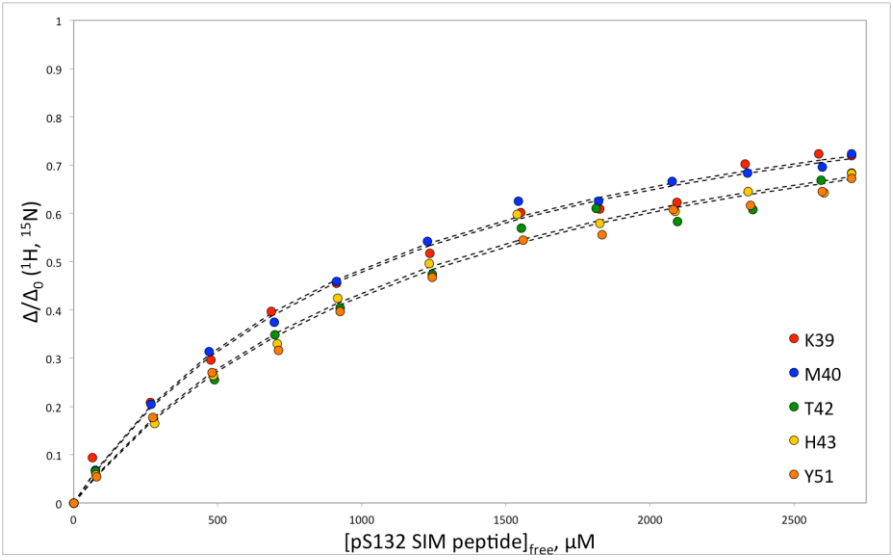



**Figure 4.12 Relative chemical shift changes from SUMO1 and pSer132 SMARCAD1 SIM peptide interaction.** Analysis of chemical shifts changes for five SUMO1 residues upon titration with pSer132 SIM peptide. In each case an individual dissociation constant (Kd) was determined using an iterative process of non-linear curve fitting (SigmaPlot from Systat Software Inc.).

**Table 4.1 Dissociation constants for SUMO1 interaction with synthetic SAMARCAD1 SIM peptides, either unmodified or mono-phosphorylated at S124, S127 or S132.**

| | SUMO1 Kd binding |
|---|---|
| Unmodified SIM peptide | (1.0±0.1)mM |
| pSer124 SIM peptide | (490±15)μM |
| pSer127 SIM peptide | ≈6mM |
| pSer132 SIM peptide | (1.2±0.1)mM |

## 4.3 Mass spectrometry analysis of SMARCAD1 phosphorylation state upon chemically induced DNA damage

Based on the differing effects that phosphorylation of Ser127 seems to have against that on Ser124 and Ser132 during SUMO1 binding, mass spectrometry analysis of post-translational phosphorylation of SMARCAD1 following chemically induced DNA damage was performed. For these, both endogenous SMARCAD1 and transiently expressed GFP fusion protein were purified from cultured human cells alternatively exposed to neocarzinostatin (NCS) as a DNA damage inducing chemical reagent, as described in Chapter 2. Figure 4.13 presents a representative western blot of the pull-down experiment for SMARCAD1 purification under stringent conditions (high ionic strength and detergent concentration) to reduce non-specific or transient interactions.

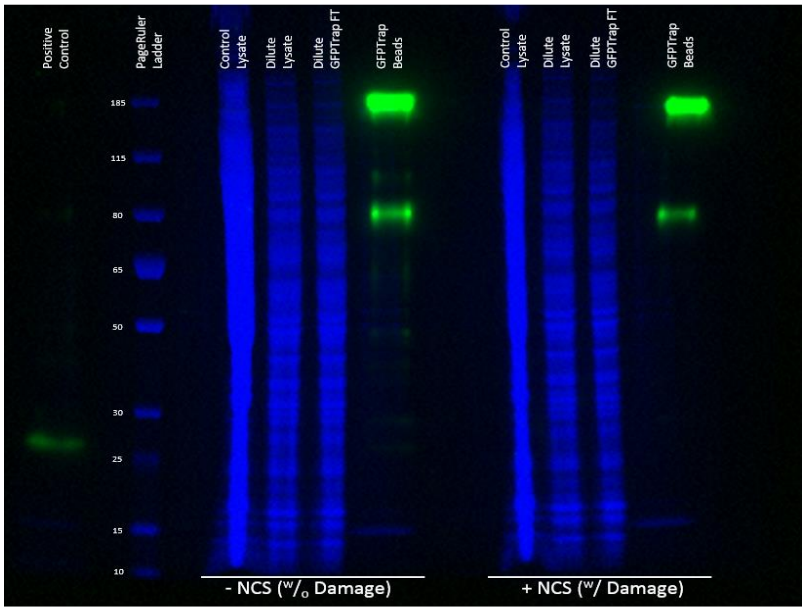



**Figure 4.13 Pull down purification of transiently expressed recombinant GFP-SMARCAD1 from HEK 293T cells alternatively exposed to chemically induced DNA damage.** A) Positive control, B) Molecular weight ladder, C) Untreated whole cell lysate, D) Untreated whole cell lysate diluted 1:3, E) Untreated pull-down flowthrough, F) GFP-Trap beads with immunoprecipitated GFP-SMARCAD1 from untreated cells, G) NCS treated whole cell lysate, D) NCS treated whole cell lysate diluted 1:3, E) NCS treated pull-down flowthrough, F) GFP-Trap beads with immunoprecipitated GFP-SMARCAD1 from NCS treated cells. Coomassie Brilliant Blue and Goat anti-GFP/Mouse anti-Goat/HRP enhanced chemiluminescence signals in blue and green, respectively.

Immunoprecipitated proteins were treated with trypsin digestion as described in Chapter 2, and analysed through nanoLC-MS/MS by the Kessler Group in the Mass Spectrometry Lab at the Target Discovery Institute, University of Oxford. Post-translational modifications were identified using the database and *de novo* search program PEAKS version 7 (Bioinformatics Solutions). Figure 4.14 shows complete sequence coverage and high redundancy of detected tryptic peptides for the SIM region, both with and without NCS induced damage. There is no difference in the phosphorylation pattern observed, suggesting no differentially induced phosphorylation event upon DNA damage.

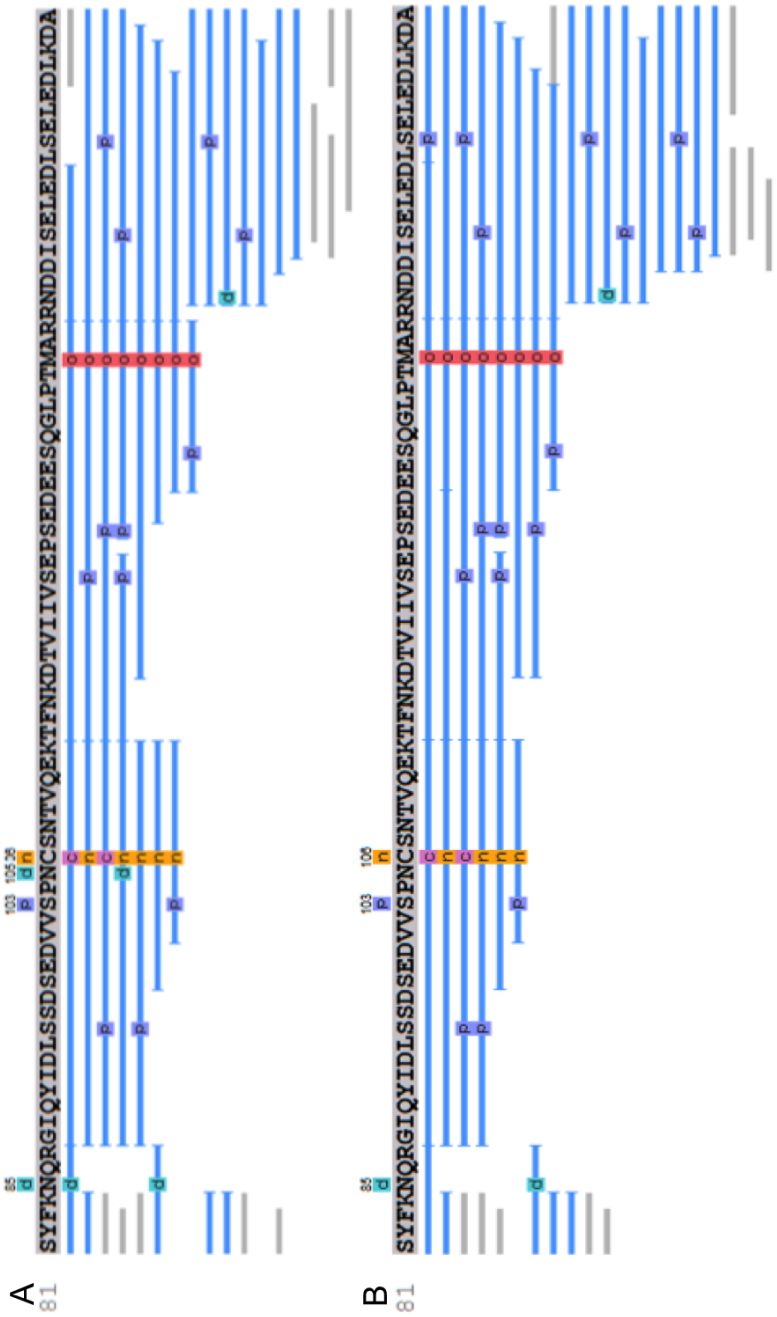



**Figure 4.14 Phosphorylation profiling of SMARCAD1 differentially exposed to chemically induced DNA damage.** Tandem mass spectrometry analysis of SMARCAD1 tryptic digest and detection of phosphorylated peptides was performed on samples without NCS induced DNA damage (A) or after 30-60min with 100µg/mL on NCS at 37°C. B) Multiple peptides covering the SIM region were detected, however no differential phosphorylation profile is evident. d: Asparagine deamidation, n: Cysteine carbamidomethylation, o: Methionine oxidation, p: Serine/Threonine/Tyrosine phosphorylation.

These results are consistent with the findings from Shiloh, Gygi and Elledge [66], though no triply phosphorylated peptides are detected. It is possible that such species are poorly represented in the pool of analysed peptides due to the well established increased in hydrophilicity of peptides bearing multiple phosphorylations, particularly if no basic residues are present in the sequence to compensate for these new charges through internal salt bridges [183]. Such increase in peptide hydrophilicity leads to poly-phosphorylated peptides with greatly decreased retention in the reverse phase chromatography column, commonly coupled to mass spectrometry ESI analysis.

It must be noted, however, that these assays were not performed with synchronised cell populations, and thus the effect of NCS treatment had through the rapid formation of double strand breaks would lead to alternative signalling cascades depending on the individual cell cycle stage (HR for cells in S phase, NHEJ for all other phases). Such stochastic averaging of a potential signalling event specific for ATM or ATR after DNA damage might account for the lack of perceived differences despite treatment.

## 4.4 SUMO1 interaction with a SMARCAD1 SIM phosphomimic.

Having established the constitutive phosphorylation state of SMARCAD1 in the extended vicinity of the SUMO interacting motif spanning Ser124, Ser127 and Ser132, a new titration experiment was performed this time with all three phospho-sites replaced by structural mimetic Glutamate residues. Such modifications, introducing a close approximation of the length and charge distribution of native phosphorylation events, are frequently used to probe structure-function relationships when *in vitro* phosphorylation of target proteins is not possible [184]. 2D $^{1}H,^{15}N$-HSQC spectra were collected for each titration point and the relative chemical shift changes were followed and plotted against SUMO1 concentration while maintaining the triple phosphomimetic SIM mutant (SIM $S_3E$) concentration constant.

Figure 4.15 shows an overlay of $^{1}H,^{15}N$ HSQC spectra with varying concentrations of unlabelled SUMO1. When comparing with Figure 4.1 it becomes instantly clear that perturbed resonances exhibit larger shift steps with increasing ligand concentration and those in the fast chemical exchange regime reached saturation. Most striking, the hydrophobic section of the SIM immediately becomes too broad for detection at even the lowest ligand titration point characteristic of intermediate chemical exchange regime, though they do not reappear even at the highest ligand concentration (~50-fold excess).

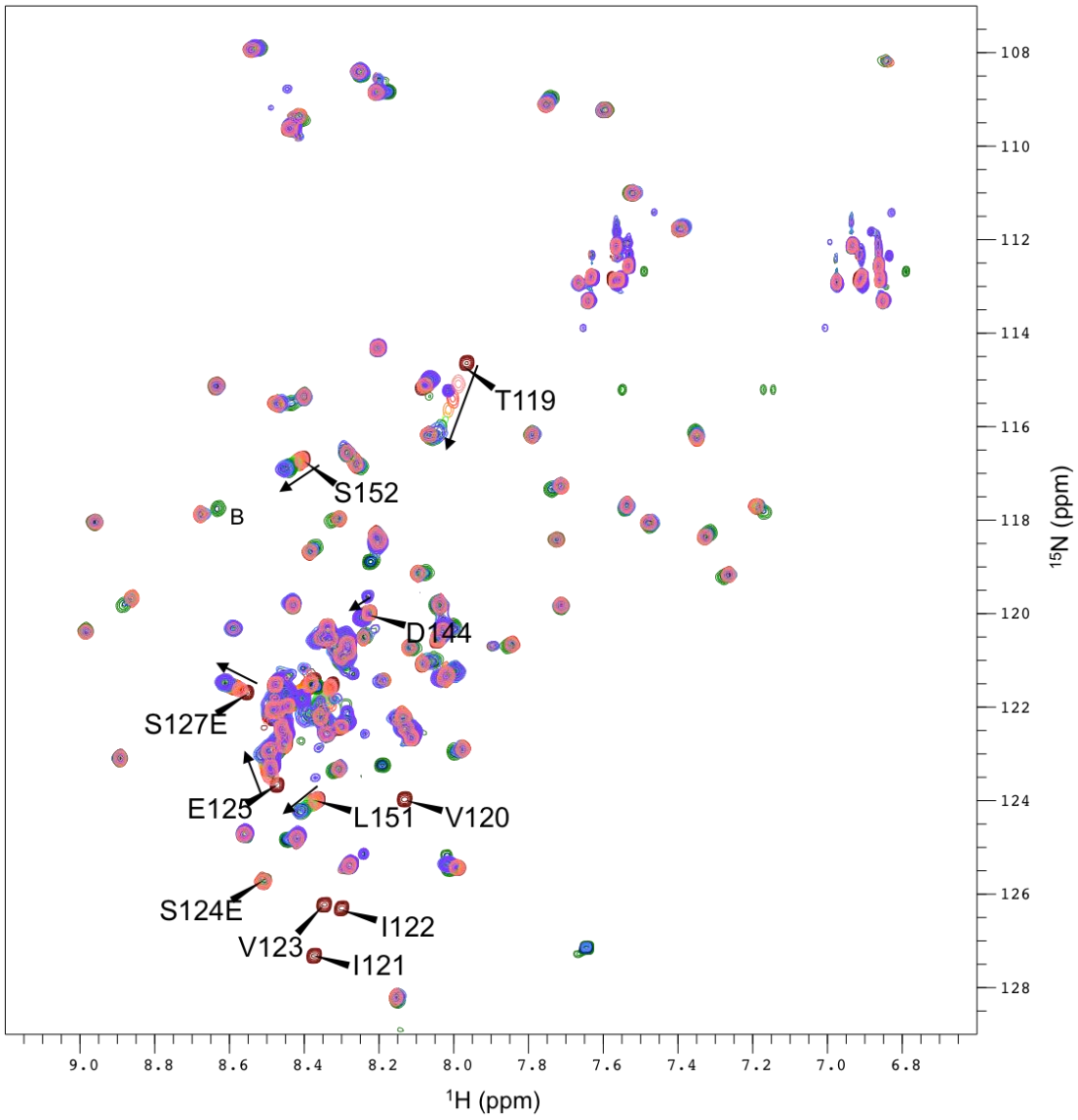



**Figure 4.15 NMR titration experiment of $^{15}N$ eCUE1/$S_3E$ with unlabelled SUMO1 (Legend on next page).**

**Figure 4.15 (Previous page) NMR titration experiment of $^{15}$N eCUE1/$S_3$E with unlabelled SUMO1.** Overlay of thirteen $^{1}$H,$^{15}$N HSQC spectra for eCUE1/$S_3$E, each with different, colour coded SUMO1 concentrations. A selection of peaks that shifted considerably is identified and the sign and magnitude of each shift is consistent with increasing ligand concentration.

Figure 4.16, a chemical shift perturbation map of the interaction with SUMO1 at the highest total ligand concentration (5mM SUMO1, 50-fold excess), once again illustrates the overall tighter interaction which cannot simply be attributed to Ser124Glu phosphomimic mutagenesis, given that both Ser127Glu and Ser132Glu exhibit a larger chemical shift perturbation than that observed for the native sequence. Furthermore, resonances previously perturbed weakly are now significantly shifted (Glu149 through Ser152) demonstrating an increase in ligand affinity when simultaneously including all three phosphorylated Serines in the extended SIM sequence. Two important features of the titration experiment are highlighted in the perturbation map. First, Ser124Glu does not display two dimensional chemical shift changes with increasing ligand concentration; rather its line intensity broadens according to a logarithmic binding isotherm, itself a characteristic of high affinity moderate chemical exchange, reaching a plateau within the ligand concentration range in the experiment. Second, due to sequence degeneracy and chemical shift overlap it is not possible to unequivocally assign Glu128, Glu130, Ser132Glu or Gln133, even with the aid of $^{15}$N-edited TOCSY and NOESY data for sequential assignment. However, at least two of these resonances are strongly perturbed, demonstrating fast chemical exchange rates. Finally, as previously mentioned, the hydrophobic core exhibits extensive line broadening even at 200µM SUMO1 concentration and thus are not included in the perturbation map.

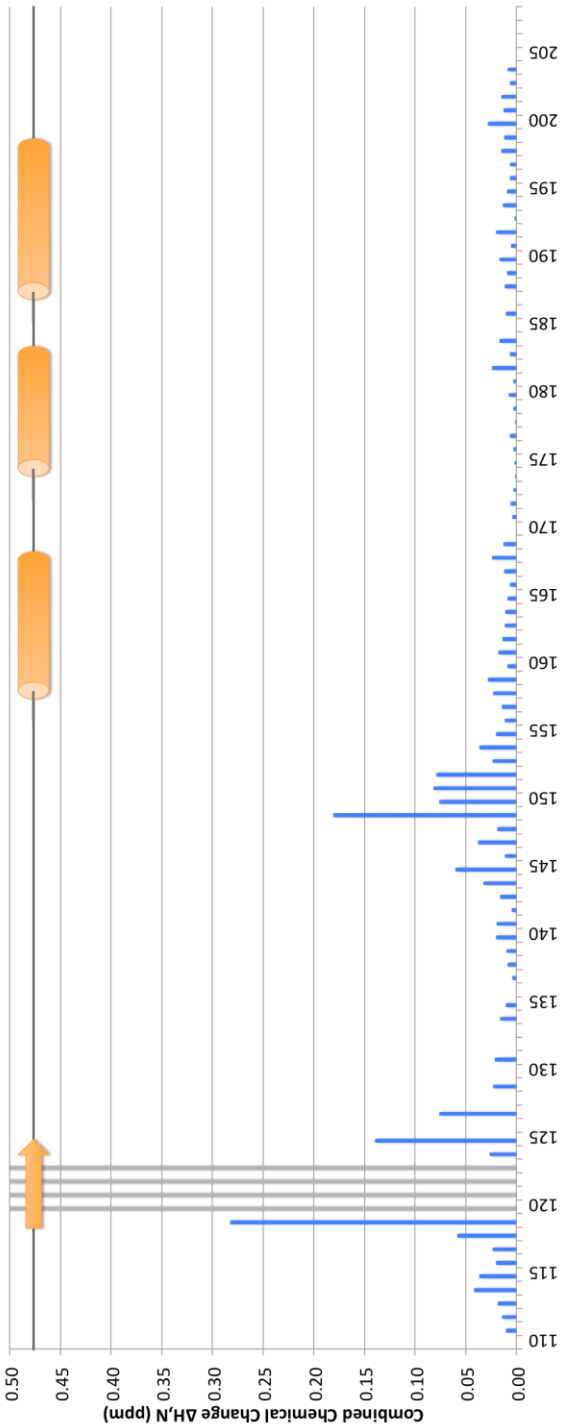



**Figure 4.16 Chemical shift perturbations for SMARCAD1 SIM/CUE1 $S_3E$ phosphomimic ($^{15}$N eCUE1/$S_3E$) bound to a 50-fold excess of SUMO1.** Maximum chemical shift change for all assigned amide resonance in the bound state, except Prolines (Pro126, Pro136, Pro170, Pro205). SIM hydrophobic residues 120 VIIV 123 broadened beyond detection at the lowest ligand concentration used and so hypothetical shifts are labelled in grey as a reference for the relative scale of perturbation throughout the sequence.

Titration data was used to determine a quantitative dissociation constant for the tighter SMARCAD1 SIM/CUE1 $S_3E$ phosphomimic interaction with SUMO1 by independently fitting to 1:1 binding isotherms the chemical shifts from residues with significant perturbations, that were simultaneously distinct and unambiguously assigned throughout the titration experiment. This excludes the hydrophobic core resonances, though polar/charged residues in the immediate vicinity including Thr119, Ser124Glu, and Glu125 likely describe the actual affinity of the SIM core sequence. Figure 4.17 presents the binding isotherms for T119, the only one successfully followed throughout the titration series.

From the perturbation plot it may be inferred that there are potentially three distinct binding modules along the sequence, stemming perhaps from complementary secondary interaction events. Following residues in the core SIM region (Thr119) results in an approximate dissociation constant (Kd) of 76μM, an affinity consistent with the vast majority of reported SIM-SUMO1 interaction, even if still somewhat weak. Couple to this binding event, the acidic patch in the extended SIM region, represented by Ser125, Ser127Glu and all three unassigned Glutamate residues which correspond to Glu128, Glu130 and Ser132Glu, appears to be moderately perturbed. Finally, a secondary binding site (within residues Glu149 to Ser152), closer to the Ubiquitin binding CUE1 domain, may not be quantitatively assessed though it appears to bind rather weakly. It is important to note that these residues have been mildly perturbed in all titration experiments, but their chemical shifts changes had been too small to exceed the significance threshold. In the present case, the larger changes to the chemical shift of this potential binding surface is likely because the relative intensity range is skewed due to the much tighter binding of hydrophobic SIM residues, which broadened upon binding and are not detectable throughout the experiment, which may indicate a transition to slower kinetic exchange rates.

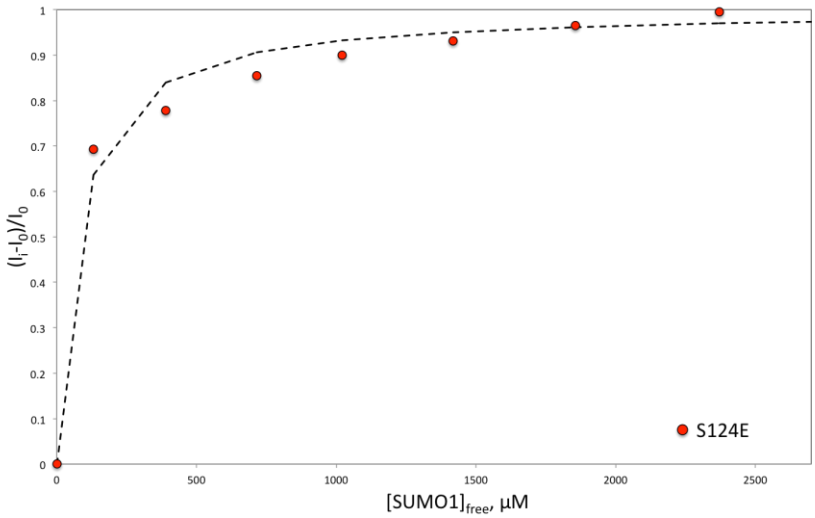



**Figure 4.17 Relative peak height intensity changes from SUMO1 and SMARCAD1 SIM/CUE1 $S_3E$ phosphomimic interaction.** Analysis of peak intensity changes for residue S124E in eCUE1/$S_3E$ upon titration with SUMO1. An individual dissociation constants (Kd) was determined using an iterative process of non-linear curve fitting (SigmaPlot from Systat Software Inc.).

## 4.5 Ubiquitin interaction with SMARCAD1 tandem CUE domain

In addition to SUMO1 binding, the canonical interaction event proposed for SMARCAD1 involves ubiquitin recognition by the tandem CUE domains, which has been previously studied, and revealed to be primarily mediated by CUE1, with CUE2 exhibiting very low affinity for monoubiquitin [135]. In order to verify that the SUMO bearing sequence, previously excluded from functional characterisation of the tandem CUE domain, does not alter this interaction, a new construct including all three domains was analysed for Ubiquitin binding through NMR titrations. Thus a uniformly $^{15}$N-labelled SIM/CUE1/CUE2 SMARCAD1 construct ($^{15}$N edCUE) was titrated with increasing concentrations of unlabelled monoubiquitin up to a 50-fold molar excess.

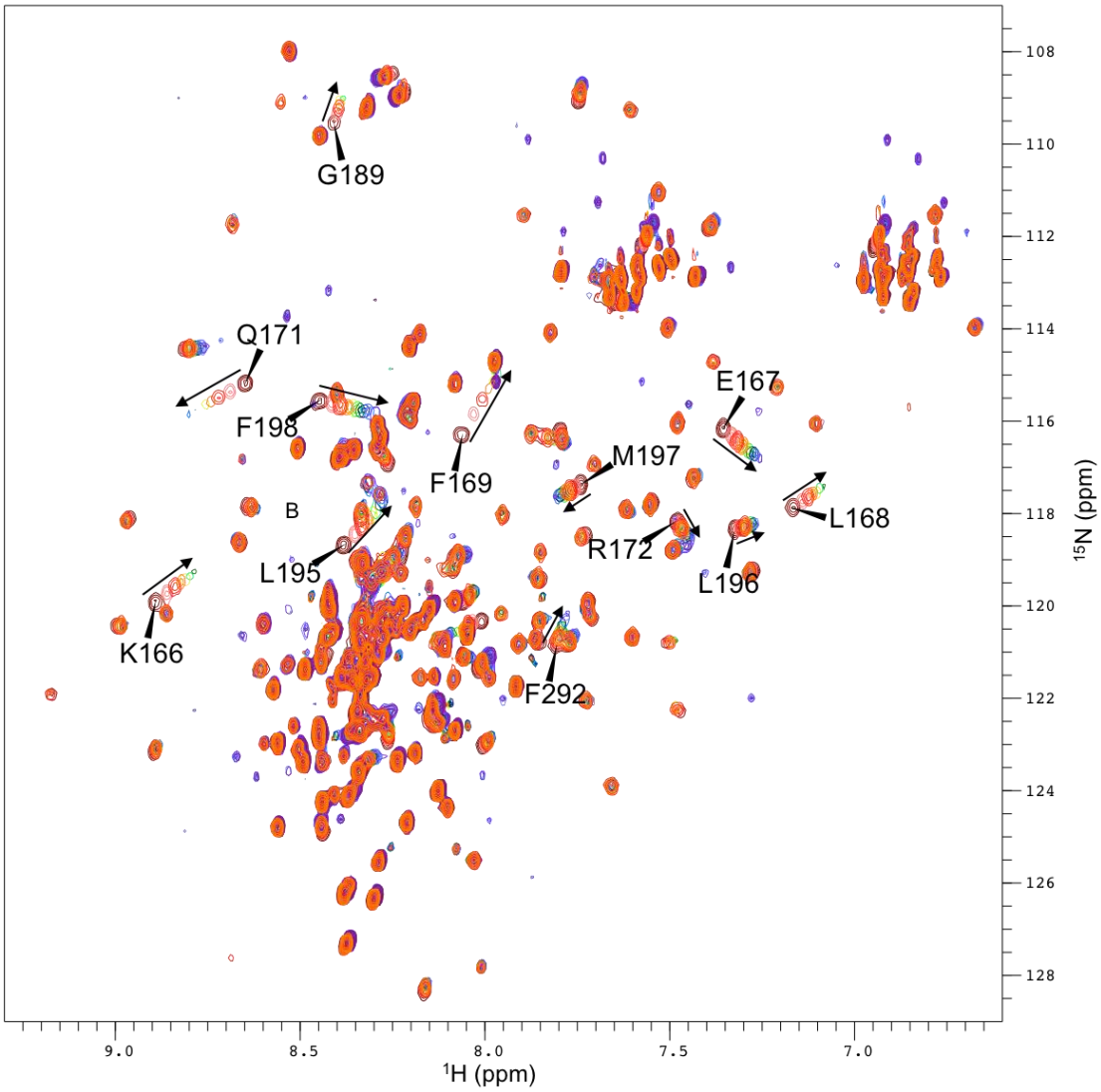



**Figure 4.18 NMR titration experiment of $^{15}$N edCUE with unlabelled Ubiquitin.** Overlay of thirteen $^{1}$H,$^{15}$N HSQC spectra for SMARCAD1 SIM/CUE1/CUE2 edCUE construct, each with different, colour coded Ubiquitin concentrations. A selection of peaks that shifted considerably is identified and the sign and magnitude of each shift is consistent with increasing ligand concentration.

Total chemical shifts perturbations were mapped against the SMARCAD1 sequence, which show in Figure 4.19 the presence of two strongly perturbed sequence tracts in CUE1 corresponding to the surface formed by the end of helix 1, the loop connecting it to helix 2 and the end of helix 3, all in close spatial proximity. Additionally, a second perturbed region in CUE2 is evident, though the extent and intensity is significantly smaller than for CUE1. As expected, no interactions are observed for the SUMO1 binding motif.

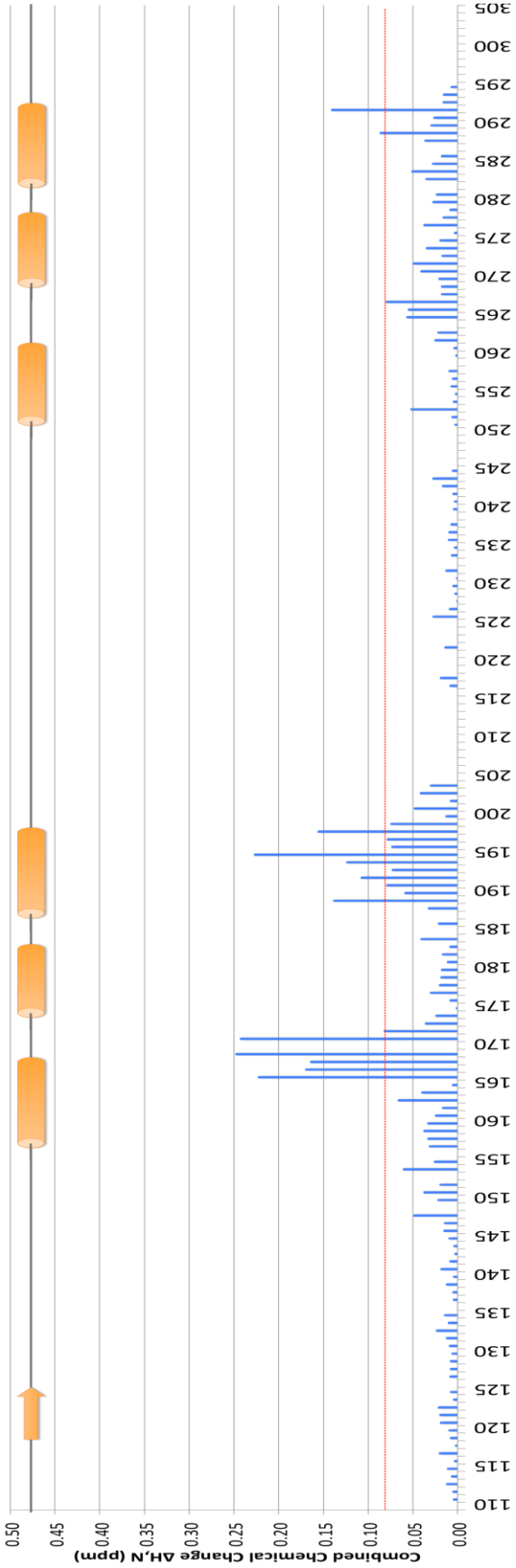



**Figure 4.19 Chemical shift perturbations for SMARCAD1 SIM/CUE1/CUE2 ($^{15}$N edCUE) bound to a 50-fold excess of Ubiquitin (Legend on next page).**

**Figure 4.19 (Previous page) Chemical shift perturbations for SMARCAD1 SIM/CUE1/CUE2 ($^{15}$N edCUE) bound to a 50-fold excess of Ubiquitin** Maximum chemical shift change for all assigned amide resonance in the bound state, except Prolines (Pro126, Pro136, Pro170, Pro205, Pro216, Pro264). Dashed line corresponds to significance threshold of one S.D. greater than the mean.

Titration data was further used to determine a quantitative dissociation constant for each CUE domain in the tandem module by independently fitting to 1:1 binding isotherms the chemical shifts from residues with significant perturbations based on the perturbation map, which were unambiguously assigned throughout the experiment. Figure 4.20 presents the binding isotherms for seven CUE1 resonances successfully followed throughout the titration series, resulting in an average dissociation constant (Kd) of (1.2±0.4)mM, a particularly low binding affinity relative to most reported CUE binding domains between 1-500µM [93]. From the curve and perturbation map it may be inferred that all residue are describing a single, simultaneous binding event or surface, with the core MetPhePro tripeptide hydrophobic recognition site expected to form the closest interaction with the complementary hydrophobic surface in Ubiquitin, especially Leu44 [92, 93].

Furthermore, Figure 4.21 presents the binding isotherms for the two CUE2 resonances which showed significant perturbation, Leu289 and Phe292, which nonetheless result in a binding constant indicative of an affinity beyond the ligand concentration range in this experiment, and thus may not be confidently deduced though it is certainly larger than 5mM and thus not of biological significance.

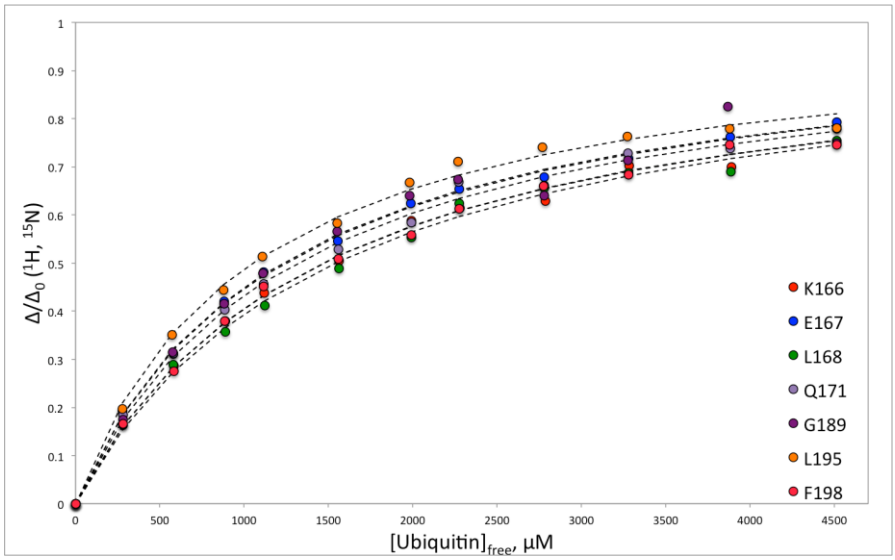



**Figure 4.20 Normalised chemical shift changes from Ubiquitin and SMARCAD1 CUE1 interaction.** Relative chemical shift changes for seven edCUE domain residues upon titration with Ubiquitin, normalised against estimated maximum shift. individual dissociation constant (Kd) was determined using an iterative process of non-linear curve fitting (SigmaPlot from Systat Software Inc.).

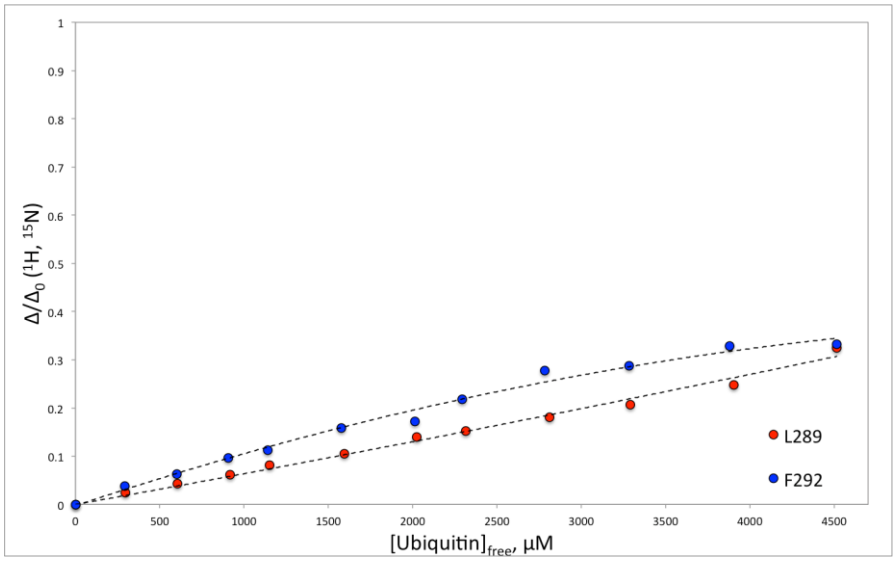



**Figure 4.21 Normalised chemical shift changes from Ubiquitin and SMARCAD1 CUE2 interaction.** Relative chemical shifts changes for two edCUE domain residues upon titration with Ubiquitin, normalised against estimated maximum shift. individual dissociation constant (Kd) was determined using an iterative process of non-linear curve fitting (SigmaPlot from Systat Software Inc.).

These results are consistent with previous studies [135] and may be understood relative to known interactions between alternative CUE domains and ubiquitin based on the extent of divergence from the consensus sequence. Multiple structures and complementary biochemical characterisation have shown that it is the MFP motif at the end of helix 1 in the CUE domain bundle, along with tandem Leucine residues at the end of helix 3, which are required for tight and stable interactions with a Ubiquitin hydrophobic axis along the β-sheet plane, usually with or around Leu8, Ile44 and Val70 [92]. Thus, it is not surprising to find a significantly reduced affinity in CUE1 where Leu168 replaces a strongly conserved Methionine residue in the core motif and an additional residue in helix 3 introduces a one-third helix turn from in the relative position of the homologous residue of the LeuLeu motif, with the actual helix surface being Ala194Leu195. On the other hand, CUE2 shows an even lower degree of sequence conservation, with Glu262Phe263Pro264 in place of the canonical helix 1 MetPhePro motif and L289K290 replacing the LeuLeu helix 3 motif. Furthermore, additional residues known to directly contact Ubiquitin [95], though present or conservatively mutated in CUE1, are for the most part lost in CUE2 (in Figure 3.2, *Sc* CUE2 Asp18 to *Hs* SMARCAD1 CUE1 Glu127 or CUE2 Lys261, and *Sc* CUE2 Leu39Asp40 to *Hs* SMARCAD1 CUE1 Met187Asp188 or CUE2 Tyr282Thr283), suggesting that despite its structural homology, CUE2 might not act as a Ubiquitin binding domain.

## 4.6 Mass spectrometry profiling of CUE1 and SIM mediated interactions.

Having analysed the binary interactions between SMARCAD1 ancillary domains and their canonical ligands (SUMO and Ubiquitin for SIM and CUE1, respectively), it was imperative that a broader search be conducted aimed at identifying specific binding partners, which were likely to be SUMOylated or Ubiquitylated proteins. For this, an affinity purification strategy was implemented, which would be coupled to mass spectrometry bottom-up analysis of co-immunoprecipitated proteins. Such approaches have been frequently and successfully employed, with the main advantage of providing unbiased detection of binding partners under physiological conditions. These hits may be easily quantified through isotopic labelling or

computational (label-free) strategies, increasing the confidence on the veracity of the identified interactor. Furthermore, this method allows cellular stimuli impacting or altering protein-protein interactions to be studied in depth.

The experimental strategy consisted in the transient transfection of GFP-fusion proteins to be used as baits in a high-affinity pull-down from a whole cell lysate, prepared under mild detergent and salt conditions so as to reduce non-specific interactions without jeopardising multi-subunit complexes. As described in detail in Chapter 2, the co-immunoprecipitated proteins were subjected to a label-free quantitative shotgun proteomics analysis and binding partners were identified based on relative abundance changes and statistical significance from replicate measurements.

Central to the proposed design were the simultaneous use of functional abrogating mutants for both the SIM and CUE1 domains, which along with a wild-type construct, would allow individually mediated interactions to be identified. Figure 4.22 presents a schematic diagram of the mutated residues in each binding domain. These constructs, cloned into the pHAGE CMV/eGFP vector produced N-terminal GFP-fusion proteins under the control of the strong CMV promotor, which were transiently transfected into HEK 293T cultured human cells alternatively exposed to neocarzinostatin (NCS) as a DNA damage inducing chemical reagent, as described in Chapter 2. Samples were harvested 48h after transfection and whole cell lysates were prepared under low detergent (0.5% NP40/Igepal) and high salt (450mM NaCl) concentrations to reduce non-specific contaminants and extract chromatin-binding protein from the nuclease-degraded genome.

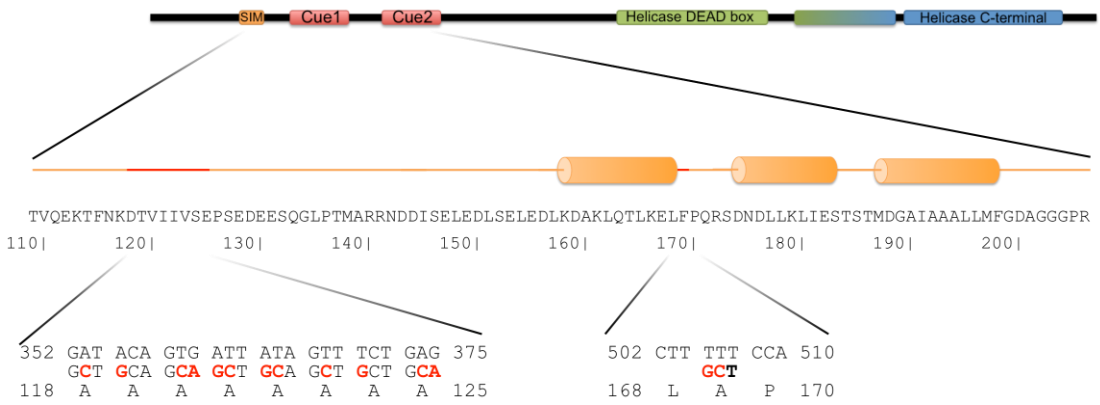



**Figure 4.22 SMARCAD1 SIM and CUE1 binding domain mutants** Schematic representation of functionally abrogating mutant constructs for the SIM (SIM8A) and CUE1 (F169A) from SMARCAD1, with nucleotide and amino acid sequence specified for each. The averaged secondary structure predicted in Chapter 3 is included, along with the primary sequence, as reference for the relative location of each mutation.

Pull-downs were executed with GFP-Trap® magnetic agarose bead (Chromotek), which has been lauded as a very high affinity and thoroughly efficient alternative to traditional GFP immunoprecipitation. Beads were boiled in SDS denaturing buffer and the eluate was processed in the Mass Spectrometry Laboratory at the University of Oxford Target Discovery Institute. Samples were reduced and alkylated, SDS removed by precipitation and the protein pellet resuspended in 6M Urea, followed by a six-fold dilution before Trypsin digestion. Peptides analysed were then separated by high performance liquid chromatography (HPLC) and directly transferred into a Thermo Q-Orbitrap tandem mass spectrometer. This instrument performs two key tasks: measuring mass to charge ratios (m/z) and intensity of the peptides eluting from the HPLC column and recording individual peptide fragmentation spectra (MS/MS). Data thus generated may then compared to protein databases for peptide and protein identification.

Protein profiles were further analysed with Progenesis LC-MS software (NonLinear) for label-free quantification based on the intensity of high-abundance precursor peptides only and normalised against the bait on each sample (GFP-SMARCAD1, either wild-type or mutant constructs) to account for expression yield variation and

pull-down efficiency. Though label-free quantitation does provide certain drawbacks compared to stable isotope-based methods such as SILAC, such as a relative loss in precision and sensitivity requiring at least a two-fold change in protein abundance, label-free methods do represent a less expensive alternative which also allows the simultaneous analysis of multiple variables, such as those evaluated in this work: SMARCAD1 protein sequence mutations and DNA damage response inducing cell treatment, leading to protein-protein interaction perturbations.

Relative abundance profiles between mutant and wild-type pull-downs, grouped based on neocarzinostatin treatment, were analysed as volcano plots. These are specialised scatter plots that relate the size of the fold change in protein abundance to its statistical significance level. In the plots, the abscissa axis indicates biological impact of a perturbation in terms of the relative change in abundance, and corresponds to $\log_2$ transformation of the abundance ratio [Treated-Mutant]/[Untreated-Wild Type]. On the other hand, the ordinate axis relates to the statistical reliability of the change, and corresponds to the $\log_{10}$ transformation of the *p*-value from a *t*-test between the two sets (on a negative log scale, so smaller *p*-values appear further up). This "double criterion" filtering removes potentially substandard hits, which would otherwise be considered relevant. These could be proteins with large scale changes caused by a few outliers with very large values in one group, or low abundance change proteins with false significance test result due to low variance, for instance. Volcano plots allow prompt identification of the most promising candidates from a very large data set.

Figure 4.23 shows the volcano plot analysis of the wild-type SMARCAD1 co-immunoprecipitation data set, analysed based on the alternative DNA damage inducing treatment with NCS. From it, it becomes immediately apparent that very few proteins seem to be significantly modified upon NCS treatment, with the greater number among those that are depleted after damage, perhaps as a consequence of signalling events leading to changes in protein localisation and transient complex composition. As noted in Table 4.2, which enumerates the identified protein

partners whose relative abundance has significantly changed by at least two-fold, only XPC and p66$_\beta$ are of evident functional relevance to SMARCAD1:

- XPC is a SUMO2 modified protein involved in the global genome nucleotide excision repair pathway as a damage sensing and DNA-binding component of the XPC complex.
- p66$_\beta$ (also known as GATAD2B) is a repressive transcription factor of the MeCP1 histone deacetylase complex known to engage in transcriptional silencing of methylated CpG islands.

These are two diverging examples of well-established roles for SMARCAD1 (DNA repair and transcriptional repression) in the context of novel potential multi-subunit complexes. It might be argued that by inducing NCS-dependent ROS damage and double-strand breaks, these protein complexes with SMARCAD1 are displaced, and the active nuclear population of this remodeler is targeted towards DNA double-strand breaks processing.

**Table 4.2 Differentially enriched SMARCAD1 binding partners, identified through label free quantitation mass spectrometry with *p*≥0.05.** Each subset of protein partners has been identified through volcano plot analysis of a pair of differentially co-immunoprecipitated protein profiles differing on a single experimental variable, either NCS treatment or bait protein primary sequence. Lists include names for each potentially interacting protein and Gene Name unique identifier.

| **≥2x fold increase, (+NCS/-NCS) WT GFP-SMARCAD1** | |
|---|---|
| Gene Name | Description |
| SMARCA2 | Isoform Short of Probable global transcription activator SNF2L2 |
| NAP1L4 | Nucleosome assembly protein 1-like 4 |
| SNRPD1 | Small nuclear ribonucleoprotein Sm D1 |
| TUBB3 | Tubulin beta-3 chain |
| **≥2x fold decrease, (+NCS/-NCS) WT GFP-SMARCAD1** | |
| Gene Name | Description |
| GATAD2B | Transcriptional repressor p66-beta |
| IFRD2 | Interferon-related developmental regulator 2 |
| EEF1B2 | Elongation factor 1-beta |
| SKP2 | S-phase kinase-associated protein 2 |
| XPC | Isoform 2 of DNA repair protein complementing XP-C cells |
| EIF3F | Eukaryotic translation initiation factor 3 subunit F |
| TBCD | Isoform 5 of Tubulin-specific chaperone D |
| TEX10 | Isoform 2 of Testis-expressed sequence 10 protein |
| NUP160 | Nuclear pore complex protein Nup160 |

**Table 4.2 (Continued) Differentially enriched SMARCAD1 binding partners, identified through label free quantitation mass spectrometry with *p*≥0.05.**

| **≥2x fold increase, (SIM8A/WT) GFP-SMARCAD1, -NCS** | |
|---|---|
| Gene Name | Description |
| SSBP4 | Single-stranded DNA-binding protein 4 |
| SLC25A4 | ADP/ATP translocase 1 |
| MMS19 | MMS19 nucleotide excision repair protein homolog |
| RBBP4 | Histone-binding protein RBBP4 |
| VPS33B | Vacuolar protein sorting-associated protein 33B |
| AIMP2 | Aminoacyl tRNA synthase complex-interacting multifunctional protein 2 |
| RPL32 | 60S ribosomal protein L32 |
| FIGNL1 | Fidgetin-like protein 1 |
| H2AFZ | Histone H2A.Z |
| QARS | Glutamine tRNA ligase |
| GPATCH4 | G patch domain-containing protein 4 |
| NUP160 | Nuclear pore complex protein Nup160 |
| MEPCE | 7SK snRNA methylphosphate capping enzyme |
| RBMX2 | RNA-binding motif protein, X-linked 2 |
| TMED10 | Transmembrane emp24 domain-containing protein 10 |
| PARP1 | Poly [ADP-ribose] polymerase 1 |
| TUBG1 | Tubulin gamma-1 chain |
| RPL29 | 60S ribosomal protein L29 |
| TTN | Titin |
| FAU | 40S ribosomal protein S30 |
| PRMT1 | Protein arginine N-methyltransferase 1 |
| SCYL2 | SCY1-like protein 2 |
| SWAP70 | Switch-associated protein 70 |
| **≥2x fold decrease, (SIM8A/WT) GFP-SMARCAD1, -NCS** | |
| Gene Name | Description |
| ACTG1 | Actin |
| LYZ | Lysozyme C |
| KRT16 | Keratin, type I cytoskeletal 16 |
| FLG | Filaggrin |
| SUMO2 | Small ubiquitin-related modifier 2 |
| TBCE | Isoform 2 of Tubulin-specific chaperone E |
| MAGOH | Protein mago nashi homolog |
| KRT9 | Keratin, type I cytoskeletal 9 |
| SUCLA | Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial |
| SMARCA5 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 |
| SUMO3 | Small ubiquitin-related modifier 3 |

**Table 4.2 (Continued) Differentially enriched SMARCAD1 binding partners, identified through label free quantitation mass spectrometry with *p*≥0.05.**

| ≥2x fold increase, (SIM8A/WT) GFP-SMARCAD1, +NCS | |
|---|---|
| Gene Name | Description |
| ACLY | ATP-citrate synthase |
| UBAC2 | Ubiquitin-associated domain-containing protein 2 |
| NAP1LA | Nucleosome assembly protein 1-like 4 |
| KRT75 | Keratin, type II cytoskeletal 75 |
| FAU | 40S ribosomal protein S30 |
| STAT1 | Signal transducer and activator of transcription 1-alpha/beta |
| IGF2BP3 | Insulin-like growth factor 2 mRNA-binding protein 3 |
| RANGAP1 | Ran GTPase-activating protein 1 |
| SMARCA5 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 |
| PRPSAP1 | Isoform 2 of Phosphoribosyl pyrophosphate synthase-associated protein 1 |
| KRTBB | Keratin, type II cytoskeletal 6B |
| DDX20 | Probable ATP-dependent RNA helicase DDX20 |
| KRT16 | Keratin, type I cytoskeletal 16 |
| PPP4C | Serine/threonine-protein phosphatase |
| RPL29 | 60S ribosomal protein L29 |
| IFRD2 | Interferon-related developmental regulator 2 |
| RPL29 | 60S ribosomal protein L36a-like |
| RBMX | RNA-binding motif protein, X chromosome |
| CDC5L | Cell division cycle 5-like protein |
| RPS25 | 40S ribosomal protein S25 |
| KRT10 | Keratin, type I cytoskeletal 10 |
| PFN1 | Profilin-1 |
| NUP153 | Nuclear pore complex protein Nup153 |
| HNRNPA2B1 | Heterogeneous nuclear ribonucleoproteins A2/B1 |
| PRRC2B | Protein PRRC2B |
| PSMC3 | 26S protease regulatory subunit 6A |
| DNAJA1 | DnaJ homolog subfamily A member 1 |
| DSC1 | Desmocollin-1 |
| C1QBP | Complement component 1 Q subcomponent-binding protein, mitochondrial |
| GAR1 | H/ACA ribonucleoprotein complex subunit 1 |
| HNRNPH3 | Heterogeneous nuclear ribonucleoprotein H3 |

**Table 4.2 (Continued) Differentially enriched SMARCAD1 binding partners, identified through label free quantitation mass spectrometry with *p*≥0.05.**

| ≥2x fold decrease, (SIM8A/WT) GFP-SMARCAD1, +NCS | |
|---|---|
| Gene Name | Description |
| PMRT3 | Protein arginine N-methyltransferase 3 |
| APEX1 | DNA-(apurinic or apyrimidinic site) lyase |
| ORC5 | Origin recognition complex subunit 5 |
| DPM1 | Dolichol-phosphate mannosyltransferase subunit 1 |
| ARMC6 | Armadillo repeat-containing protein 6 |
| USP7 | Ubiquitin carboxyl-terminal hydrolase 7 |
| RPL10A | 60S ribosomal protein L10a |
| IDH3G | Isocitrate dehydrogenase [NAD] subunit gamma |
| APEH | Acylamino-acid-releasing enzyme |
| AS3MT | Arsenite methyltransferase |
| NOC3L | Nucleolar complex protein 3 homolog |
| HSD17B4 | Peroxisomal multifunctional enzyme type 2 |
| SGPL1 | Sphingosine-1-phosphate lyase 1 |
| GSTM3 | Glutathione S-transferase Mu 3 |
| HIST2H3PS2 | Histone H3 |
| CCNT1 | Cyclin-T1 |
| MARS | Methionine-tRNA ligase, cytoplasmic |
| FLII | Protein flightless-1 homolog |
| RPN2 | Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 2 |
| WDR4 | tRNA (guanine-N(7)-)-methyltransferase non-catalytic subunit WDR4 |
| LDHA | L-lactate dehydrogenase A |
| AARS2 | Alanine-tRNA ligase, mitochondrial |
| C9orf64 | UPF0553 protein C9orf64 |
| IDE | Insulin-degrading enzyme |
| SNRPE | Small nuclear ribonucleoprotein E |
| RFC3 | Replication factor C subunit 3 |
| ATG4B | Cysteine protease ATG4B |
| NUP85 | Nuclear pore complex protein Nup85 |
| SMC2 | Structural maintenance of chromosomes protein 2 |
| KIF1B | Kinesin-like protein |
| NOL6 | Nucleolar protein 6 |
| SLC25A11 | Mitochondrial 2-oxoglutarate/malate carrier protein |
| SLC25A1 | Tricarboxylate transport protein, mitochondrial |
| SCYL2 | SCY1-like protein 2 |
| GRHPR | Glyoxylate reductase/hydroxypyruvate reductase |
| POLR3A | DNA-directed RNA polymerase III subunit RPC1 |
| TRIM25 | E3 ubiquitin/ISG15 ligase TRIM25 |
| HSD17B7 | 3-keto-steroid reductase |
| DERA | 2-deoxyribose-5-phosphate aldolase homolog (C. elegans) |
| METTL13 | Isoform 4 of Methyltransferase-like protein 13 |
| DCTN4 | Dynactin subunit 4 |
| PRKD1 | Serine/threonine-protein kinase D |
| TCEB3 | Transcription elongation factor B polypeptide 3 |

**Table 4.2 (Continued) Differentially enriched SMARCAD1 binding partners, identified through label free quantitation mass spectrometry with *p*≥0.05.**

| **≥2x fold decrease, (SIM8A/WT) GFP-SMARCAD1, +NCS** | |
|---|---|
| Gene Name | Description |
| AP3M1 | AP-3 complex subunit mu-1 |
| RPS15A | 40S ribosomal protein S15a |
| NDUFS1 | NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial |
| ABCD3 | ATP-binding cassette sub-family D member 3 |
| SEPT9 | Septin-9 |
| SRP72 | Signal recognition particle subunit |
| NOLC1 | Nucleolar and coiled-body phosphoprotein 1 |
| SRP54 | Signal recognition particle 54 kDa protein |
| **≥2x fold increase, (F169A/WT) GFP-SMARCAD1, -NCS** | |
| Gene Name | Description |
| MAD2L1BP | MAD2L1-binding protein |
| GAR1 | H/ACA ribonucleoprotein complex subunit 1 |
| MMS19 | MMS19 nucleotide excision repair protein |
| PDCD2 | Programmed cell death protein 2 |
| MOGS | Mannosyl-oligosaccharide glucosidase |
| AIMP2 | Aminoacyl tRNA synthase complex-interacting multifunctional protein 2 |
| GNE | Bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase |
| KRT36 | Keratin, type I cuticular Ha6 |
| TMED10 | Transmembrane emp24 domain-containing protein 10 |
| CELF1 | CUGBP Elav-like family member 1 |
| JUP | Junction plakoglobin |
| TTN | Titin |
| KRT35 | Keratin, type I cuticular Ha5 |
| TXNRD1 | Thioredoxin reductase 1 |
| PRPSAP2 | Phosphoribosyl pyrophosphate synthase-associated protein 2 |
| **≥2x fold decrease, (F169A/WT) GFP-SMARCAD1, -NCS** | |
| Gene Name | Description |
| YARS2 | Tyrosine--tRNA ligase, mitochondrial |
| KRT18 | Keratin, type I cytoskeletal 18 |
| HNRNPH1 | Heterogeneous nuclear ribonucleoprotein H |
| AHCY | Adenosylhomocysteinase |
| GTF3C1 | General transcription factor 3C polypeptide 1 |
| ACAT2 | Acetyl-CoA acetyltransferase, cytosolic |
| OXCT1 | Succinyl-CoA:3-ketoacid coenzyme A transferase 1, mitochondrial |
| GRSF1 | G-rich sequence factor 1 |
| ACTG1 | Actin, cytoplasmic 2 |
| HK1 | Hexokinase-1 |
| HSD17B12 | Estradiol 17-beta-dehydrogenase 12 |
| MGEA5 | Bifunctional protein NCOAT |
| SUMO2 | Small ubiquitin-related modifier 2 |
| ESD | S-formylglutathione hydrolase |
| UBAC2 | Ubiquitin-associated domain-containing protein 2 |
| PWP2 | Periodic tryptophan protein 2 |

**Table 4.2 (Continued) Differentially enriched SMARCAD1 binding partners, identified through label free quantitation mass spectrometry with *p*≥0.05.**

| **≥2x fold decrease, (F169A/WT) GFP-SMARCAD1, -NCS** | |
|---|---|
| Gene Name | Description |
| HIST1H2AJ | Histone H2A |
| RNC2 | Reticulocalbin-2 |
| AMOT | Angiomotin |
| GNL3 | Guanine nucleotide-binding protein-like 3 |
| CCT5 | T-complex protein 1 subunit epsilon |
| EPRS | Bifunctional glutamate/proline-tRNA ligase |
| RBM34 | RNA-binding protein 34 |
| ZNF638 | Zinc finger protein 638 |
| EEF2 | Elongation factor 2 |
| SUCLA2 | Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial |
| SMARCA5 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 |
| NAP1L4 | Nucleosome assembly protein 1-like 4 |
| ZFP1 | Zinc finger protein 1 |
| NPM1 | Nucleophosmin |
| KRT85 | Keratin, type II cuticular Hb5 |
| VIM | Vimentin |
| TRIM28 | Transcription intermediary factor 1-beta |
| SUMO3 | Small ubiquitin-related modifier 3 |
| BAZ1B | Tyrosine-protein kinase BAZ1B |
| **≥2x fold increase, (F169A/WT) GFP-SMARCAD1, +NCS** | |
| Gene Name | Description |
| KRT82 | Keratin, type II cuticular Hb2 |
| KRT86 | Keratin, type II cuticular Hb6 |
| KRT85 | Keratin, type II cuticular Hb5 |
| LRRC15 | Leucine-rich repeat-containing protein 15 |
| KRT36 | Keratin, type I cuticular Ha6 |
| KRT35 | Keratin, type I cuticular Ha5 |
| KRT31 | Keratin, type I cuticular Ha1 |
| TXNRD1 | Thioredoxin reductase 1, cytoplasmic |
| RPL32 | 60S ribosomal protein L32 |
| EXOSC1 | Exosome complex component CSL4 |
| **≥2x fold decrease, (F169A/WT) GFP-SMARCAD1, +NCS** | |
| Gene Name | Description |
| TRIM28 | Transcription intermediary factor 1-beta |
| GATAD2B | Transcriptional repressor p66-beta |
| ZNF638 | Zinc finger protein 638 |
| ZFP1 | Zinc finger protein 1 |

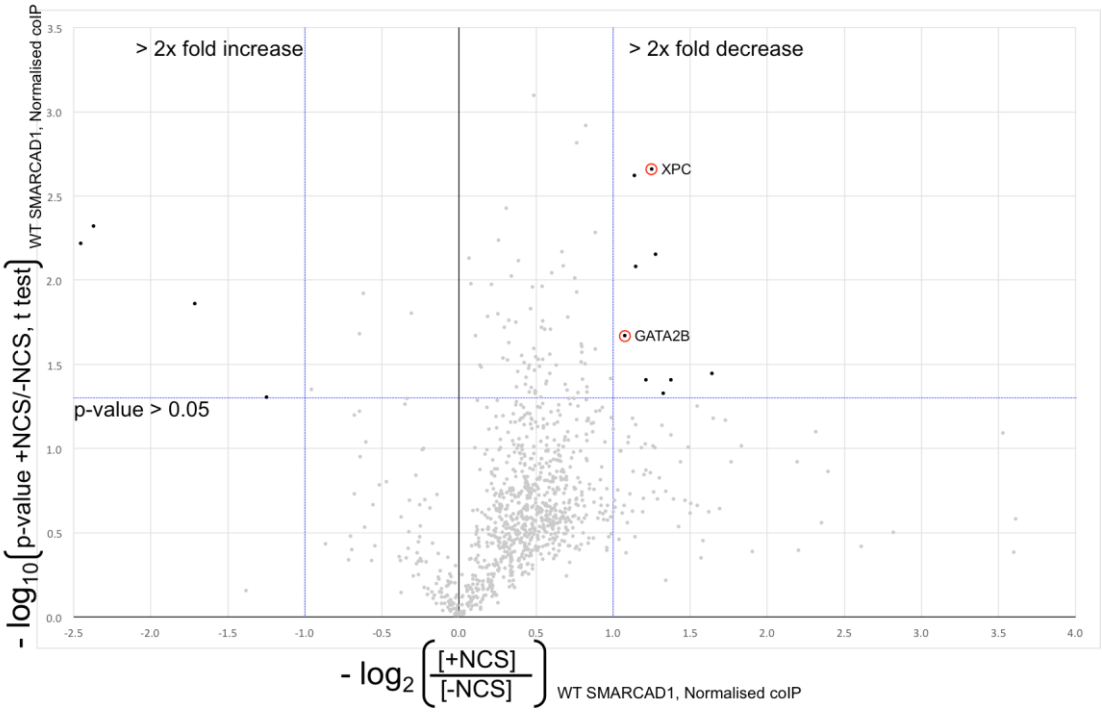



**Figure 4.23 Volcano plot of differential abundance proteins from a co-immunoprecipitated pair of wild-type GFP-SMARCAD1 samples alternatively treated with NCS.** Volcano plot illustrates significantly differentially abundant proteins, corresponding to proteins effectively recruited upon DNA damage. The -log10 [*t*-test *p*-value] is plotted against the -log2 [fold change: WT SMARCAD1,+NCS/WT SMARCAD1,-NCS]. The blue vertical lines denote ±2x-fold change while the blue horizontal line denotes *p*= 0.05. Proteins significantly more abundant after NCS treatment are to the left of the ordinate axis, and those significantly depleted are to right.

The use of functional nullifying mutant sequences for each of the SMARCAD1 binding domains permitted the examination of individually mediated interactions, as mentioned above. The analysis of differential protein profiles was implemented through volcano plots between samples equivalently treated with NCS, thus samples had been either submitted or not to chemically induced DNA damage were compared relative to the differential ability of the wild-type protein to recruit a given protein partner which failed to happen in the mutant. Figure 4.24 presents the volcano plots of the SIM dependent protein profiles of binding partners that failed to be recruited, with and without NCS treatment.

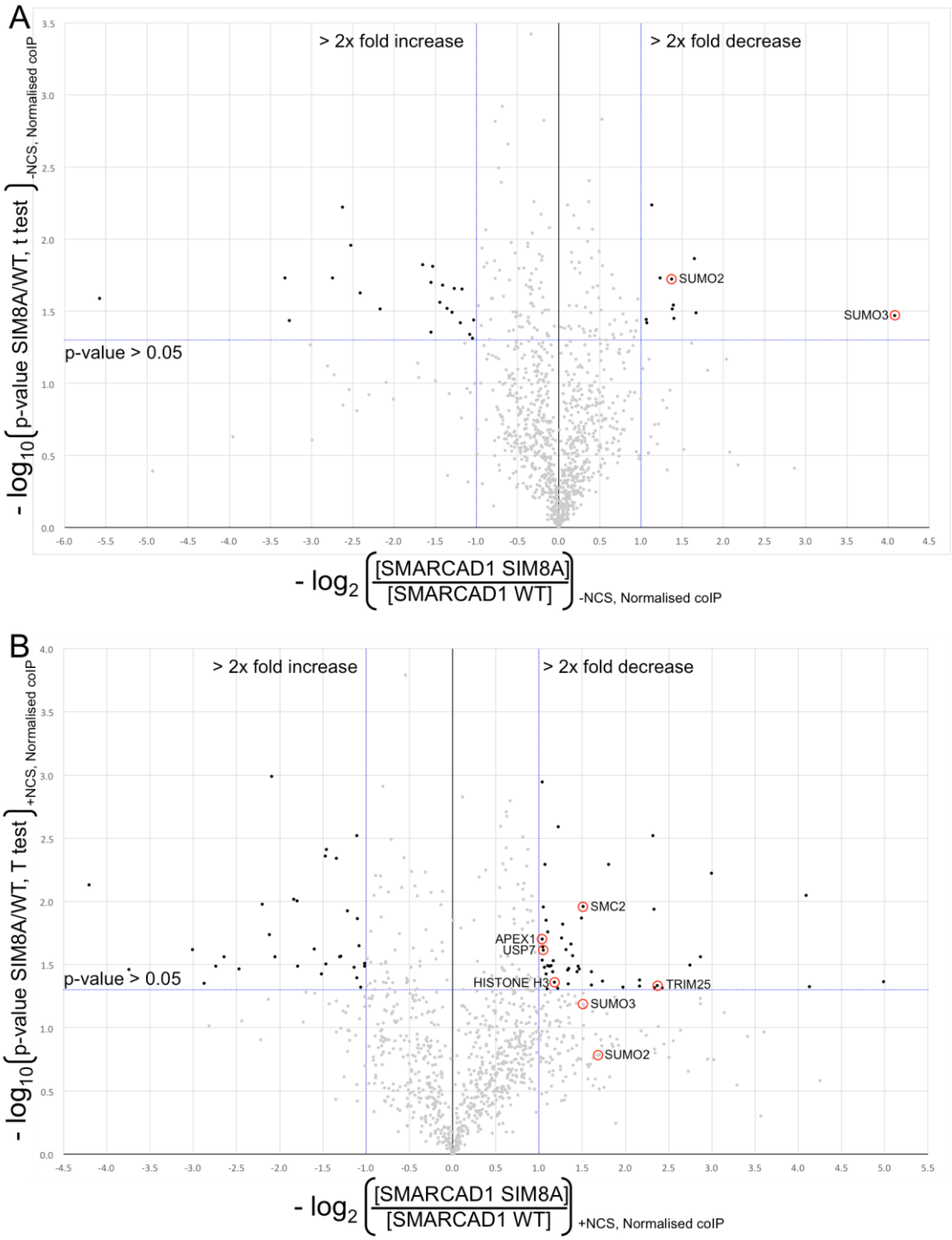



**Figure 4.24 Volcano plots of differentially abundant proteins from a co-immunoprecipitated set of wild type and SIM mutant GFP-SMARCAD1 samples alternatively treated with NCS (Legend on next page).**

**Figure 4.24 (Previous page) Volcano plots of differentially abundant proteins from a co-immunoprecipitated set of wild type and SIM mutant GFP-SMARCAD1 samples alternatively treated with NCS.** A) Relative abundance ratio of significantly perturbed binding proteins which failed to interact with a SMARCAD1 SIM mutant (SIM8A), without NCS treatment, B) Significant abundance change in proteins which failed to interact with a SMARCAD1 SIM mutant (SIM8A), having been subjected to NCS treatment. The -log10 [*t*-test *p*-value] is plotted against the -log2 [fold change: SIM8A SMARCAD1/WT SMARCAD1]. The blue vertical lines denote ±2x-fold change while the blue horizontal line denotes *p*= 0.05. SIM8A binding proteins significantly less abundant relative to wild-type co-immunorpecipitation samples are to the right of the ordinate axis, and those significantly enriched are to right.

From these two plots it was determined that, without NCS treatment only SUMO2/3 is significantly reduced in the pull-down of SIM mutant bait relative to wild type, whereas after NCS treatment a considerable number of protein targets were significantly lost, that is their interaction is both DNA damage dependent and SIM mediated. A few of these proteins, listed in Table 4.2 stand out, including:

- SMC2, a central component of the Condensin complex, required for the condensation of interphase chromatin prior to mitosis.
- APEX1, a central player in the cellular response to oxidative stress as part of the base excision repair pathway.
- USP7, a SUMO2-modified ubiquitin hydrolase that regulates the MDM2-p53/53BP1 axis though stabilisation of de-ubiquitylated forms of its target protein, favouring cell-growth arrest and apoptosis.

Overall, it would appear that interaction through SIM might target SMARCAD1 for involvement in the DNA damage response.

Finally, Figure 4.25 shows the volcano plots of the CUE1 dependent protein profiles of binding partners that failed to be recruited, with and without NCS treatment. In contrast to the differential profile of SIM mediated interactions, the number of

binding partners significantly perturbed and lost when abrogating the CUE1 domain is larger when the samples was not treated with NCS. When comparing the list of identified hits in Table 4.2, it is easy to see that, with the exception of p66$_\beta$ the three remaining proteins significantly reduced are also present in the subset of binding partners without NCS treatment. This suggests that these interactions are not dependent or altered by chemically induced DNA damage:

- KAP1, the transcriptionally repressive, heterochromatin-bound master regulator.
- ZNF638, a repressive transcription cofactor, involved in the regulation of proadipogenic genes.
- ZFP1, a SUMO2-modified KRAB transcription regulator, known to interact with KAP1.

These proteins represent the canonical cellular role of SMARCAD1 as a member of transcriptionally repressive complex with KAP1.

**Figure 4.25 (Next page) Volcano plots of differentially abundant proteins from a co-immunoprecipitated set of wild type and CUE1 mutant GFP-SMARCAD1 samples alternatively treated with NCS.** A) Relative abundance ratio of significantly perturbed binding proteins which failed to interact with a SMARCAD1 CUE1 mutant (F169A), without NCS treatment, B) Significant abundance change in proteins which failed to interact with a SMARCAD1 CUE1 mutant (F169A), having been subjected to NCS treatment. The -log10 [*t*-test *p*-value] is plotted against the -log2 [fold change: F169A SMARCAD1/WT SMARCAD1]. The blue vertical lines denote ±2x-fold change while the blue horizontal line denotes *p*= 0.05. F169A binding proteins significantly less abundant relative to wild-type co-immunorpecipitation samples are to the right of the ordinate axis, and those significantly enriched are to right.

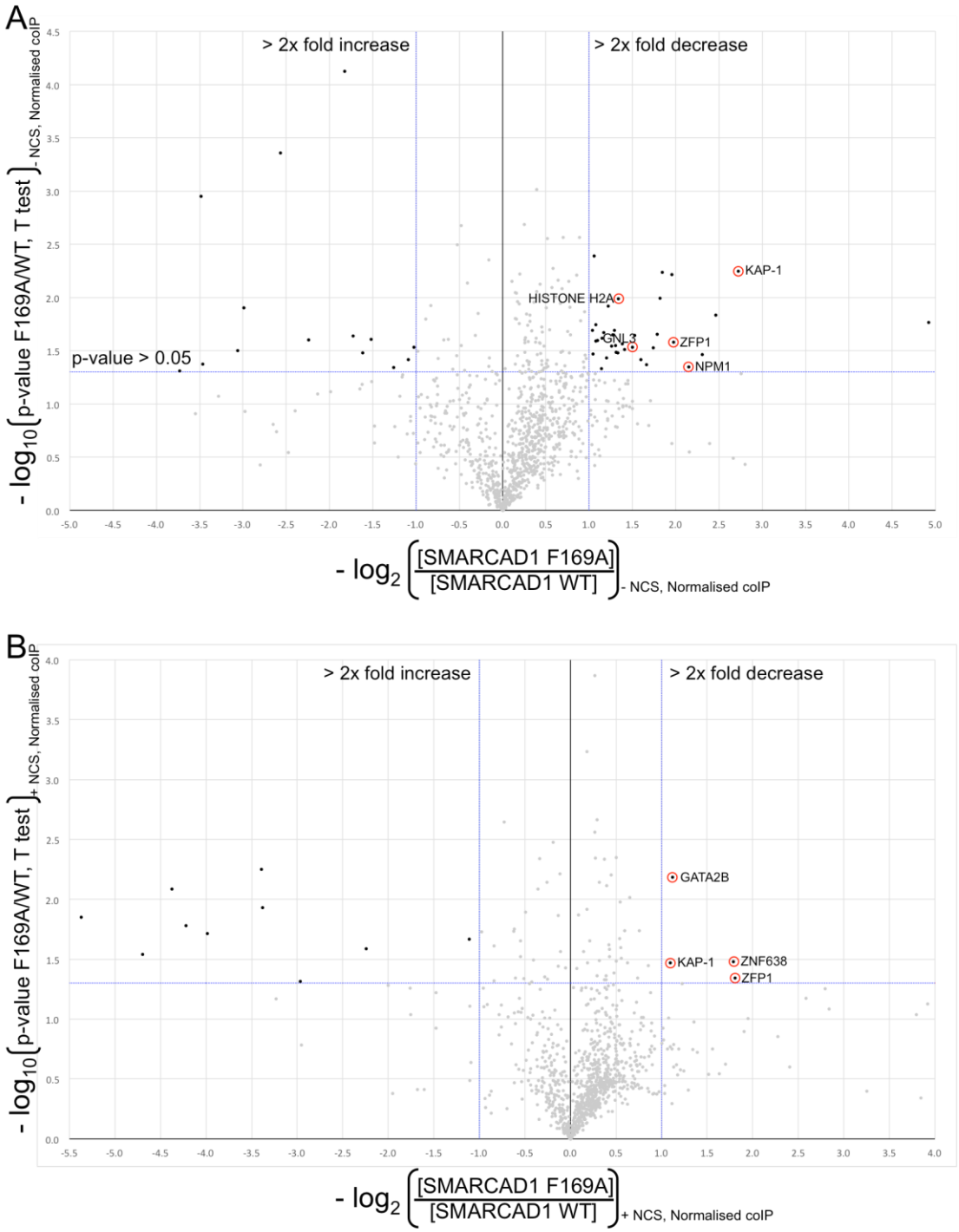



**Figure 4.25 Volcano plots of differentially abundant proteins from a co-immunoprecipitated set of wild type and CUE1 mutant GFP-SMARCAD1 samples alternatively treated with NCS (Legend on previous page).**

Three additional hits are noteworthy from among the CUE1 mediated binding partners. These are all significantly altered without NCS treatment, but not so when DNA damage is induced, suggesting that these interactions at least might be subject to negative regulation as part of the DNA damage response:

- Histone H2A, a core nucleosome component, which has recently been proposed as a target for SMARCAD1 CUE1 when ubiquitylated downstream of BRAC1-BARD1 E3-ligase activity in DNA double-strand breaks as part of the homologous recombination repair pathway.
- NPM1, a binding partner of APEX1 (shown above to be a target for SIM specific binding), regulating its endonuclease activity while favouring apurinic repairs in dsDNA.
- GNL3, a SUMO2-modified KAP1-MDM2-p53/53BP1 binding partner, which regulates the MDM2-p53/53BP1 axis though stabilisation of its ubiquitylated state and favours p53/53BP1 ubiquitylation and degradation.

The overarching pattern observed from CUE1 mediated interactions is that it appears to recruit SMARCAD1 to chromatin, through a myriad of binding partners that fine-tune the specific context for it remodelling activity.

## 4.7 Conclusions and future work

The work described in this chapter integrates *in vitro* analysis of binding events mediated by the SIM and CUE1 ancillary domains in SMARCAD1, creating a robust framework to interpret and further hypothesize about the functional role in which this non-canonical remodeler might engage.

Thorough characterisation of the SMARCAD1 SUMO interacting motif is a prime example of the natural evolution a structural premise undergoes when challenged by a cumulative body of evidence. The working hypothesis started with a recombinant unmodified construct as a model for the SUMO binding system. It followed the use of phosphorylated synthetic peptides to assess individual contributions from each

modified residue, through the identification of its constitutive poly-phosphorylated state *in vivo* by mass spectrometry, finally evolving into a phosphomimetic model with appropriate charge distributions and a biologically relevant affinity constant. This system exemplifies the importance of considering the role that post-translational modifications likely adopt in establishing and stabilising transient protein interactions in cells.

Similarly, detailed analysis of ubiquitin binding domain CUE1 (and to a lesser extent, CUE2) shows that functional assumptions based on sequence and structural homology may be misleading. These domains do not match the behaviour of other mono-ubiquitin binding CUE domains such as *Sc* CUE2 or *Hs* TOLLIP, which bind to its ligand with affinities in the low micromolar range, contrasting the low millimolar range association found for either CUE domain in SMARCAD1. Moreover, CUE2 appears to interact with Ubiquitin at considerably weaker affinities, likely due to key primary sequence changes, which alter the predicted binding surface so that residue complementarity is lost. On the other hand CUE1 residue perturbation upon Ubiquitin interaction is consistent with the established mode of molecular recognition for other CUE domains, if only with a weaker affinity. It is important to note the crucial benefit that NMR offers in analysing these otherwise undetectable interactions, which might nonetheless be relevant as transient contacts under specific cellular contexts.

Though complete, these *in vitro* analyses would be enhanced by complementary titrations with additional ligands that may represent structural and functional variants, namely SUMO2/3 for SMARCAD1 SIM, and poly-ubiquitin for the tandem CUE domains. The argument for each is sound, and could result in significant improvement of our understanding of this remodeler, driving the interpretation of functional analyses such as cellular localisation and pathway-specific recruitment:

- SUMO proteins have been shown to play an important role in the DNA damage response signalling pathways, with SUMO1 stimulating all forms of DNA double-strand break repair and SUMO2 specifically affecting the

conservative NHEJ mechanism, central to the repair response against double-strand breaks throughout the cell cycle. Though only sharing approximately 45% sequence homology with SUMO1, SUMO2/3 display an equivalent tertiary fold and have been shown to interact with SIMs in much the same way as SUMO1. In fact, selective cases have substantially demonstrated that certain SIM proteins specifically favour SUMO2/3 binding. A notable example is the BRCA1-A complex subunit RAP80, central to Lys63-linked poly-ubiquitin chains at double-strand break sites leading to HR processing. RAP80 presents domain architecture equivalent to SMARCAD1, though in place of tandem CUE domains it possess a double Ubiquitin Interacting Motif (UIM) creating a single elongated α-helix capable of spanning poly-ubiquitin chains. Furthermore, RAP80 is known to be SUMO2- modified and possesses a phosphorylated SIM, which was recently demonstrated to preferentially bind to SUMO2 in poly-Ubiquitin/SUMO hybrid chains. Given the structural parallels with SMARCAD1, and the recent report where this remodeler has been proposed to enhance double-strand breaks resection in HR downstream of BRAC1-A complex, it would be a simple, sensible and potentially enlightening effort to assess the potential binding affinity between SMARCAD1 SIM and SUMO2.

- As mentioned above, poly-ubiquitylation is a structural staple of DNA damage response signalling cascades, with many proteins presenting tandem Ubiquitin or poly-ubiquitin chain modifications. Considering SMARCAD1 has a tandem CUE domain arrangement, with a flexible linker potentially capable of spanning linked Ubiquitin chains, or even distinguishing them as has been shown before for *Sc* Cue1, makes this mode of modified protein recognition plausible. Moreover, it might be the case that despite not being able to bind mono-ubiquitin with biologically relevant affinity, SMARCAD1 CUE2 might still bind to poly-ubiquitin chains once local effective concentration increase due to avidity is accounted for. Though mere speculation, the possibility of poly-ubiquitin chain binding remains feasible, and a potential strategy to address it could be the recombinant expression of end-to-end ubiquitin

repeats as models for oligomeric forms of varying composition. Once verified whether concerted binding of poly-ubiquitin chains do take place, and depending on potential oligomeric size preferences, it would be possible (if arduous) to synthesize *in vitro* poly-ubiquitin chains with single Lysine mutants to guide the composition and linkage of the connection between individual repeats.

Finally, mass spectrometry results decisively point towards specific protein interactions mediated through either of these binding domains. Of particular interest is the trend observed for differential pull-down profiles of mutant versions of SMARCAD1 when DNA damage was induced. It would appear that the SIM, but not CUE1 domain, is engaging in protein interactions that could place it within signalling pathways of various mechanisms for repair in the DNA damage response. On the other hand, CUE1 seems to mediate multiple interactions, notably KAP1, which presumably might be responsible for chromatin recruitment within the context of SMARCAD1 remodeling roles in replication and transcription. Interestingly, some of the CUE1 mediated interactions, including Kap1, were independent of DNA damage treatment, suggesting an enduring, constitutive interaction that is lost when the binding domain is abrogated. The next logical step is to verify these hits through complementary techniques.

Despite their great promise and robustness, there are important ways in which these results could be strengthened overall and might lead to further work, such as synchronizing the cell population so as to simplify the DNA damage repair pathway response to double-strand breaks and improve the SNR by reducing the heterogeneity of cell states prior to (and during) treatment. Another relatively simple way in which this assay could have been improved is to switch from a whole cell lysate to a nuclear extract preparation as input for co-immunoprecipitation, which would likely reduce the number of biologically irrelevant non-specific interactions, screening noise out and simplifying data processing. Both of these factors likely would not have drastically changed the outcome of the experiment, but they might have improved SNR and variance enough to strengthen potential

interacting hits which are currently borderline on the statistical significance threshold despite having greater than two-fold protein abundance changes.

As mentioned above, the main outlook of these results is the validation of individual hits as directly binding partners through *in vitro* and *in vivo* biophysical characterisation with assays such as two-hybrid screening or reciprocal co-immunoprecipitation and, though, a generalised or high-throughput analysis is beyond the scope of this work, one candidate does stand out: the interaction with *bona fide* binding partner KAP1 has now been shown to be mediated by the ubiquitin binding CUE1 domain in SMARCAD1, independent of DNA damage treatment. If a direct interaction, these results suggest that KAP1 might be ubiquitylated, a novel modification on an extensively characterised protein, known to be primarily SUMOylated and only modified through RNF4 mediated ubiquitylation as part of its cell cycle dependent degradation. Alternatively, it is possible that an "ubiquitin-like" fold is featured in KAP1 (despite the absence of the short sequence homology to any of the known ubiquitin-like domains in proteins such as rad23A/B, Bag1, UBQLN1-3 among others [99], nor to ubiquitin itself) or this interaction might signify a diverging binding function for this loosely canonical CUE domain. These possibilities are explored in Chapter 5.

Chapter 5

# Biochemical and biophysical analysis of the SMARCAD1-KAP1 protein complex

The identification of KRAB-associated binding protein 1 (KAP1, also known as TIF1-β or TRIM28) as key binding partner of SMARCAD1 mediated specifically through the CUE1 domain, evidenced by its statistically significant abundance decrease following CUE1 abrogation mutagenesis in Chapter 4, creates an exciting opportunity to further elucidate the molecular mechanism establishing this interaction. KAP1 has been extensively studied [43, 44, 185-188], is known to be a master transcription regulator [189, 190] and, along with SMARCAD1, is part of a protein complex re-establishing silent heterochromatin epigenetic marks on newly replicated sections of the genome [51]. Furthermore, KAP1 has recently been linked to DNA damage repair, alternatively as an active (though indirect) participant in early stages of the signalling cascade by affecting the choice of double strand breaks (DSB) repair pathway [42, 191, 192] or as a target for ATM phosphorylation leading to relaxation of heterochromatin regions neighbouring DSB [43, 44]. As such, it could represent the linking factor placing SMARCAD1 within the DNA repair pathway in a chromatin context and this interaction was characterised through a combination of mass spectrometry, pull-down assays and biophysical methods.

The binary complex is shown to be direct and independent of a ubiquitination modification in KAP1 (Section 5.1), though the SMARCAD1 CUE1 mutation (F169A) leads to partial unfolding of the helical bundle (Section 5.2), so both binding events might not be mutually exclusive. Following the identification of the KAP1 domain responsible for SMARCAD1 binding (Section 5.3), a recombinant hexahistidine

construct was expressed and purified (Section 5.4) after which the complex was biochemically characterised using ITC (Section 5.4) and NMR (Section 5.5). Of particular relevance are the preliminary structural studies involving crystallisation of the protein complex (Section 5.6) and selective immunofluorescence co-localisation of SMARCAD1-Kap1 active complexes in cultured cells with a Proximity Ligation Assay (Section 5.7).

## 5.1 Ubiquitylation state of KAP1

Considering that the ubiquitin-binding CUE1 domain in SMARCAD1 directly mediates KAP1 interaction, the first assumption about the molecular mechanism sustaining the complex is that KAP1 might be constitutively ubiquitylated. This is directly deduced from mass spectrometry protein profiling results in Chapter 4, which indicate that decreased abundance of SMARCAD1 CUE1 mutant F169A relative to that of wild type is independent of chemically induced DNA damage and would constitute a novel report of a potentially crucial modification required for KAP1 function. To investigate this, post-translational modification analysis was performed on SMARCAD1 co-immunoprecipitated KAP1 samples from cultured human cells alternatively exposed to neocarzinostatin (NCS) as a DNA damage inducing chemical reagent, as described in Chapter 2 and presented in Chapter 4. Post-translational modifications (PTMs) were identified using the database and *de novo* search program PEAKS version 7 (Bioinformatics Solutions). Figure 5.1 shows sequence coverage and high redundancy of detected tryptic peptides for KAP1, both with and without NCS induced damage.

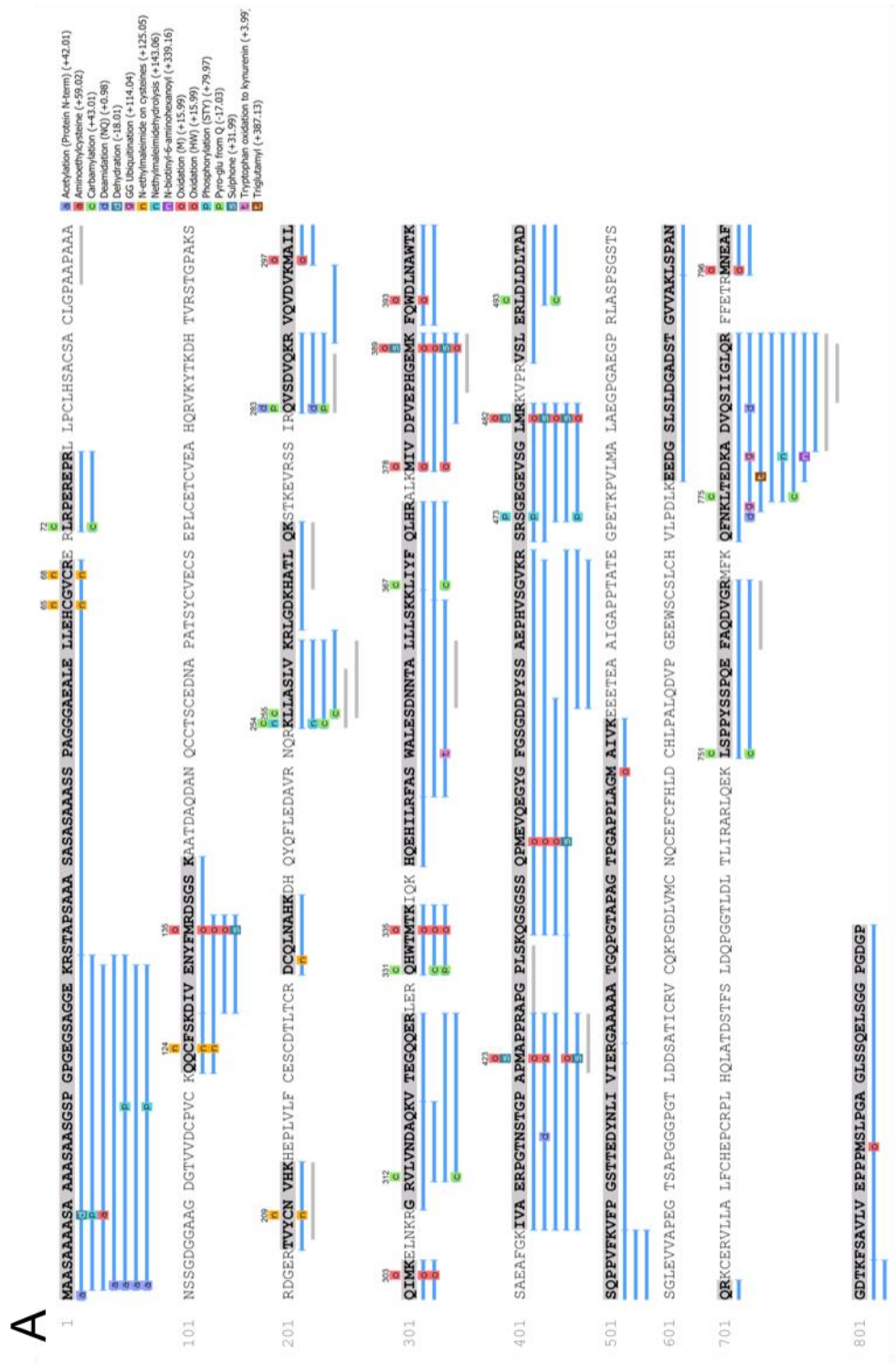



**Figure 5.1 Post-translational modifications profiling of KAP1 differentially exposed to chemically induced DNA damage and co-immunoprecipitated with GFP-SMARCAD1.** Tandem mass spectrometry analysis of tryptic digests and peptide detection was performed on samples without neocarzinostatin induced DNA damage (A). Multiple peptides bearing modifications are identified including potentially ubiquitylated Lys774, and Lys779.

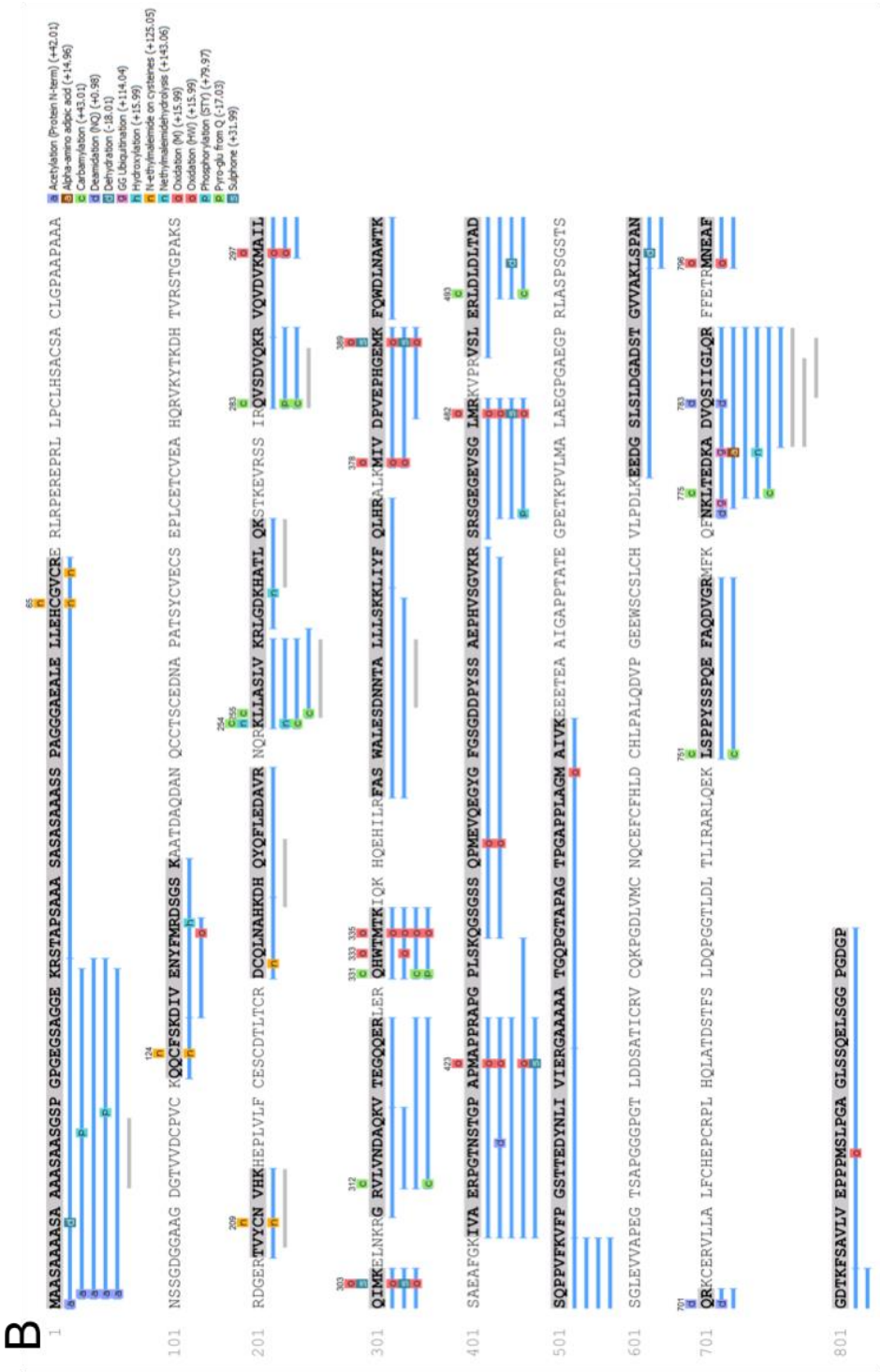



**Figure 5.1 (Continued) Post-translational modifications profiling of KAP1 differentially exposed to chemically induced DNA damage and co-immunoprecipitated with GFP-SMARCAD1.** Tandem mass spectrometry analysis of tryptic digests and peptide detection was performed on samples after 30-60min with 100µg/mL on NCS at 37°C (B). Multiple peptides bearing modifications are identified including potentially ubiquitylated Lys774, and Lys779.

These results appear to suggest that KAP1 is ubiquitylated along surface exposed residues in the Bromo domain [193], alternatively modified before and after chemically induced DNA damage. However, all three Lysine residues have been previously reported to be SUMOylated, with extensive studies, including Lysine to Arginine point mutations, to elucidate their impact on KAP1 mediated transcription repression [194], which show that Lys779 and to a lesser extent Lys750, are essential in sustaining the transcriptional repressive role through their interaction with histone methyl transferase SETDB1. Furthermore, with the exception of SUMO-dependent RNF4 poly-ubiquitylation and subsequent proteasome-mediated degradation [192, 195], such modification has not been reported for KAP1 thus far.

In order to confirm these finding, a pull-down of ubiquitylated proteins was performed under highly denaturing conditions to reduce the likelihood of post-lysis degradation. As described in Chapter 2, cultured human HEK 293T cells were transfected with pOPINF/Ubiquitin Nt MAHx6G, which would transiently overexpress a recombinant ubiquitin fusion protein with an uncleavable hexahistide N-terminal tag. Following 48h after transfection, and assuming successful incorporation of this alternative ubiquitin form in target proteins without chemically induced DNA damage, a whole cell lysate was prepared under 8M Urea, 300mM NaCl and pH8.0, with the supernatant (after sonication and centrifugation to clear the insoluble debris) being incubated with TALON® magnetic beads. The input lysate and Cobalt affinity pull-down was analysed by immunoblot against KAP1, Penta-His and Ubiquitin, as shown in Figure 5.2

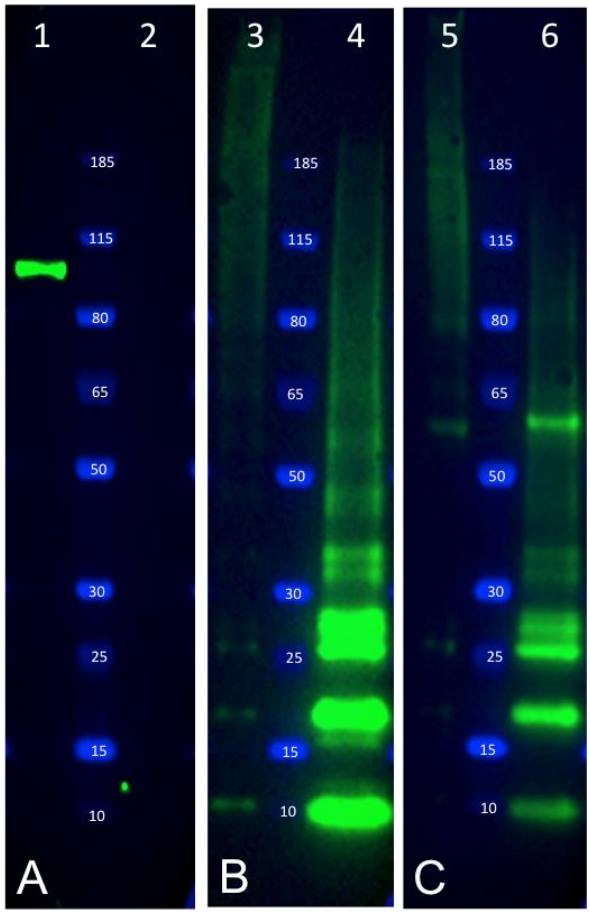



**Figure 5.2 Cobalt TALON® affinity purification of Nt MAH$_6$G/Ubiquitin modified proteins from HEK 293T cells under denaturing conditions.** A) Rabbit anti-KAP1/Goat anti-Rabbit, B) Mouse anti-Ubiquitin/Goat anti-Mouse, C) Mouse anti-PentaHis/Goat anti-Mouse. Lanes 1, 3 and 5: Whole cell lysate input, 50ng total protein, Lanes 2, 4 and 6: TALON® magnetic beads pull-down. Page Ruler molecular weight ladder (Thermo) and HRP enhanced chemiluminescence signals in blue and green, respectively.

As is evident in from Figure 5.2, KAP1 does not appear to be ubiquitylated under the experimental conditions employed throughout this work, at least not in the absence of chemically induced DNA damage. In the blots there is evidence for the presence of high molecular weight ubiquitylated proteins along with poly-ubiquitylated species from the smeared signals in both the anti-Ubiquitin and anti-PentaHis blots, which are to a large extent equivalent.

## 5.2. Structural consequences of SMARCAD1 CUE1 F169A mutation

Given that it does not seem likely that a ubiquitin modification in KAP1 is responsible for the interaction with SMARCAD1, it is possible that the CUE1 domain possesses a secondary binding surface specific to KAP1, independent of the canonical ubiquitin recognition module. Alternatively, KAP1 may exhibit a ubiquitin-like fold within its structure, appropriating the SMARCAD1 CUE1 domain. Either possibility could justify the subtle though important divergence from the consensus sequence for ubiquitin-binding CUE domains, with CUE1 having potentially evolved as a KAP1 specific binding domain.

In order to assess whether the CUE1 inactivation mutant F169A leads to a structurally homologous fold to the wild type, which would allow additional interpretation of the functional lost of KAP1 binding, a SMARCAD1 SIM/CUE1 F169A construct equivalent to that extensively used in NMR titrations to characterise SUMO1 and Ubiquitin binding, was designed and expressed in *E.coli* as a uniformly $^{15}N$- labelled recombinant protein. A 2D heteronuclear single quantum correlation (HSQC) spectrum was collected and compared to that of an unmodified SIM/CUE1 construct. Figure 6.3 shows the overlaid spectra, collected at 600MHz.

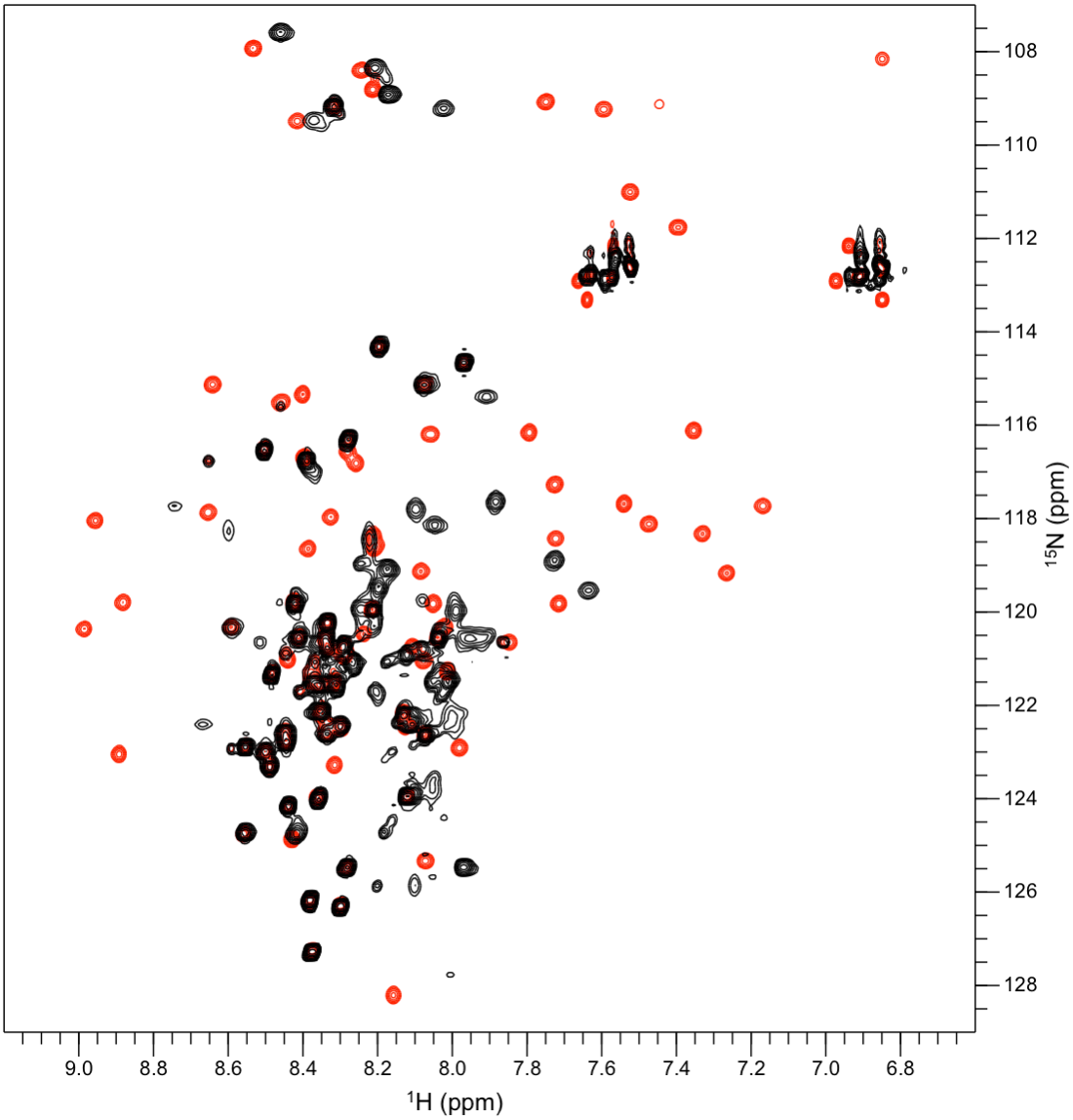



**Figure 5.3 $^{1}H,^{15}N$ HSQC analysis of SMARCAD1 SIM/CUE1 F169A.** Overlay of $^{15}N$ eCUE1 F169A (black) and $^{15}N$ eCUE1 wild type (red) $^{1}H,^{15}N$ HSQC spectra at approximately 1mM each, collected at 600MHz.

It becomes immediately evident from Figure 5.3 that the overall fold in the F169A construct is lost, though it does not seem likely that individual helices are unfolded since the point mutation is at a linker between helices 1 and 2. Indeed, there is smaller though still considerable distribution of amide resonances, consistent with folded secondary structures. It is not possible to unequivocally identify most peaks solely from the HSQC spectra, but based on the overall distribution there does not seem to be many additional unstructured peaks, with equivalent distribution of peaks upfield (e.g. Lys166, Ser173, and Leu178) or downfield (e.g. Glu167, Leu168,

Asp176, Leu177, Leu180, and Met197) relative to unstructured resonances (e.g. Val120, Ile122, Thr137, Gly204) over a smaller and closer range of 1H chemical shifts. The broad signals and reduced chemical shift dispersion might be an indication that the F169A mutation leads to a molten-globule state of the otherwise globular CUE1 domain, but further characterisation would be required through additional NMR [196], circular dichroism [197] or small angle X-ray scattering [198] analysis.

Whether a molten globule or simply a partially unfolded protein, it is nonetheless certain that no sequence specific correlation may be drawn between F169A mutagenesis of CUE1 and loss of KAP1 binding, though this does not detract from the domain specificity of the interaction.

## 5.3 Identification of SMARCAD1 binding domain of KAP1

With no clear indication of a ubiquitylation event, there is little indirect evidence to suggest which subdomain of KAP1 could be involved in SMARCAD1 binding. To identify such interacting protein region, a series of truncation mutants of KAP1 were designed and expressed as C-terminal eGFP/Hexahistidine fusion proteins with an additional N-terminal triple repeat of the SV40 nuclear localisation signal [199]. Overall, constructs were designed based on the general distribution of functional domain clusters in KAP1: the KRAB binding RBCC domain, the HP1 binding mid-section and the PHD/Bromo dual domain involved in histone methylation and deacetylation through SETDB1 and HDAC1/2 association, respectively. Furthermore, based on the mass spectrometry analysis, to account for the possibility of ubiquitylation of the Bromo domain despite no clear evidence to support it, two additional truncation mutants were included, one without the Bromo domain and one without the entire PHD/Bromo region. All constructs were designed based on previously publish data [186, 193, 194, 200], in combination with secondary structure and disorder prediction based on bioinformatics tools [108-110]. Along with a near full length construct, these truncation mutants were transiently expressed in HEK 293T cultured cells and harvested 48h after transfection as described in Chapter 2. The soluble supernatant of a whole cell lysate, cleared

through centrifugation and prepared under mild condition (ionic strength and detergent concentration) to preserve transient protein interactions, were subjected to a GFP pull down using Chromotek GFP-Trap® magnetic agarose beads. Normalised lysate input and pull-downs were analysed with SDS-PAGE and immunoblotting against GFP, SMARCAD1, PCNA and Histone H3, as shown in Figure 5.4.

**Figure 5.4 (Next page) Identification of SMARCAD1 binding domain in KAP1. GFP pull down of transiently transfected KAP1 truncation mutants.** A) Schematic representation of KAP1 truncation mutants as C-terminal GFP/Hexahistidine fusion proteins, B) Western blot analysis of normalised whole cell lysate input (50ng total protein) and GFP-Trap® co-immunoprecipitation. Lane 1/9: KAP1 near full length, construct 1, Lane 2/10: KAP1 RBCC, construct 2, Lane 3/11: KAP1 HP1 binding mid-region, construct 3, Lane 4/12: KAP1 PHD/Bromo, construct 4, Lane 5/13: KAP1 RBCC/HP1/PHD ΔBromo, construct 5, Lane 6/14: KAP1 RBCC/HP1 ΔPHD/Bromo, construct 6, Lane 7/15: NLS-GFP, negative control. Page Ruler molecular weight ladder (Thermo) in blue, and HRP enhanced chemiluminescence signals from Goat anti-GFP/Mouse anti-Goat, Rabbit anti-H3/Goat anti-Rabbit, Rabbit anti-SMARCAD1/Goat anti-Rabbit, and Rabbit anti-PCNA/Goat anti-Rabbit in green.

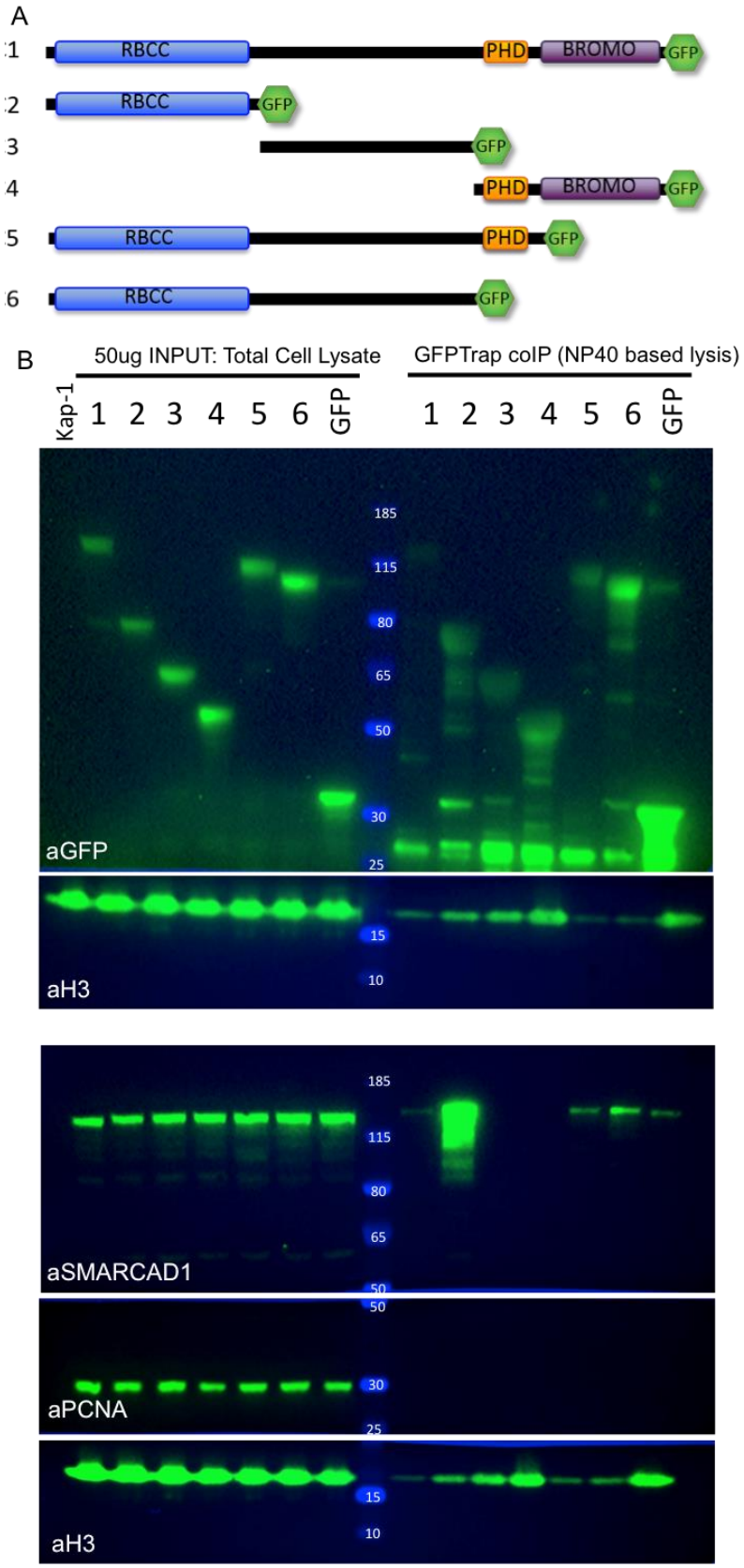



**Figure 5.4 Identification of SMARCAD1 binding domain in KAP1. GFP pull down of transiently transfected KAP1 truncation mutants (Legend on previous page).**

Figure 5.4 makes evidently clear that KAP1 establishes an interaction with full length SMARCAD1 through its RBCC domain, with all constructs that include it capable of immunoprecipitating endogenous SMARCAD1 from HEK 293T lysates. Though the signal derived from chemiluminescence detection is not linearly proportional to the abundance of the epitope, it is clear that the construct formed only by the RBCC domain has recruited a significantly larger amount of SMARCAD1 relative to the other KAP1 constructs. This is perhaps due to fewer competing binding partners vying for space across the bait-coated GFP-Trap® beads, considering that each subdomain is known to have stable binding partners specific to epitopes within its sequence such as HP1 and SETDB1. Additionally, it might be possible that the full length architecture of KAP1 imposes conformational restraints which regulate the protein-protein interaction between KAP1 RBCC and SMARCAD1, be it through reduced affinity or a fixed protein partners ratio which is not observed when limited only to the effective binding domain.

Considering that the RBCC is approximately 400 amino acids long and close to 45kDa in size, it is safe to assume that SMARCAD1 CUE1, at one fifth the size, likely interacts with one of the four RBCC sub-domains: Ring, B-Box 1, B-Box 2 or Coiled-Coil. To determine if any of them (individually or as a subset combination) specifically interact with SMARCAD1, a new set of truncation mutants was designed. Once more, the constructs were C-terminal eGFP/Hexahistidine fusion proteins with an N-terminal double repeat of the SV40 nuclear localisation signal. The boundaries for each subdomain were designed based on previously published data for the KAP1 B-Box2 (PDB 2YVR) [201] and structural homology models based on sequence alignment, including alternative RBCC subdomains of related TRIM proteins whose structures are known, such as Bre1 C3HC4 Ring domain (PDB 4R7E) [202], MID1/TRIM18 tandem B-Box (PDB 2JUN) [203] and TRIM25 Coiled-Coil and L2 segments (PDB 4LTB) [204], while still considering KAP1 specific secondary structure and disorder propensities. Along with the full RBCC domain and NLS-eGFP as positive and negative controls, respectively, a total of nine KAP1 RBCC truncation constructs were transiently expressed in HEK 293T cells, harvested 48h after transfection and whole cell lysates (prepared as described above and in Chapter 2) were subjected to

GFP mediated co-immunoprecipitation using Chromotek GFP-Trap® magnetic agarose beads. Normalised lysate input and pull-downs were analysed with SDS-PAGE and immunoblotting against GFP, SMARCAD1 and Histone H3, as shown in Figure 5.5.

**Figure 5.5 (Next page) Identification of SMARCAD1 binding subdomain in KAP1 RBCC.** GFP pull down of transiently transfected KAP1 RBCC truncation mutants. A) Schematic representation of KAP1 RBCC truncation mutants as C-terminal GFP/Hexahistidine fusion proteins, B) Western blot analysis of normalised whole cell lysate input (50ng total protein) and GFP-Trap® co-immunoprecipitation. Lane 1/8: KAP1 Ring, construct 7, Lane 2/9: KAP1 B-Box1, construct 8, Lane 3/10: KAP1 B-Box2, construct 9, Lane 4/11: KAP1 Coiled-Coil, construct 10, Lane 5/12: KAP1 RBCC domain, construct 2, Lane 6/13: NLS-GFP, negative control. C) Western blot analysis of normalised whole cell lysate input (50ng total protein) and GFP-Trap® co-immunoprecipitation. Lane 1/9: KAP1 Ring/B-Box1, construct 11, Lane 2/10: KAP1 Ring/B-Box1/B-Box2, construct 12, Lane 3/11: KAP1 B-Box1/B-Box2, construct 13, Lane 4/12: KAP1 B-Box1/B-Box2/Coiled-Coil, construct 14, Lane 5/13: KAP1 B-Box2/Coiled-Coil, construct 15, Lane 6/14: KAP1 RBCC domain, construct 2, Lane 7/15: NLS-GFP, negative control. Page Ruler molecular weight ladder (Thermo) in blue, and HRP enhanced chemiluminescence signals from Rabbit anti-SMARCAD1/Goat anti-Rabbit, Goat anti-GFP/Mouse anti-Goat, and Rabbit anti-H3/Goat anti-Rabbit in green.

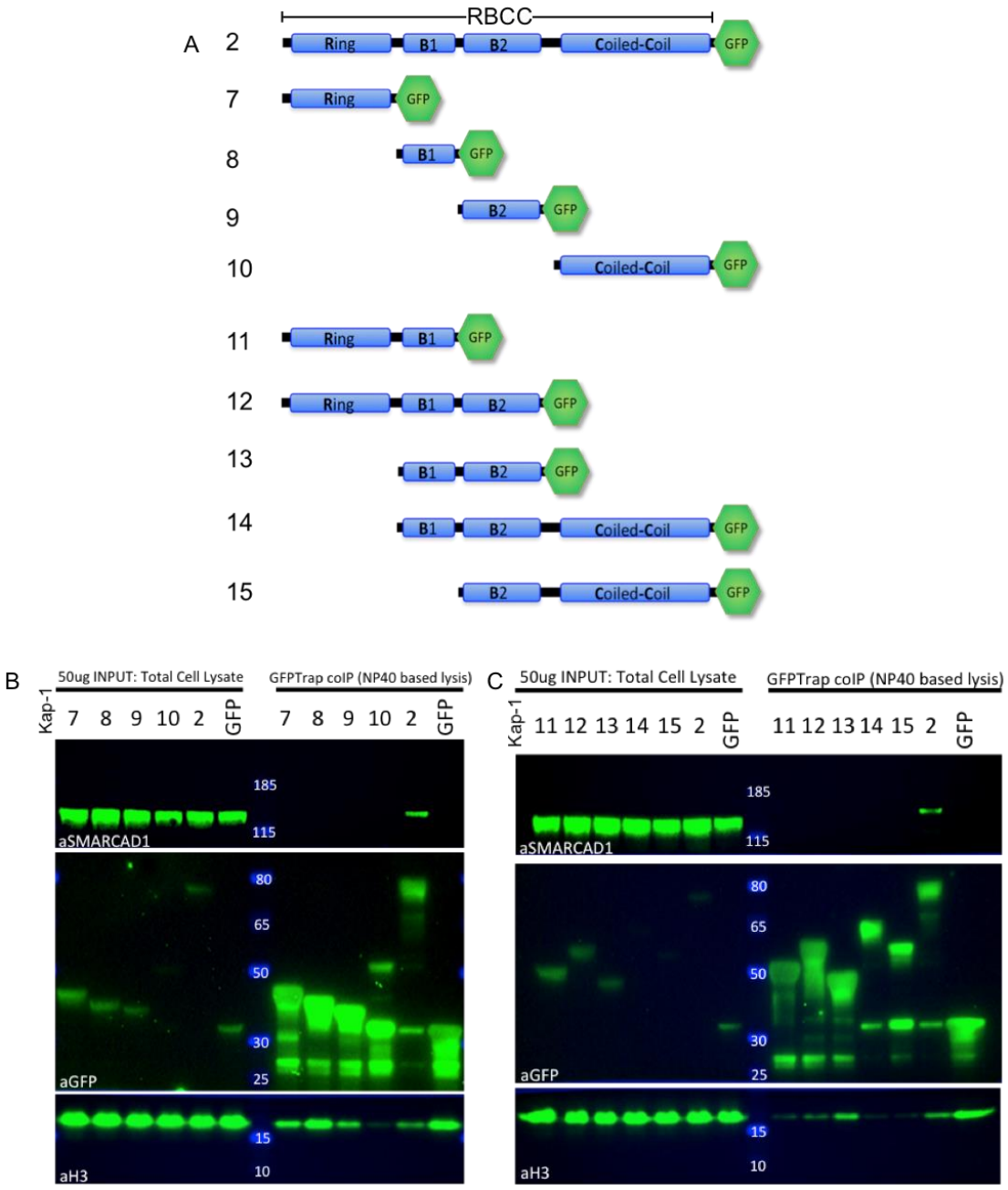



**Figure 5.5 Identification of SMARCAD1 binding subdomain in KAP1 RBCC. GFP pull down of transiently transfected KAP1 RBCC truncation mutants (Legend on previous page).**

Based on Figure 5.5 it may be concluded that only the complete RBCC domain is able to interact with endogenous SMARCAD1 in HEK 293T lysates. This could be interpreted as evidence that the epitope involved in SMARCAD1 recognition is not monomeric KAP1 but rather the oligomeric state characteristic of most other TRIM proteins [205-208], likely a homo-dimer though KAP1 has been proposed to potentially act as a homo-trimer in solution [200]. Whether the interaction involves equimolar protein ratios or not cannot be directly inferred, and would require direct *in vitro* characterisation of the complex. Careful analysis of the immublot also suggests that the Coiled-Coil domain of KAP1 is poorly soluble and may lead to uncontrolled aggregation if the Ring domain is not part of the construct. This implies that the intact RBCC might pack in such a way that the Ring domain inhibits non-specific aggregation by engaging otherwise exposed residues. In contrast, the tandem B-Box domain do not seem to regulate or alleviate the non-specific aggregation propensity of the Coiled-Coil domain, though they are well expressed and soluble whether individually, as a tandem construct or together with the Ring domain. From this it may be deduced that SMARCAD1 CUE1 domain recognition requires an epitope only available in the correctly folded oligomeric state of KAP1 RBCC, either a binding interface formed by the integration of more than one sub-domain or surface contacts only appropriately exposed when correctly folded in a defined quaternary structure.

## 5.4 Recombinant protein expression and oligomeric state determination

In order to further investigate the minimal complex between KAP1 RBCC and SMARCAD1 CUE1 domains, it was necessary to produce and purify recombinant proteins. In an extension of the successful over-expression of SMARCAD1 SIM/CUE1/CUE2 constructs employed in solution NMR studies described in Chapters 3 and 4, a C-terminal hexahistidine fusion protein of SMARCAD1 CUE1 (pOPINE/WT CUE1 WCtHis) was expressed in *E.coli* and purified through Nickel affinity, anion exchange and size exclusion chromatography, as detailed in Chapter 2. On the other hand and despite a previous report to contrary [200], KAP1 RBCC hexahistidine

fusion proteins (either N- or C-terminal) did not express in *E.coli*, which motivate the transition to heterologous expression in baculovirus-infected *Spodoptera frugiperda* (Sf9) insect cells, as described in Chapter 2. As evidenced in Figure 5.6, the use of baculovirus expression system in insect cells resulted in overexpression of KAP1 RBCC constructs, with a clear advantage in yield and stability exhibited by the N-terminal hexahistidine fusion protein, compared to equivalent C-terminal hexahistidine or C-terminal eGFP/Hexahistidine fusion proteins.

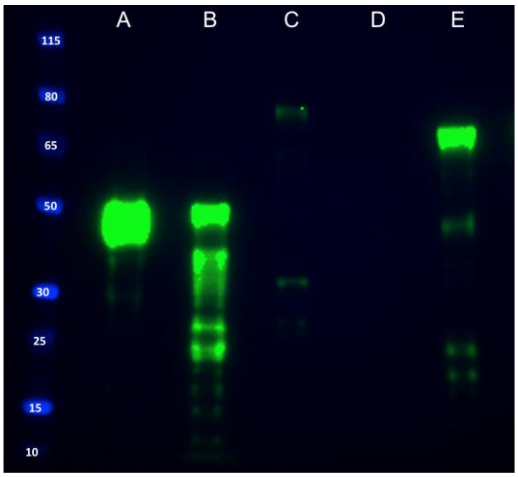



**Figure 5.6 Heterologous expression of KAP1 RBCC fusion proteins in baculovirus infected Sf9 insect cells.** Western Blot analysis of KAP1 RBCC constructs lysed from Sf9 cells after baculovirus amplification to produce P2 virus stock. A) pOPINF/Kap1 RBCC NtHis, B) pOPINE/Kap1 RBCC CtHis, C) pOPINGFP/Kap1 RBCC NLSx3 CtGFP, D) Bacmid DNA only, Negative control, E) Kindlin3 pOPINE/*FERMT3*, positive control. Page Ruler molecular weight ladder (Thermo) in blue, and HRP enhanced chemiluminescence signals from Mouse anti-PentaHis/Goat anti-Mouse in green.

Large-scale expression of pOPINF/Kap1 RBCC NtHis was followed by sequential purification with Nickel affinity and size exclusion chromatography, as described in Chapter 2. Optimisation of purification conditions through repetitive trial-and error made evident the distinct solubility sensitivity of the protein to low ionic strength and high imidazole concentration, which seemed to be (at least partially) concentration-dependent. The final purification strategy avoided long incubations between purification steps and limited imidazole concentrations during Nickel

affinity elution by adjusting a linear gradient to a step-wise transition with long elution steps at lower concentrations. Analysis by SDS-PAGE allowed identification of appropriate fractions, which were immediately diluted in binding buffer to reduce the imidazole concentration of the pooled samples, concentrated using centrifugation filter and further purified using size exclusion chromatography. The SDS-PAGE analysis corresponding to the elution profile from a Superdex S200 16/600 column is shown in Figure 5.7. Given the high level of purity, fractions containing the recombinant KAP1 RBCC construct were combined and concentrated to approximately 80µM (monomer). Yields ranged between 4.7mg-5.4mg per litre of cultured cells harvested after expression.

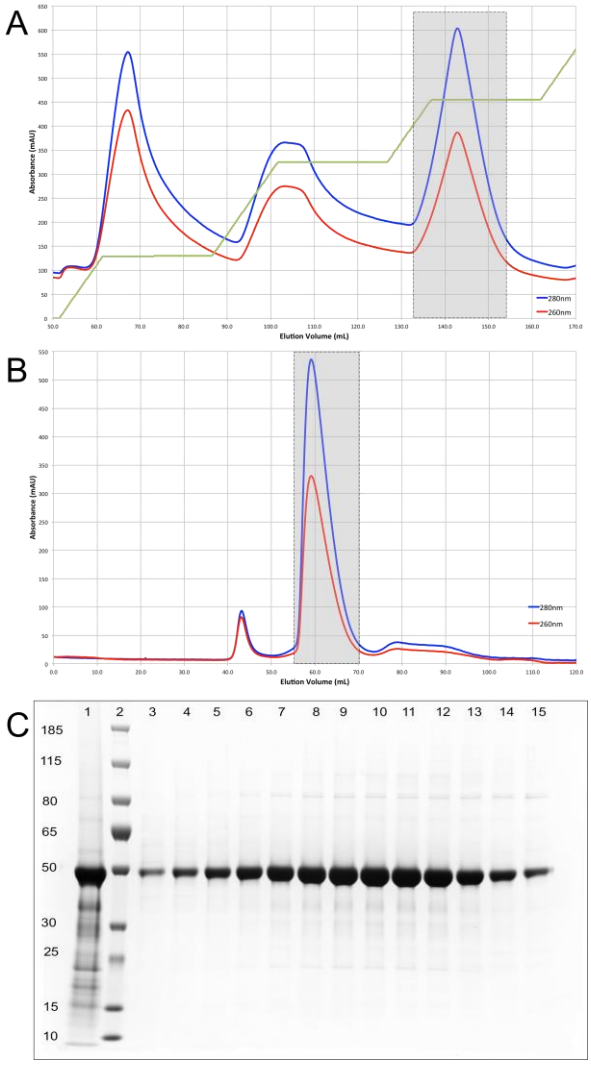



**Figure 5.7 (Next page) Strategies for KAP1 RBCC N-terminal hexahistidine fusion protein overexpressed in Sf9 insect cells.**

**Figure 5.7 Strategies for KAP1 RBCC N-terminal hexahistidine fusion protein overexpressed in Sf9 insect cells.** SDS-PAGE analysis and chromatographic traces of elute absorption at 280nm and 260nm. Samples analysed by PAGE are marked on the chromatogram: A) Nickel affinity purification, B) Size exclusion chromatography, and C) Size exclusion chromatography SDS-PAGE analysis. Lane 1: Nickel affinity eluate, Lane 2: Page Ruler Molecular Weight ladder, Lanes 3-15: SEC elution fractions. SDS-PAGE gels were stained with Coomassie Brilliant Blue; molecular weight ladder is Page Ruler (Thermo Fisher). Chromatogram traces are: absorption at 280nm in blue, absorption at 260nm in red, elution buffer composition in green.

To assess the oligomeric state of recombinant KAP1 RBCC protein size exclusion chromatography coupled to multiple angle laser-light scattering (SEC-MALLS) was carried out by the Biophysics Facility in the Department of Biochemistry, University of Oxford. A dilution series of five samples of KAP1 RBCC NtHis between 3.7mg/mL and 0.7mg/mL was analysed as described in Chapter 2 and processed using ASTRA 6.1 (Wyatt Technology). Figure 5.8 shows individual chromatographic traces from the differential refractive index detector, with slight concentration dependence in elution volume and an asymmetric peak with a late eluting tail suggesting a potential reversible equilibrium between oligomeric species. The average molar mass estimation is (94±6)kDa, indicating a strong tendency towards stable dimeric species, in direct contrast to previous reports suggesting KAP1 RBCC exists as a trimer in solution [200, 205].

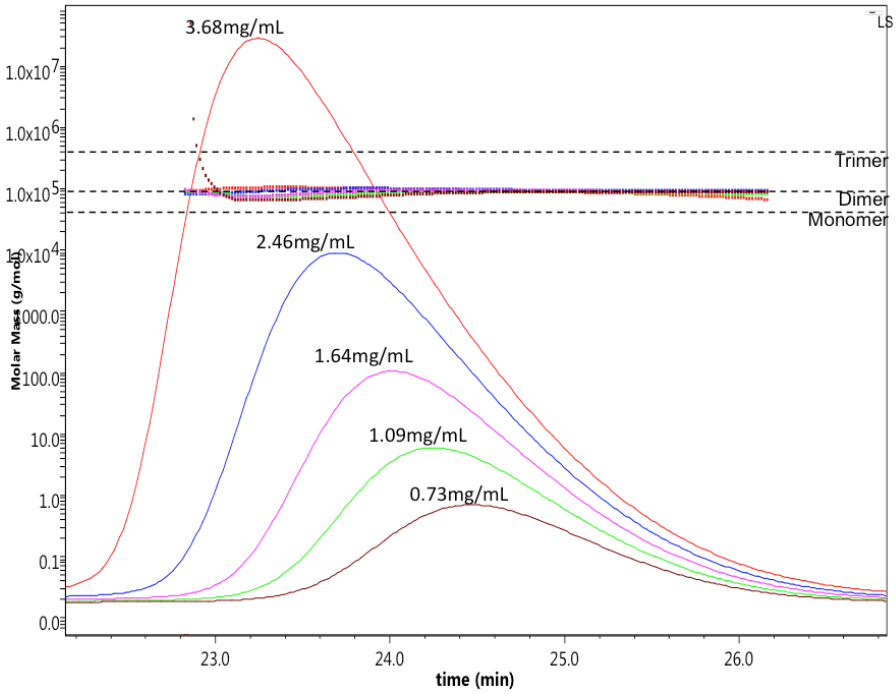



**Figure 5.8 Multiple angle laser light scattering analysis of KAP1 RBCC NtHis.** Chromatograms from the differential refractive index detector are shown at five protein concentrations. Calculated molar masses measured across the peak corresponding to a dimeric species are shown as segmented lines, with theoretical monomer (46.15kDa), dimer (92.30kDa) and trimer (138.45kDa) molecular weights given as reference.

The disparity in oligomeric size likely derives from the non-native purification of recombinant proteins from bacterial expressions systems as reported by Peng *et al* [189, 200, 205]. Furthermore, the reports make note that even under optimized buffer conditions, the protein sample tended to precipitate indicating poor stability in solution. An additional discrepancy, which favours the results presented in this work, arises from the method for molecular mass determination, given that previous reports relied almost exclusively on UV absorbance from size exclusion chromatography elution profiles relative to molecular weight standards. Though useful for everyday analysis, this kind of relative interpolation with calibration curve does not take into account additional factors that may skew the elution profile of proteins diverging from the assumed model. In particular, proteins that do not comply with a globular tertiary structure or may interact non-specifically with the

resin, usually display delayed retention volumes with consequently larger estimated molecular masses [209]. In comparison, the use of static light scattering in connection with refractive index measurements yields an accurate determination of the molecular mass of proteins in solution, which is independent of molecular shape or hydrodynamic parameters [210, 211].

In support of the dimeric nature of KAP1 RBCC domain, the structures of related TRIM protein coiled-coil domains have been exclusively that of antiparallel homodimers, including TRIM5α (PDB: 4TN3 [212]), TRIM25 (PDB: 4LTB, [204]), TRIM69 (PDB: 4NQJ, [213]). It must be considered, however, that higher order oligomers cannot be entirely dismissed in light of the concentration dependent oligomerisation exhibited by TRIM5a (whose RBCC domain has been studied beyond its coiled-coil domain, including the Ring and B-Box domains, in the form of truncated coiled-coil chimeras [214, 215]) when analysed at higher concentrations than those in the present work, and shown to be mediated by weak self-association of the TRIM5α B-Box domain into a trimer [214, 216]. However, it seems unlikely that the trimeric form of TRIM5α, recently determined to be crucial for the establishment of the functional hexameric species required for retrovirus capsid recognition and subsequent ubiquitin-dependent degradation will necessarily be conserved in a functionally distinct and evolutionary distant member of the TRIM family such as KAP1 [217, 218]. Consistent with such considerations, in a very recent study SAXS data from the entire RBCC domain and associated truncation mutants of the closely related TRIM25 ubiquitin ligase, clearly show a stable dimer in solution, with the Ring/B-Box1/B-Box2 module folding back over the antiparallel homodimeric coiled-coil domains and creating additional stabilising contacts by dimerization of the Ring domains, just as our current model proposes [208]. Nevertheless, it would be advisable to further analyse the KAP1 RBCC domain with complementary analytical ultracentrifugation experiments (both sedimentation velocity and equilibrium) to confirm its oligomeric state and better assess any potential dynamic transitions to higher order oligomers [219].

## 5.5 Binary protein-protein interaction between SMARCAD1 CUE1 and KAP1 RBCC domains

Having successfully expressed and purified a stable, soluble construct of KAP1 RBCC domain, it was possible to characterise its interaction with SMARCAD1 ubiquitin binding CUE1 domain *in vitro*. Isothermal titration calorimetry (ITC) analysis was carried out after dialysing both recombinant proteins against the same batch of Phosphate Buffer (20mM $NaH_2PO_4$ pH7.0, 100mM NaCl) supplemented with 1mM β-Mercaptoethanol as reducing agent. KAP1 RBCC NtHis protein concentration was limited due to reduced solubility at low ionic strength, and as such was used as the analyte protein in the ITC experiment at 20.6μM (monomer). SMARCAD1 CUE1 CtHis was consequently used as the titrant protein at 113.1μM and injected over 18 fractions of 2μL each. The heat rate change over each injection was recorded as a thermogram, from which integrated heat data was plotted as an isotherm relative to analyte-titrant ratio. The titration differential enthalpy, affinity constant and interaction stoichiometry were derived from an iterative nonlinear regression analysis of the isotherm [220], using the MicroCal PEAQ-ITC Analysis Software (Malvern) assuming a single binding site per analyte. Figure 6.9 shows the heat rate for power compensation throughout the titration experiment and the binding thermogram. Based on it, a 2:1 stoichiometry is derived from this interaction; that is, one KAP1 RBCC dimer per each SMARCAD1 CUE1 domain, binding with a dissociation constant of (1.4±0.5)μM. The binding model between one RBCC dimer as a functional unit recognising the CUE1 from SMARCAD1 has precedent in additional evidence from other interaction between KAP1 RBCC and alternative binding partners including MDM2 [221], MAGE A/C [222], and notably KRAB transcription factor KOX1 [189, 200]. A systematic analysis of individual binding subdomains showed that swapping with homologous sequences from related TRIM proteins (TRAF proteins 1-6, MID1, TIF1-α and TIF1-γ) did not sustain the oligomeric form of KAP1 nor the interaction with its binding partner, and each subdomain contributed to the specificity and selectivity for its substrate. Together, these data strongly suggest that KAP1 RBCC binds as an integrated unit to its ligand, and that each subdomain contributes to protein recognition and oligomerisation.

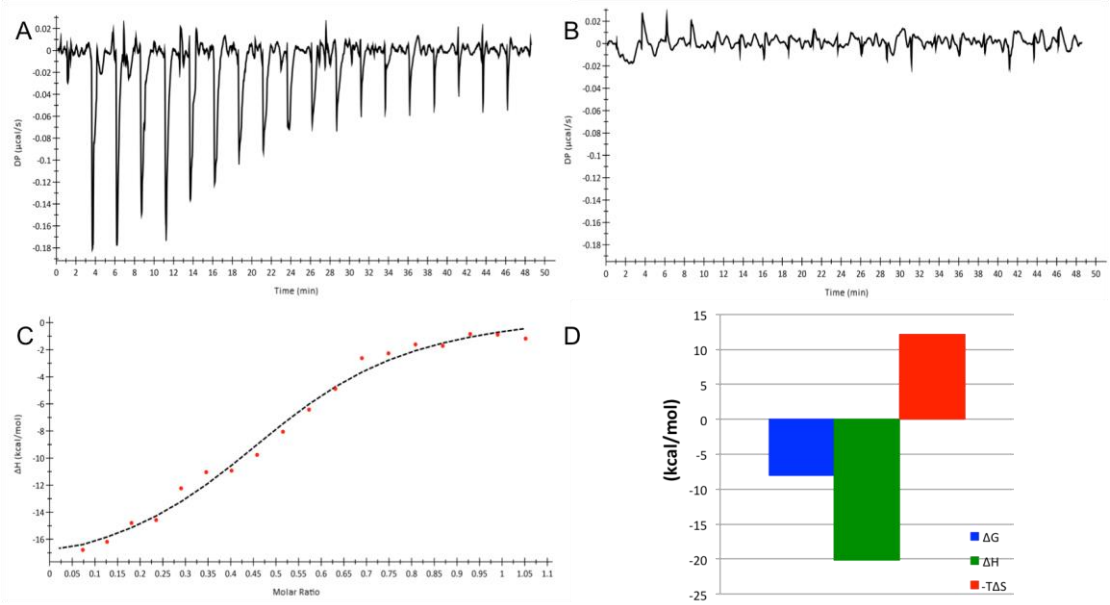



**Figure 5.9 Isothermal titration calorimetry (ITC) analysis for binary binding between KAP1 RBCC NtHis and SMARCAD1 CUE1 CtHis.** A) Heat rate change for each titration point in the ITC binding experiment between KAP1 RBCC NtHis and SMARCAD1 CUE1 CtHis, B) Heat rate change for each titration point between KAP1 RBCC NtHis and SMARCAD1 CUE2 CtHis, C) Binding thermogram for KAP1 RBCC and SMARCAD1 CUE1, from integrated heat data along with a nonlinear regression fit to a one-site binding model using MicroCal PEAQ-ITC Analysis Software (Malvern), D) Thermodynamic constants derived from the binding curve of KAP1 RBCC and SMARCAD1 CUE1 interaction.

Based on the thermodynamic constants derived from the binding isotherm it can be inferred that the interaction between KAP1 RBCC and SMARCAD1 CUE1 domains involves the formation of salt bridges or hydrogen bonds resulting in an overall favourable enthalpic change with the observed molar enthalpy change ($\Delta H_{obs}$) at (-84±8)kJ/mol. Similarly, the observed molar entropy change at 298K ($\Delta S_{obs}$) is 170J/mol, indicating an unfavourable change in the disorder propensity of the system, which might indicate a more structured bound state, perhaps involving flexible linkers or exposed hydrophilic residues in the free state which become locked in a fixed conformational arrangement once bound. Overall, the change in free energy is consistent with an exothermic spontaneous process, which is reflected by on the dissociation constant.

An equivalent assay, assessing the potential interaction between KAP1 RBCC and SMARCAD1 CUE2 was performed and resulted in no detectable heat exchanges beyond the potentially insignificant heat of dilution. The recorded heat rate changes for the mixture are part of Figure 5.9. From this, it can be inferred that despite physical proximity and structural homology between both CUE domains in SMARCAD1, their biological function is likely to be distinct and independent.

In an attempt to define the binding interface in SMARCAD1 CUE1 domain, a partial NMR titration was carried out with five concentrations of unlabelled KAP1 RBCC between 11µM and 108µM (monomer), against 45µM SMARCAD1 SIM/CUE1. With a focus of small chemical shift perturbations early in the titration experiment, individual concentration points were prepared as described in Chapter 2 and all titration experiments in Chapter 4. Briefly, two solutions of uniformly labelled SMARCAD1 SIM/CUE1 ($^{15}$N eCUE1) were prepared with one of them containing the highest concentration point of unlabelled KAP1 RBCC as ligand, and mixed in fixed ratios maintaining the labelled protein concentration constant through the experiment. 2D heteronuclear single quantum correlation (HSQC) spectra were collected of each titration point and the relative chemical shift changes were followed and plotted against ligand concentration.

Figure 5.10 presents an overlay of all six titration points, as well as a histogram of the relative intensity changes of $^{15}$N eCUE1 at two KAP1 RBCC domain concentrations. From this data it may be easily appreciated that most peaks corresponding to the unstructured regions of SMARCAD1 SIM/CUE1 are not significantly affected by KAP1 RBCC binding, whereas amide resonances belonging to residues in the structured CUE1 helical bundle become increasingly broadened upon binding and eventually disappear all together. Because the interaction affinity is fairly high, it is to be expected that residues involved in ligand recognition will not exhibit fast exchange between free and bound states within the time frame of the HSQC experiment, and so are more likely to lose intensity than to shift [223]. Additional differences between the free and bound state might be contributing to the broadening effect of

amide resonances, such as the relative ten-fold increase in apparent molecular weight once the CUE1 domain binds to the RBCC dimer, limiting its tumbling rate in solution.

On the other hand, the relative intensity of resonances perturbed upon binding supports the stoichiometry deduced for this interaction by ITC, given that most peaks are not, in fact, completely lost when ligand (KAP1 RBCC) and analyte ($^{15}$N eCUE1) are at a monomeric molar ratio close to unity (60µM KAP1 RBCC monomer relative to 45µM SMARCAD1 SIM/CUE1, ≈1.3 ratio) compared to fully bound state at 108µM KAP1 RBCC (monomer). This can easily be explained if the effective ligand concentration is halved by reconsidering it as a homodimer, in which case only when fully bound has the dimeric ligand to analyte molar ratio reached unity (54µM KAP1 RBCC dimer relative to 45µM SMARCAD1 SIM/CUE1, 1.2 ratio).

A histogram of relative intensity changes for resonances in the structured CUE1 domain of SMARCAD1 at each titration point allows for a more detailed and careful consideration of potential interacting surfaces recognising KAP1 RBCC dimer in solution. From Figure 5.11A, it is possible to appreciate the rate differences in broadening tendencies for a set of residues in SMARCAD1 CUE1, though they do not form a continuous patch along the primary polypeptide sequence. However, mapping these more strongly perturbed residues on the solution structure of CUE1 does appear to reveal a cluster of symmetrically distributed amino acids centred along the buried, solvent inaccessible surface of the second α-helix in the bundle. Figure 5.11B consist of alternate views of the cartoon representation of SMARCAD1 CUE1 domain with residues most strongly perturbed upon KAP1 RBCC domain binding coloured red. It becomes apparent from analysing these structures that one face of the helix bundle, specifically along the axis of CUE1 helix 2, may interact preferentially with the RBCC dimer, with additional contacts made by the solvent inaccessible surfaces of helix 1 and 3. Such perturbation patterns are consistent with the pull-down data in section 5.3, which shows that SMARCAD1 is specifically recognised by features exclusively present in the dimeric RBCC domain, possibly the

Coiled-Coil or the Ring domains, being the likeliest candidates to form homodimer-stabilising contacts [207, 208].

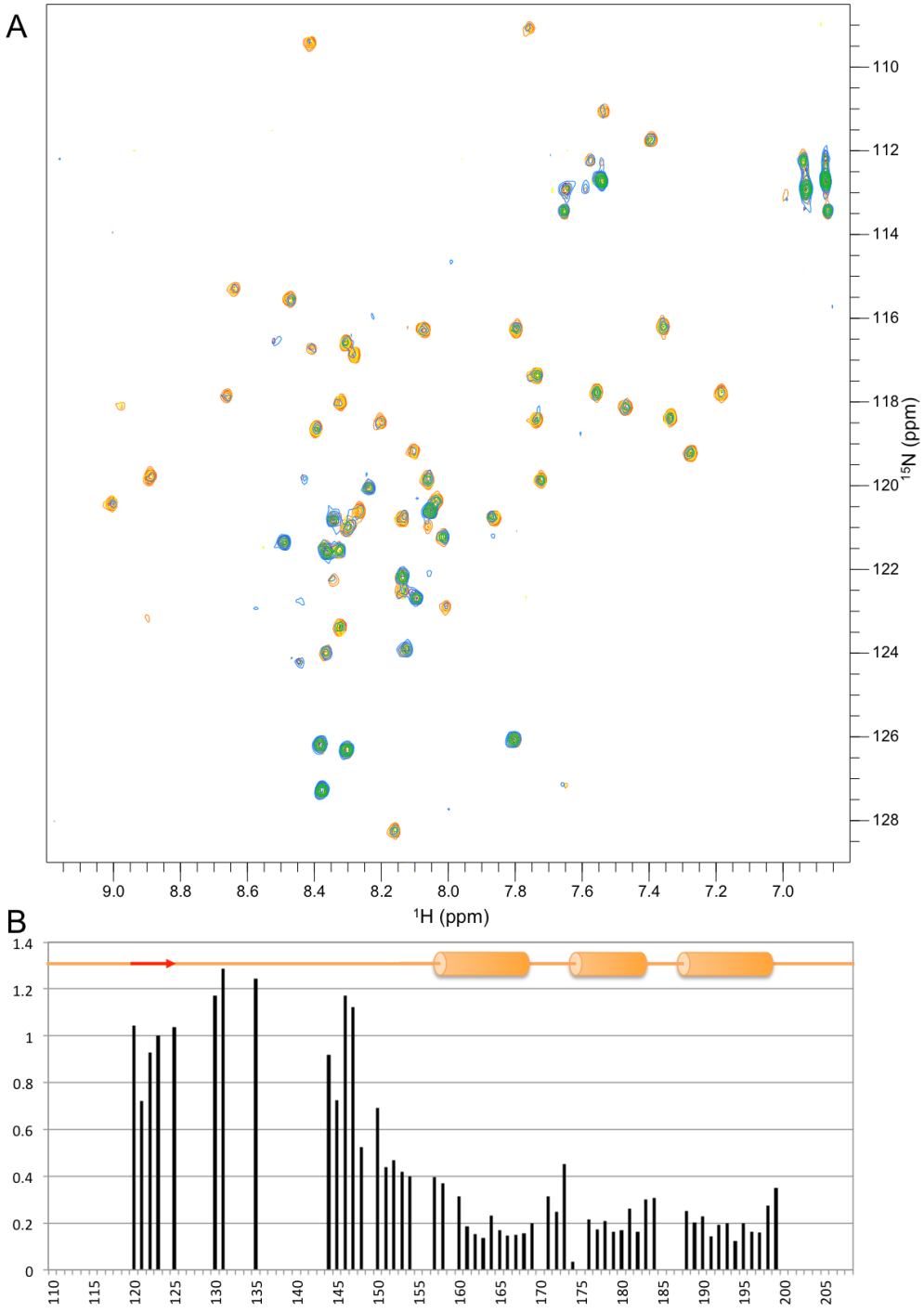



**Figure 5.10 NMR titration experiment of $^{15}$N eCUE1 with unlabelled KAP1 RBCC.** A) Overlay of six $^{1}$H,$^{15}$N HSQC spectra for SMARCAD1 SIM/CUE1, each with different, colour coded KAP1 RBCC concentrations. B) Histogram maximum the relative intensity change between two KAP1 RBCC titration points (5µM and 30µM, dimer).

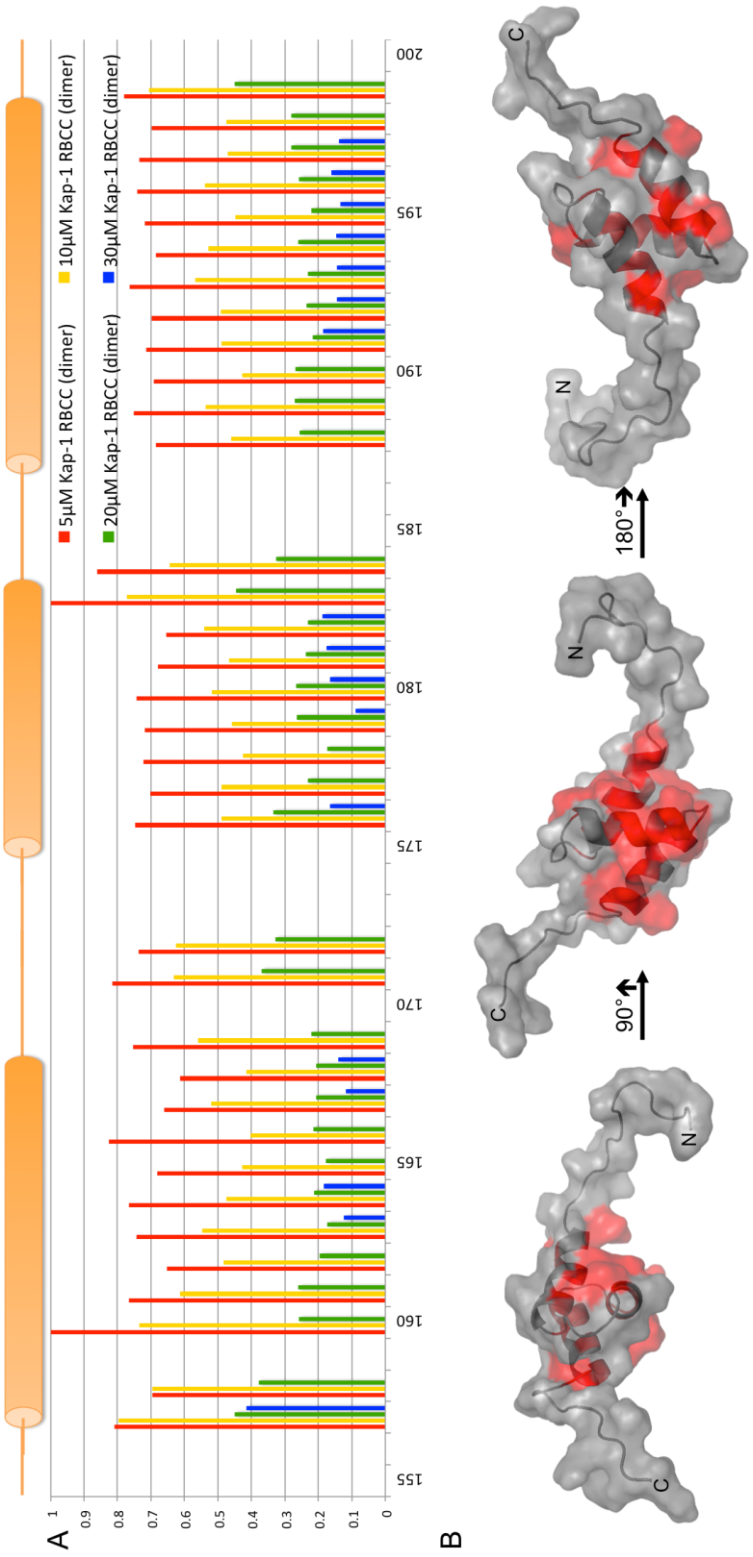



**Figure 5.11 Identification of KAP1 RBCC binding surface in SMARCAD1 CUE1 domain.** A) Relative intensity changes of residues in the structured region of SMARCAD1 SIM/CUE1 for each concentration point throughout the titration with KAP1 RBCC. B) Alternative views of SMARCAD1 CUE1 domain solution structure with strongly perturbed residues upon KAP1 RBCC domain binding coloured in red.

## 5.6 Crystallisation of the protein complex between SMARCAD1 CUE1 and KAP1 RBCC

Crystallisation trials were performed with purified recombinant SMARCAD1 CUE1 CtHis and KAP1 RBCC NtHis proteins, as described in Chapter 2. Briefly, freshly purified KAP1 RBCC NtHis domain, concentrated to approximately 80µM in 50mM TrisHCl pH 8.0, 500mM NaCl, 10% Glycerol, 1mM TCEP was mixed with SMARCAD1 CUE1 CtHis at 4.3mM in 20mM $NaH_2PO_4$ pH7.0, 100mM NaCl at a 1:1 equimolar ratio. This mixture proved adequate for crystallisation trials based on PCT™ Pre-crystallisation Test kit (Hampton Research), suggesting a particular susceptibility to polyethylene glycol precipitant conditions. Crystallization screening with commercially available reagents and vapour diffusion methods using nanolitre sitting drops were carried out at 20°C and 4°C, without removal of the hexahistidine affinity tag. Given the unusually high salt concentration in the protein drop, reservoirs were adjusted to approximately 500mM NaCl with a saturated sodium chloride stock solution (~5.5M).

Three conditions out of >1400 tested produced crystals within a week, all at 20°C. These were allowed to grow until ten days after setting up the trials before harvesting and flash freezing in liquid nitrogen prior to data collection at Diamond Light Source (Didcot, UK), using mother liquor supplemented with 30%$^{v}/_{v}$ glycerol or 20%$^{v}/_{v}$ Ethylene glycol as cryo-protectants. Figure 6.12 presents images of representative crystals for each condition, along with the specific formulation for the commercial crystallization buffer solution.

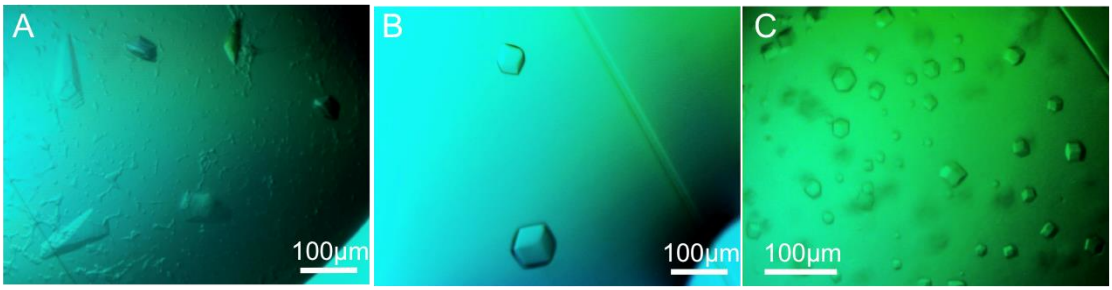



**Figure 5.10 Crystallisation of SMARCAD1 CUE1 and KAP1 RBCC protein complex. Images of Nanoliter crystallisation experiments using sitting drops and the vapour diffusion method.** A) Index (Hampton Research) Solution 59: 100mM HEPES pH 7.5, 20mM $MgCl_2$, 22%$^{w}/_{v}$ Polyacrylic Acid 5100. B) JBScreen Nuc-Pro (Jenna Bioscience GmbH) Solution 2/A6: 50mM MESNa pH 6.5, 80mM MgAcO, 30%$^{w}/_{v}$ PEG4000, C) ProPlex™ (Molecular Dimensions) Solution 31: 100mM Sodium Citrate pH 5.6, 20%$^{v}/_{v}$ 2-Propanol, 20%$^{w}/_{v}$ PEG4000.

Diffraction data was collected at Diamond Light Source (Didcot, UK) on Beamline I24. Crystals were poorly diffracting, particularly those from JBScreen Nuc-Pro (Jenna Bioscience GmbH) Solution 2/A6. Crystals from the other two conditions between 12-18Å, with those derived from Index (Hampton Research) Solution 59 consistently producing diffraction to 14Å, whereas those from ProPlex™ (Molecular Dimensions) Solution 31 oscillated between more extreme diffraction limits. The best sample, derived from this latter condition, diffracted to 11.6Å and automatic data analysis using Mosfilm suggests the crystal belongs to space group C 2 2 $2_1$, with a unit cell of 198Å x 279Å x 332Å 90°x90°x90° (Figure 5.13). Though the quality of the diffraction pattern is very poor, it is possible to estimate a potential unit cell content. Thus, assuming a 50% solvent content and a Matthews coefficient of 2.47 [224], the unit cell contains a protein complex of 900kDa, which could be interpreted as eight dimers of approximately 112.5kDa, roughly consistent with the molecular weight of the KAP1 RBCC domain dimer bound to the SMARCAD1 CUE1 domain, though this is mere speculation.

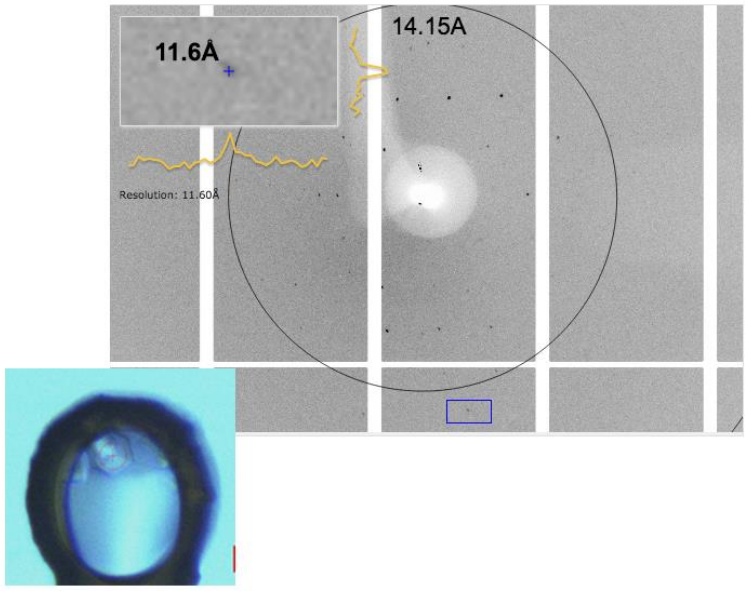



**Figure 5.11 Data collection of initial crystal hits for the SMARCAD1 CUE1 and KAP1 RBCC protein complex.** Diffraction image from a single cryo-frozen crystal analysed at 100K using synchrotron radiation at Beamline I24 (Diamond Light Source, UK). The crystal was cryo-protected in 20%$^{v}/_{v}$ Ethylene glycol before being exposed to X-ray radiation at a wavelength of 0.9686Å. Resolution circles are shown as continuous lines and are labelled on the diffraction image.

In an effort to improve the quality of diffracting crystals all conditions were subjected to optimisation rounds using additives and crystal seeding [128, 130], the latter making use of crystals previously used for data collection. Additionally, all conditions were further optimised using a three-row dilution step optimisation, as described in Chapter 2. Optimisation trials failed to yield any crystals even after several months. Though the original conditions were reproducible, crystals thus produced were of the same poor quality as those previously tested.

Multiple factors were likely at play preventing the crystallisation of the SMARCAD1 CUE1 and KAP1 RBCC protein complex to a sufficient quality for successful high-resolution diffraction, from high salt content to multiple flexible linkers connecting the various subdomains in the RBCC domain of KAP1. It could be possible that dialysis of the RBCC construct, even at high concentrations, might have allowed lowering salt concentrations while maintaining a stable soluble fraction, though

while dialysing the protein in preparation for ITC experiments against a low salt buffer, approximately 50% of the protein content on average (as measured by UV/Visible absorbance quantitation) would precipitate. Perhaps a more effective alternative would be the re-engineering of the RBCC construct into truncated forms with shorter linkers between subdomains. Design of these constructs could be guided by limited proteolysis experiments with the goal of identifying a structured core [225], coupled to pull-down experiments (such as those described in Section 5.3) to screen against truncation mutants incapable of adopting the native fold and thus would fail to recruit SMARCAD1.

## 5.7 Immunofluorescence analysis of SMARCAD1-KAP1 interaction in chromatin

In order to expand on the biological role of the SMARCAD1-KAP1 complex beyond that of the established context in DNA replication [51] immunofluorescence assays were performed to determine cellular co-localisation in cultured human cells. In the first instance, cultured HeLa cells were fixed, permeabilized and stained with primary antibodies against SMARCAD1, KAP1, Histone H2A.X phosphorylated on Ser140 (γH2A.X) or Histone H3 trimethylated on Lys9 (H3K9me3), followed by appropriate host-specific secondary antibodies conjugated to Alexa Fluor® dyes and nuclear counterstaining using DAPI (Chapter 2). Figure 6.14 depicts representative co-localisation images for each set of stained proteins in three immunofluorescence experiments performed in parallel.

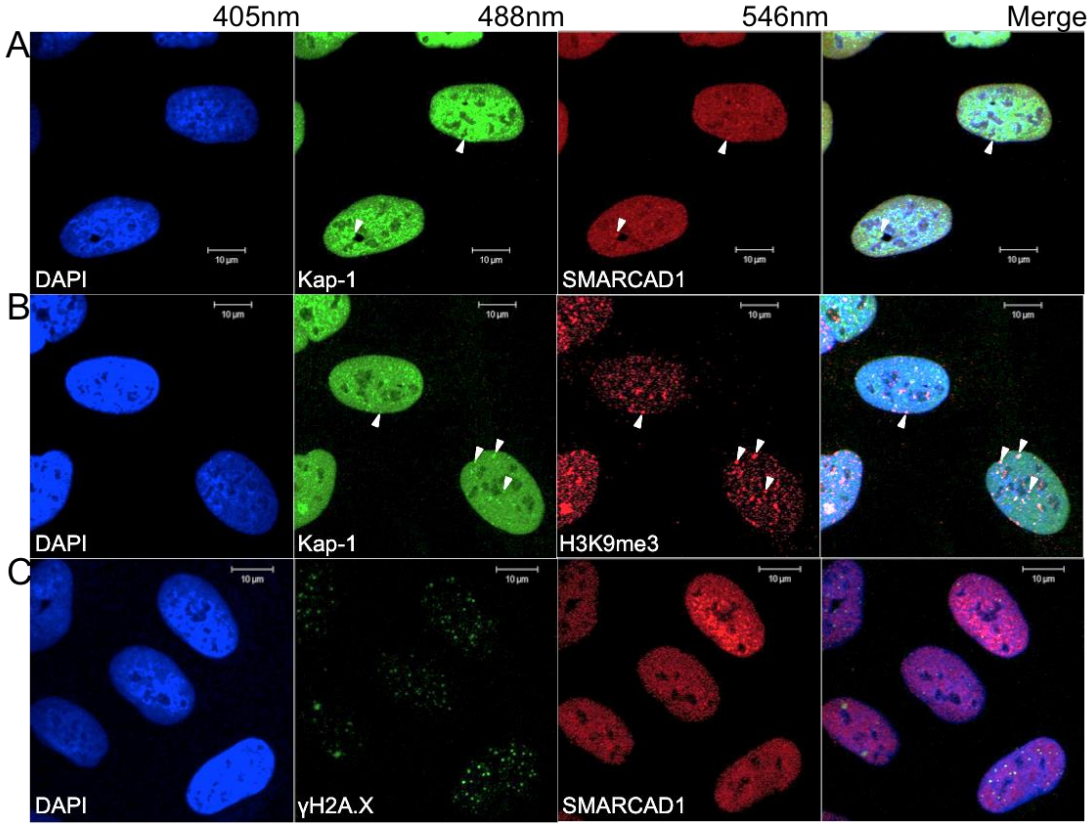



**Figure 5.12 Immunofluorescence microscopy co-localisation of SMARCAD1 and Kap1 on fixed and permeabilized HeLa cells.** Panels correspond to a single excitation wavelength or an overlay of all three channels, as labelled. HeLa cells stained for A) DAPI/KAP1/SMARCAD1, B) DAPI/KAP1/H3K9me3 or C) DAPI/γH2A.X/SMARCAD1. White arrowheads correspond to foci that overlay well on both target specific channels, with a correlated arrow signalling the same foci on the merged image. Images were collected on a Zeiss LSM510 confocal microscope.

Figure 5.14 makes immediately apparent how the uniform labelling of samples stained for KAP1 or SMARCAD1 make almost impossible the confident assignment of co-localised foci. Even in a sample with clearly formed stained clusters of H3K9me3, its co-localisation with KAP1 comes from signals whose intensity hardly rises above the background as evident in Figure 5.14B (white arrowheads). This is particularly detrimental in Figure 5.14A where the entirety of the nucleus could be said to indicate co-localisation for KAP1 and SMARCAD1, yet very few foci are confidently detected to overlay in the merged image. Finally, Figure 5.14C could be interpreted to indicate that SMARCAD1 does not co-localise with γH2A.X foci, though the fact that these samples have not been treated to induce DNA damage means that these foci may not represent active repair factories but rather a non-canonical role in chromatin regulation during mitosis [18, 226] and thus leave room for further elucidation.

In order to overcome the limitations on immunofluorescent assignment of co-localised signals by an overcrowded nuclear environment, a Proximity Ligation Assay (PLA®, Sigma-Aldrich) was performed to selectively visualise only the nuclear sub-populations of KAP1 and SMARCAD1 involved in a binary complex. Samples were fixed, permeabilized and stained simultaneously with primary antibodies against KAP1 and SMARCAD1, followed by proximity probe conjugated *in situ* PLA® secondary antibodies, with the detection of protein pairs within 30nm of each other [227] through the hybridization of circularization oligonucleotides, which are then joined by ligation into a circular DNA molecule. Rolling circle amplification creates a concatemeric amplification product that remains covalently attached to one of the proximity probes and is subsequently detected with complementary fluorophore-labeled hybridization oligonucleotides [228]. Preliminary PLA® analysis of the binary SMARCAD1-KAP1 complex, along with technical and biological negative controls are presented in Figure 5.15.

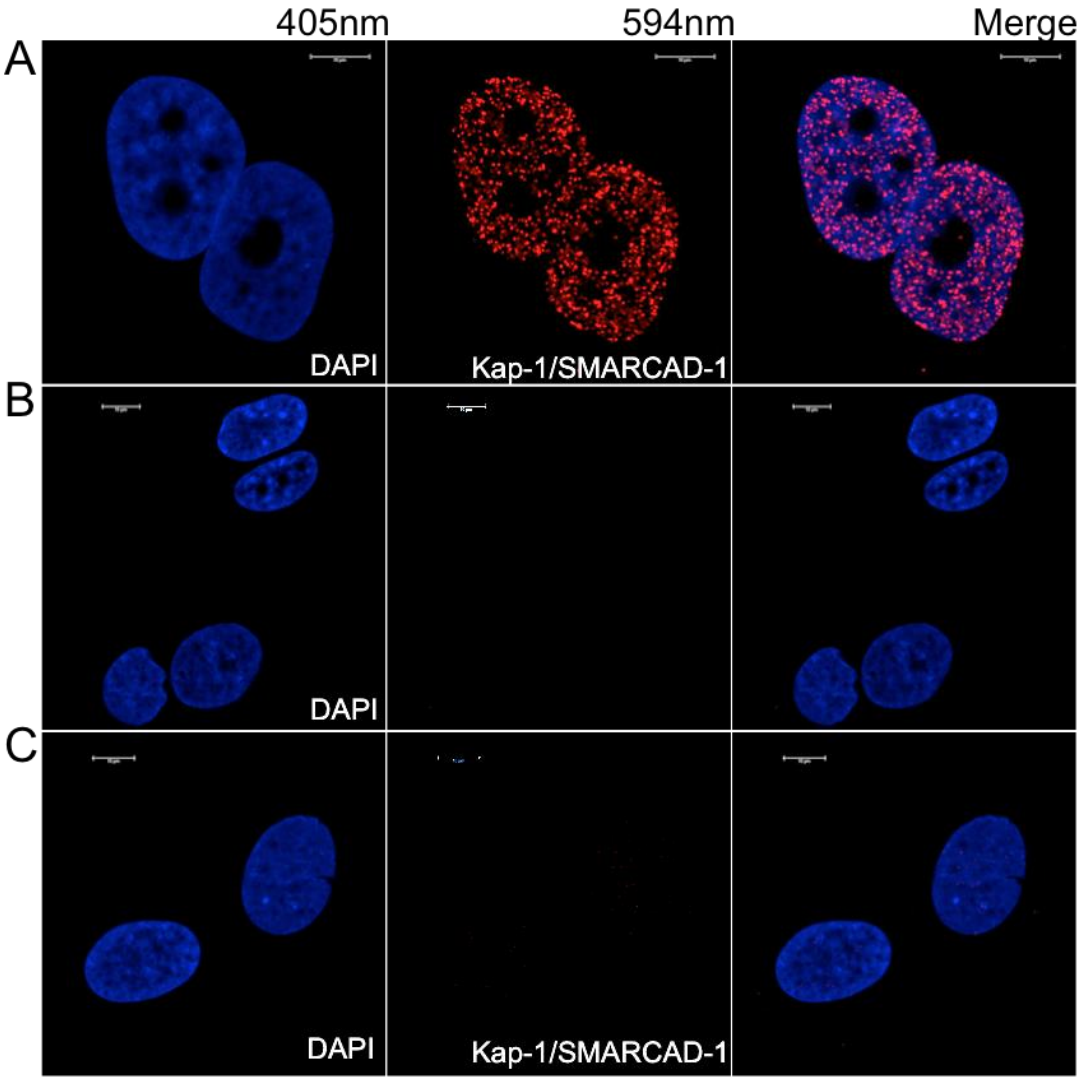



**Figure 5.13 Proximity Ligation Assay of the SMARCAD1-KAP1 complex.** Panels correspond to a single excitation wavelength or an overlay of both channels. A) HeLa cells stained for DAPI/KAP1/SMARCAD1, B) Technical negative control, HeLa cells stained for DAPI but not any primary antibodies, C) Biological negative control, HeLa cells, +72h after transfection with siGENOME siRNA SMARTpool against SMARCAD1, stained for DAPI/KAP1/SMARCAD1. Images were collected on a Leica TCS/WLL SP8 confocal single molecule detection microscope.

PLA® analysis of SMARCAD1-KAP1 protein complex displays a large number of fluorescent spots uniformly distributed across the nucleus, indicating this complex is abundant and evenly represented. Comparison of the overlaid panel in Figure 5.14A and 5.15A emphasises the distinct superiority of the PLA to classic IF with respect to spatial resolution of co-localising agents. By measuring close proximity, the PLA allows a validation of the molecular proximity between endogenous proteins and effectively screens out the fractions of interactive partners not actively involved in complex formation. Furthermore, the specificity of the assay is evidenced by the technical and biological negative controls performed in parallel. The technical control, Figure 5.15B, consists of the omission of primary antibodies so as to exclude cross-reactivity of secondary PLA® antibodies. Similarly, the biological control involved siRNA mediated knock-down of SMARCAD1. In this case both primary and secondary antibodies were applied but due to the absence of one of the antigens, no signal derived from the PLA® should be detected.

Finally, it is possible to extend the use of the PLA® selective detection of SMARCAD1-KAP1 in complex by analysing co-localisation patterns in cells through the use of target specific antibodies generated in alternative host systems, different from those used in the PLA, or with direct labelling of primary antibodies. The latter method was used for co-staining with γH2A.X and H3K9me3, which were labelled using Zenon® anti-rabbit Fab fragments, each conjugated to a different Alexa Fluor® dye. Figures 5.16 and 5.17 show the immunofluorescence co-localisation of PLA® detected SMARCAD1-KAP1 complex in representative HeLa and U2OS cultured cells, respectively.

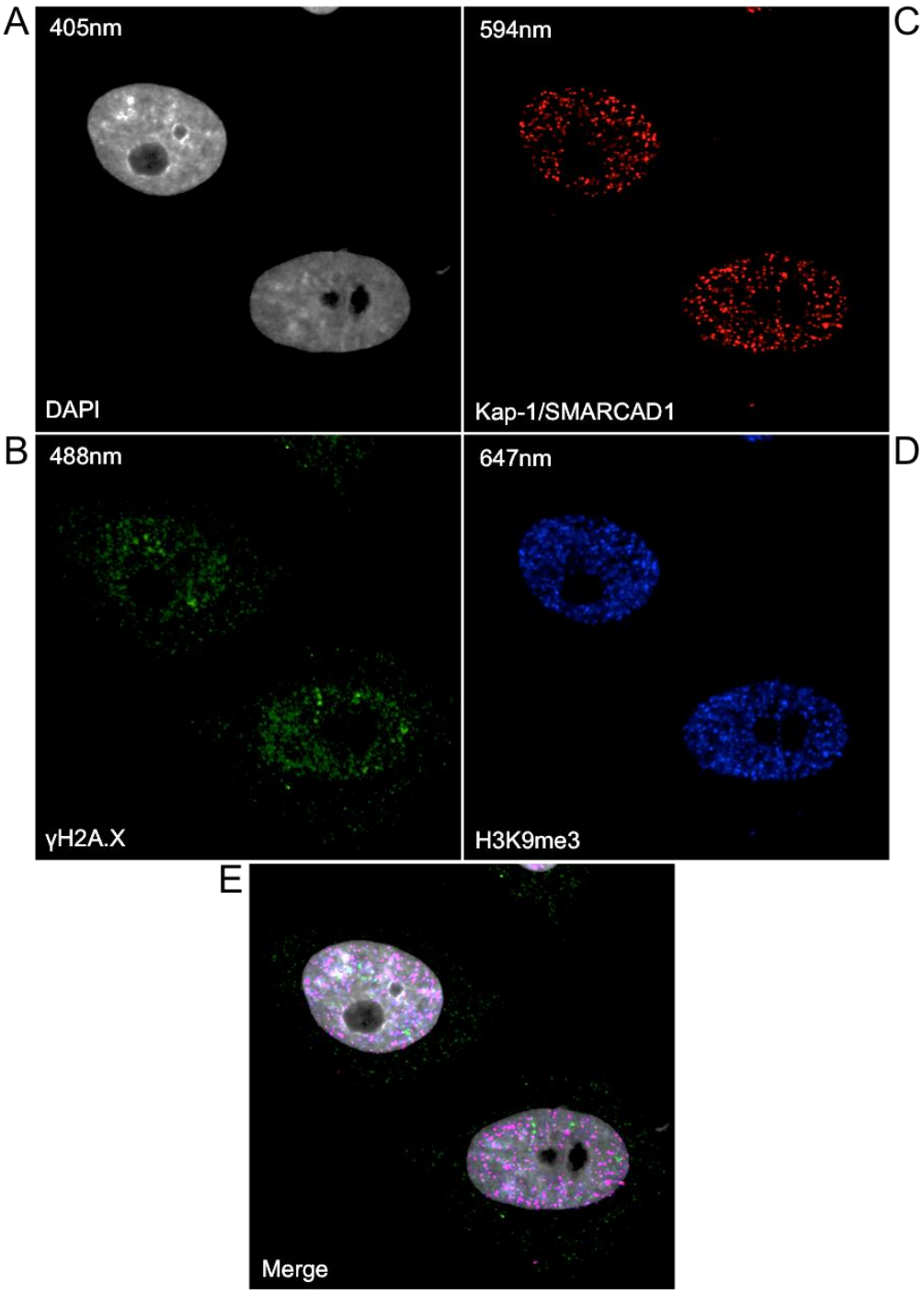



**Figure 5.14 HeLa cells analysed by immunofluorescence co-localisation of PLA®-detected KAP1-SMARCAD1 complex with γH2A.X and H3K9me3 directly labelled using Zenon® anti-rabbit Fab fragments, each conjugated to Alexa Fluor® 488 and 647, respectively.** Panels correspond to individual channels or an overlay of all four, target antigens are labelled accordingly.

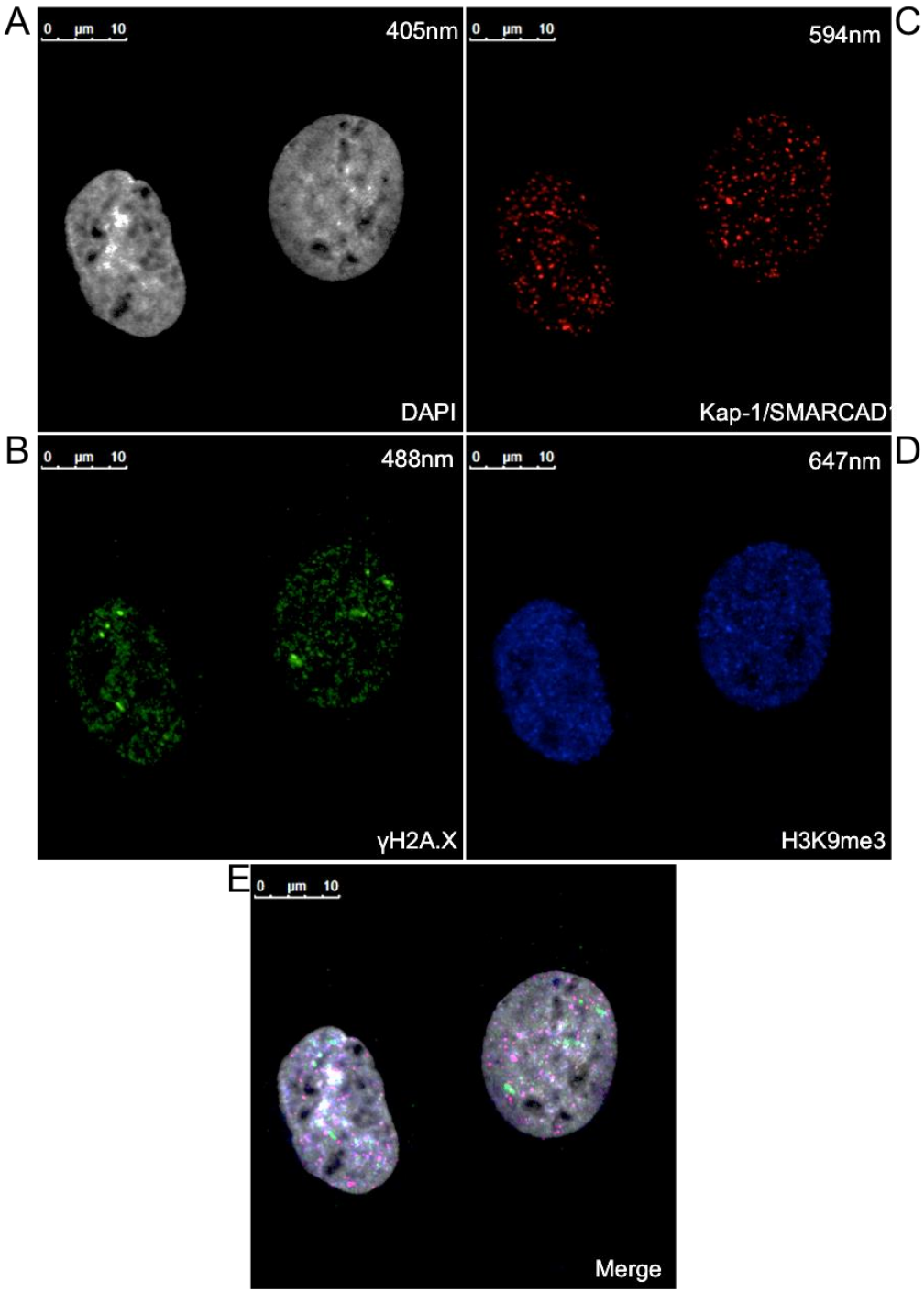



**Figure 5.15 U2OS cells analysed by immunofluorescence co-localisation of PLA®-detected KAP1-SMARCAD1 complex with γH2A.X and H3K9me3 directly labelled using Zenon® anti-rabbit Fab fragments, each conjugated to Alexa Fluor® 488 and 647, respectively.** Panels correspond to individual channels or an overlay of all four, target antigens are labelled accordingly.

It is easily appreciated that the PLA® targeted SMARCAD1-KAP1 complex co-localises completely and perfectly with the heterochromatin marker H3K9me3. This is consistent with its composition, given the well-established cellular localisation for KAP1, however these results indicate that at least a fraction of such heterochromatin bound KAP1 exists as a functional complex with SMARCAD1, at least in cells that have not been treated to induce the DNA damage response. Such a clear and unequivocal description of a three-partner interaction is in sharp contrast with the ambivalent and the poorly defined immunofluorescence co-localisation on Figure 5.14 between KAP1 and H3K9me3. On the other hand, no apparent co-localisation may be appreciated between SMARCAD1-KAP1 complex and γH2A.X. As pointed out above, these foci (present in cells without a DNA damage response induced by external sources such as ionizing radiation or a chemical agent such as Neocarzinostatin) do not necessarily represent active repair factories and may in fact be an example of non-canonical γH2A.X function. These results would benefit from complementary experiments in which equivalent analysis was performed on cells whose DNA repair programme has been induced, to compare contrasting phenotypes and co-localisation patterns under such cellular response.

## 5.8 Conclusions and future work

The biochemical characterisation of the binary interaction between SMARCAD1 and its *bona fide* interacting partner KAP1 described in this chapter has important repercussions for our understanding of this chromatin remodeler, challenging long standing assumptions of its modular binding domains. The unequivocal fact that the CUE1 domain in SMARCAD1 mediates this interaction with a non-ubiquitylated binding partner suggests that perhaps the structural homology to the CUE family of proteins may not limit its functional role solely to that of an ubiquitin-binding domain. This is further highlighted by the sharp difference in affinities against recombinant Ubiquitin or KAP1 RBCC domain (close to three orders of magnitude) as determined by NMR and ITC, respectively. What consequences such disparity may have in the nuclear distribution of SMARCAD1 or how it limits the type of multi-subunit protein complexes with which it is capable of interacting still needs to be

assessed. To do this, however, it is necessary to identify and decouple the Ubiquitin and KAP1 recognition surfaces given that each interaction likely engages distinct (though possibly overlapping) residue clusters.

One crucial step in achieving such understanding is the successful expression of stable, recombinant constructs of KAP1 RBCC from baculovirus infected insect cells. This construct has been shown to be dimeric in solution by SEC-MALLS and immunoprecipitation assays demonstrate that this oligomeric state is required for SMARCAD1 recruitment. Further studies of the KAP1 RBCC domain by complementary techniques such as SAXS and AUC would complement the biophysical characterisation of this domain, and significantly expand on the limited molecular description of the core domain of the large and varied TRIM family of proteins.

Furthermore, NMR perturbation analysis of the interaction between 15N-labelled SMARCAD1 CUE1 domain and unlabelled KAP1 RBCC suggests that the molecular recognition in SMARCAD1 takes place through symmetric patches along two sides of the pyramidal volume created by the CUE1 three-helix bundle, consistent with the hypothesis of a symmetric surface in the dimeric RBCC structure. These results indicate that KAP1 binding surface on CUE1 is considerably larger that that for Ubiquitin, and does share a few key residues involved in recognising the latter, which could mean that both binding events are mutually exclusive, though it could also mean that there are at least two subpopulations of SMARCAD1 engaged in mutually exclusive interactions. Though crystallisation of the binary complex failed, initial hits suggest that optimisation of the RBCC domain, whether through chemical modification of the current construct (surface Lysine methylation or controlled partial trypsination) or requiring re-cloning to remove the longer and more flexible linkers connecting the RBCC subdomains, could potentially lead to better quality crystals which might diffract to high resolution. Such endeavours are commendable and highly desirable since they would represent the first high-resolution structure for a full-length TRIM domain, in addition to providing key information on the specific case of SMARCAD1-KAP1 interaction.

Finally, it has been possible to confirm the co-localisation of the binary SMARCAD1-KAP1 complex on trimethylated nucleosomes, a key feature of heterochromatin. This distribution probably involves a functional role for this complex beyond the re-establishment of silent epigenetic marks upon replication of transcriptionally repressed sections of the genome. Identifying what these functions might be or how the co-localisation phenotype might change upon DNA damage requires further work. Moreover, understanding how these phenotypes depend on each of the binding domains needs to be addressed. This question in particular could benefit from the PLA® immunofluorescence detection employed in this work, given that it could bypass the need to selectively knock down the endogenous protein through RNA interference followed by rescue with resistant mutants by use of primary antibodies that recognise a tagged version of the protein (with or without functional mutations), effectively screening out the endogenous population. Describing the domain dependency of various SMARCAD1 phenotypes following a stimulus (e.g. DNA damage) would provide multiple layers of information to better understand the role that this unique, non-canonical remodeler plays in the cell.

Chapter 6

# Discussion

## 6.1 SMARCAD1 CUE1 domain adopts a triple-helix bundle conformation

The solution structure for the CUE1 domain from SMARCAD1 was determined through a molecular dynamics based simulated annealing protocol in ARIA [124] to minimize a global energy function with distance constraints derived from aliphatic and aromatic $^{13}$C-edited NOE assignments, empirical residual dipolar couplings (RDC), TALOS+ predicted dihedral angles and α-helical Hydrogen bond networks delimited based on experimentally determined three-bond $H^N$-Hα coupling constants ($^3J_{HNHA}$). The ensemble model of 25-lowest energy structures shows low intermolecular variation within the structured region of the sequence, particularly for backbone atoms (<0.5Å). Similarly, the globular fold and helical packing is consistent with structures for other known members of the CUE family, with root-mean-square deviation (RMSD) between 2.5-3.5Å as evidence of relatively low intermolecular variation within the compared structures. As shown in Figure 6.1, the main difference between all CUE domain models derives from the specific angles between the two somewhat parallel helices in the domain fold, from fairly flat well-aligned helices as in *Hs* CUEDC1 through to more concave packing such as *Hs* AMFR. SMARCAD1 CUE1 shows a clear concave helical packing diverging by approximately 53° from ideal parallel conformation.

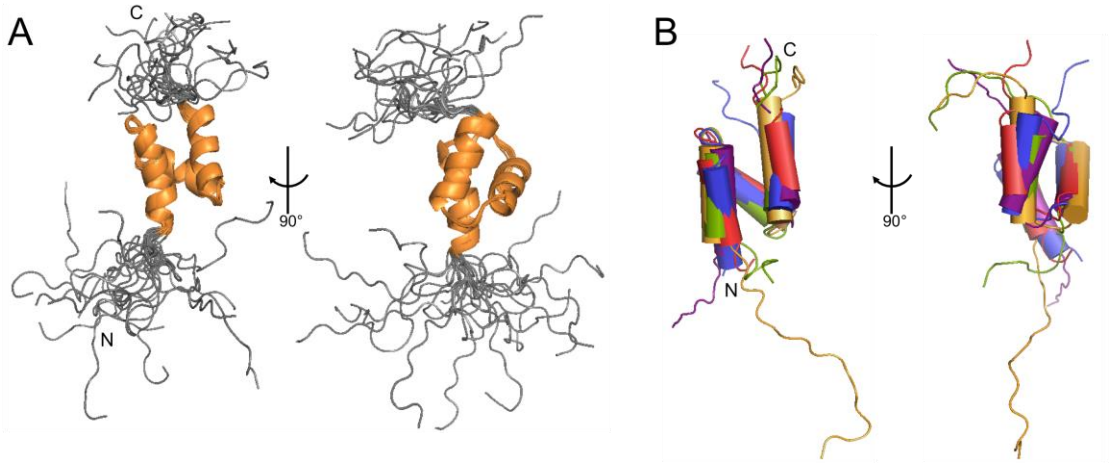



**Figure 6.1 SMARCAD1 CUE1 low-energy ensemble.** A) Cartoon representation of the SMARCAD1 CUE1 domain, with 25 structures shown as an ensemble with ordered residues (158-198) in orange and more flexible, non-helical residues in grey, B) Energy minimised model for SMARCAD1 CUE1 (purple) was superimposed with the deposited structures for the N-terminal CUE domain from yeast Cue2 protein (red), the CUE domain from human CUEDC1 (CUE Domain Containing Protein 1; yellow), the CUE domain form human TOLL-interacting protein (TOLLIP, green) and the CUE domain from human E3 ubiquitin-protein ligase AMFR (blue), with helices drawn as cylinders.

Furthermore, this model addressed the conformational ambiguity that was the result of previous efforts in the Mancini lab to resolve the SMARCAD1 CUE1 domain solution structure. Poorly defined structural elements, likely the unintentional consequence of insufficient data, resulted in potentially equivalent families of structures as illustrated in Figure 6.2. Unlike those calculations, the present structure benefited from additional sources of structural restraints including a larger number of methyl-methyl and aromatic NOE assignments, which reference the internal packing of buried residues within the helix bundle, and an extended set of RDCs that better define the general orientation of structured residues relative to an external common frame, which considerably sharpens the definition of NMR-derived structures.

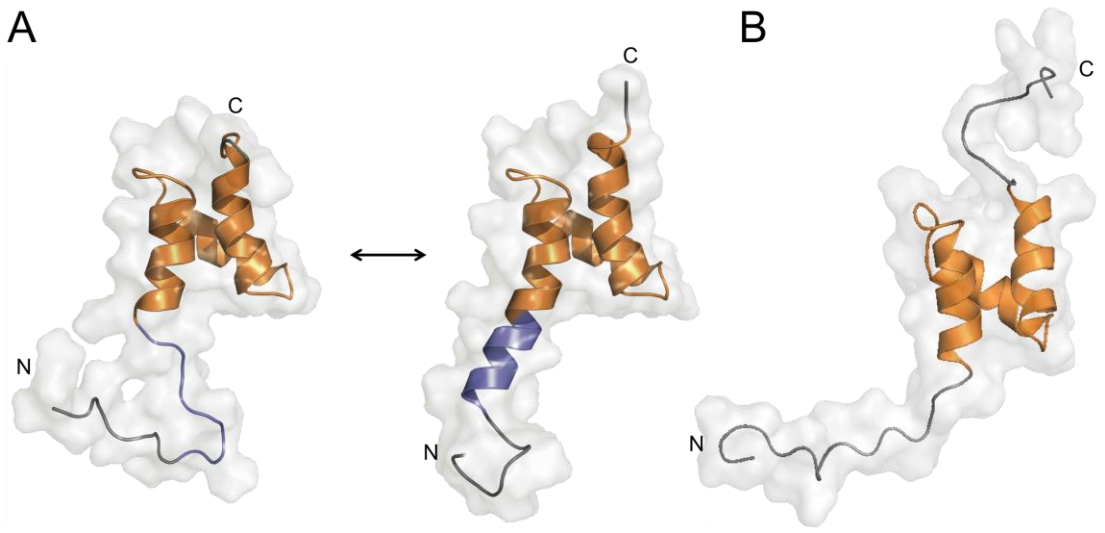



**Figure 6.2 Comparison between SMARCAD1 CUE1 domain solution structure models.** A) Cartoon representation of CUE1 solution structure from West 2012 [135], with structured residues in orange, flexible ones in grey and ambiguously defined residues 150-158 in mauve. B) Cartoon representation of improved CUE1 structure, with structured residues 158-198 in orange and flexible coil regions in grey.

## 6.2 Novel SMARCAD1 SIM is largely flexible and constitutively phosphorylated

Bioinformatics and solution NMR analysis of the SMARCAD1 primary sequence immediately N-terminal to the first CUE domain revealed the presence of a previously unassigned SUMO interacting motif (SIM), which was confirmed to bind *in vitro* to SUMO1. The unmodified recombinant protein sequence (including the 119TVIIVSE125 SIM region) binds to SUMO1 with an average dissociation constant (Kd) of 1.3mM.

Moreover, this region also includes the three Serine residues that were previously identified to be targeted by DNA damage response (DDR) kinase ATM through phospho-proteomics analysis [66], and have been previously proposed to target SMARCAD1 to the homologous recombination (HR) pathway for DNA repair [61, 62, 229]. Given that phosphorylation of regions flanking the hydrophobic patch of SIMs is a well-established mechanism to increase its affinity by introducing ionic charges complementary to those in SUMO [100], a hypothesis was proposed, that such an

event might act as a regulatory mechanism through which SMARCAD1 might be recruited to the DDR. NMR titrations with mono-phosphorylated peptides, that included both the SIM and potentially targeted Serine residues, indicated that phosphorylation of Ser124 (closest to the hydrophobic SIM centre) led to a two-fold increase in SUMO1 affinity. The effect of alternative phosphorylation on Ser127 and ATM canonical target site Ser132 was not clear, with these modifications resulting in significantly lowered affinity for SUMO1 or no change in affinity at all. The apparent ambiguity in the effect of each phosphorylation might be justified when considering the mechanisms for SUMO recognition of SIM sequences. It has been well established that for SIM interactions of particularly high affinity the combined hydrophobic and ionic interaction restrict the orientation of the SIM sequence to a single conformation [182], whereas in cases with fewer contacts the SIM might be able to bind in both a parallel or antiparallel orientation relative to the SUMO β-sheet, likely in a dynamic equilibrium between the two states [100]. In the case of SMARCAD1, phosphorylation at Ser124 likely contributes to increased ionic contacts within the SUMO1 binding cleft, where phosphorylation at Ser127 and Ser132 would not.

Considering that all three Serine residues where identified to be simultaneously phosphorylated upon DNA damage, the effect that such cellular process might have on SMARCAD1 phosphorylation state was evaluated through mass spectrometry analysis of immunoprecipitated GFP-SMARCAD1, transiently expressed in HEK 293T cells which were exposed to NCS, a DNA damage inducing chemical agent. The comparison between NCS treated and un-treated samples indicated that SMARCAD1 is constitutively phosphorylated on all three Serine residues in question, and no difference was detected upon NCS treatment. This led to the design of a triple-phosphorylation mimic in which Ser124, Ser127 and Ser132 were mutated to Glutamate and its analysis of SUMO1 binding, resulting in a dramatically higher affinity with an average dissociation constant (Kd) of 80μM. These results clearly indicate that SMARCAD1 is capable of mediating interactions with SUMO at biologically relevant affinities and rules out the ATM phosphorylation of these sites as the targeting mechanisms for DDR recruitment.

## 6.3 SMARCAD1 CUE1 domain is a poor mono-ubiquitin binder but mediates a high affinity interaction with KAP1 dimeric RBCC domain

CUE domains, ascertained as canonical ubiquitin binding domains (UBDs), have been extensively studied and the interaction of many of its members with ubiquitin well characterised. In an effort to similarly corroborate their predicted function the tandem SMARCAD1 CUE domains (CUE1 and CUE2) were analysed as a single construct, which also included the SIM sequence immediately N-terminal to them, against recombinant ubiquitin in an NMR titration experiment. This showed that both CUE domains were selectively perturbed along the predicted binding residues according to sequence conservation alignment with other members of the family, while SIM was not affected in accordance with the distinct modes of interaction by SUMO and ubiquitin despite sharing a tertiary fold. However, these results also show that SMARCAD1 CUE domains are particularly poor mono-ubiquitin binding modules, with CUE1 revealing an average dissociation constant (Kd) of 1.2mM, and CUE2 betraying an interaction so weak it could not be unequivocally determined within the ligand concentration range and may only be extrapolated to approximately 6mM. Neither of these binding affinities appears to be within biological relevant ranges, even for notoriously weak interactions such as those with ubiquitin, where affinities higher than low micromolar might result in unspecific or unregulated associations with overly abundant ubiquitin inside the cell [92-94]. Despite this, recent evidence [62] indicates that specific mutagenesis on both domains compromises the ability of SMARCAD1 to both bind to C-terminally ubiquitylated H2A histones and evict 53BP1 from double-strand break sites, required for HR directed resection and repair. This might indicate that SMARCAD1 is capable of binding to a specific ubiquitylated protein with higher affinity than that shown for mono-ubiquitin *in vitro*. Alternatively, it might be a reflection of binding to poly-ubiquitylated H2A, including at multiple mono-ubiquitylation sites or poly-ubiquitin chains [230, 231].

Additional evidence in the form of domain specific protein interaction profiles determined through mass spectrometry analysis of GFP-SMARCAD1 WT and CUE1 F169A mutant, transiently transfected in HEK 293T cells and exposed to chemically induced DNA damage through NCS treatment, unequivocally indicates that CUE1 domain mediates the interaction with its *bona fide* stoichiometric partner KAP1. This complex was shown not to be dependent on post-translational ubiquitylation, but rather it is a direct interaction with KAP1 reproduced *in vitro* with recombinant proteins.

Interestingly, SMARCAD1 co-immunoprecipitation with GFP-fusion proteins of KAP1 truncation mutants not only show that it is the RBCC multi-subunit domain which sustains this interaction but that all subdomains are necessary, with individual Ring, tandem B-Box 1 and 2, and Coiled-Coil constructs and combinations thereof failing to pull-down endogenous SMARCAD1 from HEK 293T cell lysates. This suggested that the binary interaction with SMARCAD1 was dependent on the oligomerisation of the TRIM motif, which in itself required further elucidation given conflicting reports suggesting that KAP1 was a trimer in solution while most TRIM proteins have been reported to act as anti-parallel dimers based on crystal structures from various coiled-coil domains [204, 212, 214]. The notable exception being TRIM5α, which has recently been characterised as a trimer or dimers forming a hexameric lattice along the capsid of viruses such as HIV-1 and triggering the innate immune response upon target recognition [215, 218, 232].

Heterologous expression of a recombinant KAP1 RBCC construct and subsequent biophysical characterisation of this complex identified this domain as a stable dimer in solution and defined the stoichiometry of the interaction at 2:1, that is one RBCC dimer per SMARCAD1 CUE1 domain. In contrast to its interaction with mono-ubiquitin, the association between SMARCAD1 CUE1 domain and KAP1 RBCC presented an equilibrium constant (Kd) of 2μM, determined by ITC. In support of the proposed organisation of KAP1 RBCC domain, a recent SAXS model of the homologous TRIM25 RBCC domain was shown to be an antiparallel dimer with two surface contacts stabilising it: the coiled-coil and Ring domains, while the tandem

Zinc-finger B-Box was fairly flexible connecting the two protein ends while folding in on itself (Figure 6.3) [208].

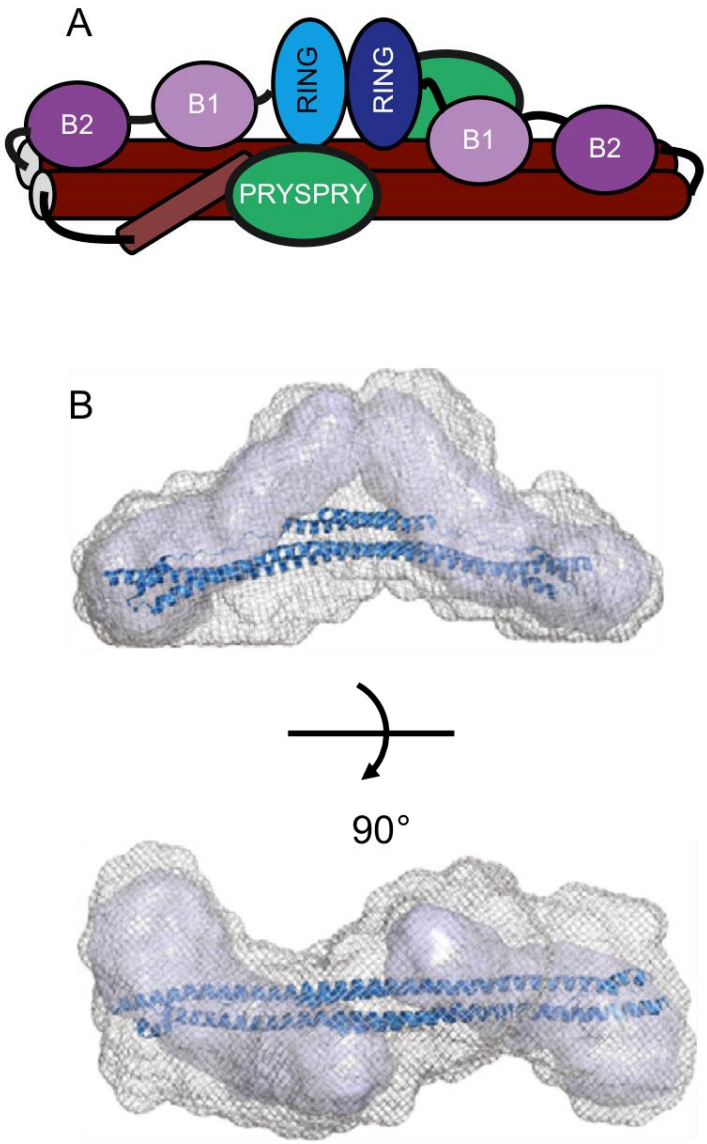



**Figure 6.3 Solution structure of KAP1 homologue TRIM25.** A) Schematic representation of the dimeric TRIM25 RBCC solution structure, with individual subdomains identified as Ring (in blue), B-Box 1 and 2 (as B1 and B2 respectively, in purple) and the coiled-coil domain with its C-terminal PRYSPRY fold in brown and green. B) Front and top views of the SAXS envelope for the RBCC domain in grey, the monomeric Ring/B-Box1/B-Box2 RBCC fragment in purple and the cartoon representation of TRIM25 coiled-coil (PDB: 4LTB) in blue. Adapted from Koliopoulos *et al* 2016 [208].

## 6.4 SMARCAD1 regulates p53 transcriptional repression

SMARCAD1 has been frequently proposed to have a direct role in DNA repair through functional extrapolation from its yeast ortholog FUN30, a well characterised chromatin remodeler known to influence long-range resection of DSBs during HR repair. This, coupled with the proposed targeting by ATM kinase, led our group and others to attempt to define the potential role SMARCAD1 might play within this broad range of pathways. Two reports [61, 62] have shown consistent evidence that places SMARCAD1 at the stage of DSB resection, specifically downstream from BRCA1-BARD1 ligase activity. Additional recent studies have indicated a SMARCAD1 might play a role in breast cancer metastasis and could potentially be targeted as a therapeutic agent [229]. These considerations motivated us to develop a proteomics approach to identify DNA damage specific binding partners from co-immunoprecipitated GFP-SMARCAD1 samples. Moreover, the domain specificity of each potential interaction was probed through the use of CUE1 and SIM functional abrogating mutants.

Mass spectrometry profiling of domain dependent interactions based on differential abundance relative to changes due to chemically induced DNA damage suggest that SMARCAD1 SIM mediates interactions with a host of proteins that potentially recruit it to diverse DNA repair pathways, specifically at the early stages of nucleotide excision repair (NER) and base excision repair (BER). In contrast, CUE1 domain (specifically mediating a binary interaction with KAP1 as demonstrated in Chapter 5) appears to sustain additional protein interactions with various effector proteins involved in regulation of p53 transcriptional repression through MDM2/MDM4 ubiquitylation and subsequent degradation, including Nucleophosmin (NPM1), Guanine Nucleotide-binding Protein-like 3 (GNL3) and KAP1 itself. These interactions are selectively compromised when CUE1 is abrogated by mutagenesis, prior to chemically induced DNA damage. Consistent with this proposed role, upon genotoxic treatment SMARCAD1 SIM is shown to interact with deubiquitinase Ubiquitin Carboxyl-terminal Hydrolase 7 (USP7, an established SUMO2-modified modulator of

the p53-MDM2 axis) favouring transcriptional activation by de-ubiquitylating both p53 and MDM2. This, in turn, stabilises the p53 and impedes MDM2 ubiquitin E3-ligase activity. The differential protein profile identified by co-immunoprecipitation and mass spectrometry is illustrated in Figure 6.4. Protein targets have been clustered according to the specific signalling pathway they belong to, while some overlap or are known to be directly influenced by each other, reflecting the dynamic and transient cellular response to DNA damage (and thus, ATM/ATR activation).

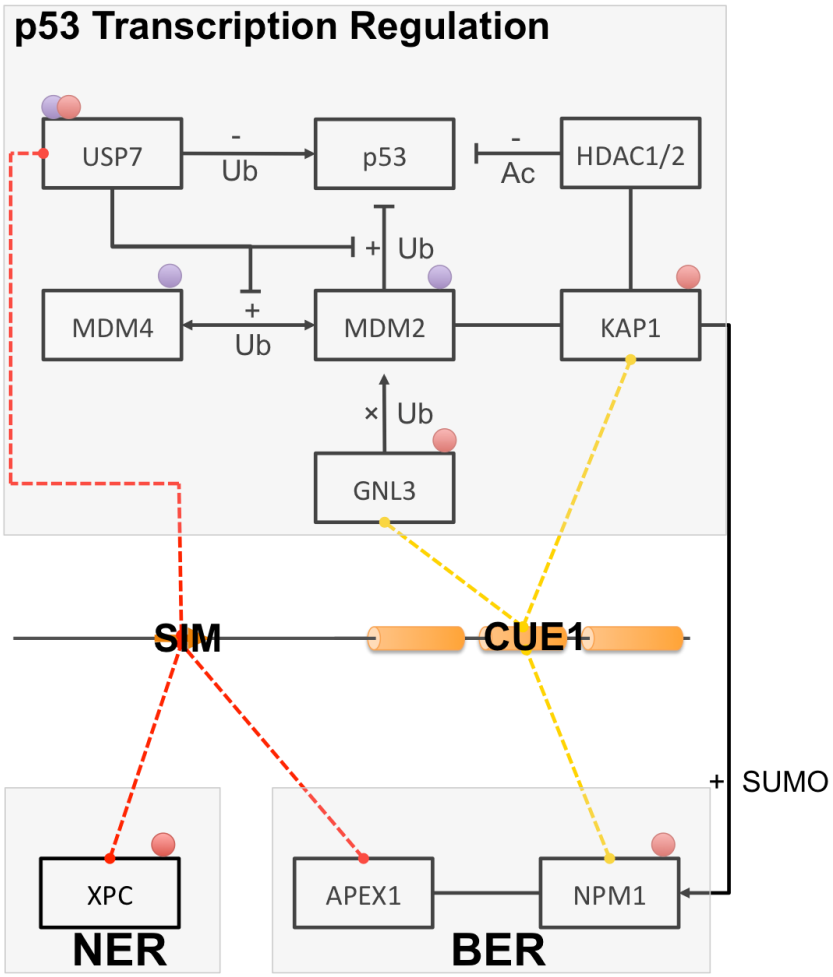



**Figure 6.4 Alternative pathways involved in SMARCAD1 response to chemically induced DNA damage.** Proteins are clustered based on the specific signalling cascade involved. Proteins experimentally detected are connected to either SIM or CUE1 domain. Targets detected prior to DNA damage are linked by yellow dashed line, those detected after DNA damage are connected in red. PTMs: SUMO as red spheres, Ubiquitylation as purple spheres.

The diversity of pathways associated with SMARCAD1 suggests there might be multiple functionally distinct subpopulations of it in the nucleus, simultaneously (and perhaps autonomously) enforcing its remodeling role.

Of the detected proteins, most are known to be SUMOylated including NPM1, GNL3, XPC and KAP1, while USP7 goes from being Ubiquitylated (and thus degraded) to being SUMOylated and stabilised in an ATM-dependent manner [233-235]. Due to low coverage, however, it was not possible to empirically confirm the predicted PTMs on any of the partner proteins with the exception of KAP1 (as discussed in Chapter 5), further highlighting the stoichiometry and high-affinity character of this interaction compared to all others proposed. The fact that the SMARCAD1 CUE1 domain did not present any DNA damage induced interactions might be a reflection of a biologically relevant process such as selective stripping of all alternative partners and specific recruitment to a given pathway, for instance. However, it seems more plausible to suppose technical limitations to the assay might be responsible for reduced sensitivity and thus an apparent stringently limited interactome, which could be woefully misleading.

The one irrefutable fact, the CUE1 mediated interaction with KAP1 RBCC domain, does inspire some potentially enlightening hypotheses, which could be further tested. First, at the molecular level, the 2:1 stoichiometric interaction suggests that, at the very least, CUE1 might be able to recognise motifs beyond ubiquitin. Moreover, our perturbation assay clearly indicates a much broader interface against KAP1 compared to the pointedly few residues significantly affected when involved in mono-ubiquitin binding. As may be inferred, this does not exclude the possibility of tighter interactions with poly-ubiquitin, as has been observed for other UBDs (notably, members of the structurally homologous UBA family). The possibility that such interactions might not be mutually exclusive with KAP1 binding would imagine a potential scenario in which ubiquitylated substrates for KAP1 E3-ligase Ring domain are immediately recognised (and perhaps, remodelled) by SMARCAD1. This, however, is quite a speculative extrapolation. Additionally, though SMARCAD1 CUE2 does not bind mono-ubiquitin under the conditions tested, it does not bind the KAP1

RBCC domain either, so it is still certainly possible that it could selectively recognise a poly-ubiquitin chain.

These data also address the role of SMARCAD1 in complex with KAP1 within the DDR. Beyond the established model for heterochromatin relaxation preceding HR repair of DSBs, evidence now suggests that KAP1 might be antagonistically associated with BRCA1 foci, the signalling stage that effectively commits the DSB repair to HR through resection of the broken DNA ends. Kuo *et al* [192, 195] show that KAP1 foci are depleted in a cell cycle dependent manner during the transition to S phase, and are inversely proportional to BRCA1 foci though both co-localise with 53BP1. This is explained as a consequence of S/G2 accumulation of RNF4, a SUMO-guided ubiquitin ligase that selectively targets SUMOylated KAP1 only once ATM has phosphorylated it at Ser824. As this modification is associated with initial local chromatin relaxation by disrupting the KAP1 SUMO-CHD3 SIM interaction, it would appear that KAP1 itself is also evicted from DSB flanking chromatin through RNF4 ubiquitylation and subsequent proteosomal degradation.

The key factor, however, is the fate of KAP1 bound SMARCAD1. Is the evicted KAP1 inexorably coupled to SMARCAD1? Is SMARCAD1 itself ubiquitylated by RNF4? Is KAP1 ubiquitylation mutually exclusive with SMARCAD1 interaction? As mentioned before, it is certainly true that SMARCAD1 has been firmly associated with DSB resection during HR, specifically following BRCA1-BARD1 ubiquitylation of H2A histone. Does this mean that SMARCAD1 must dissociate from KAP1 before it can recognise H2A-Ub, presumably a lower affinity ligand if the *in vitro* characterisation with mono-ubiquitin holds valid?

Alternatively, these events, though proposed and studied in isolation, might be snapshots of a concerted process, as presented in Figure 6.5. In this scenario, eviction of 53BP1 by SMARCAD1 through H2A-Ub remodeling is enhanced by RNF4 ubiquitylation of KAP1, which while bound to SMARCAD1 enforces translocation of the entire multi-subunit complex, paving the way for chromatin relaxation, DSB end remodeling and resection.

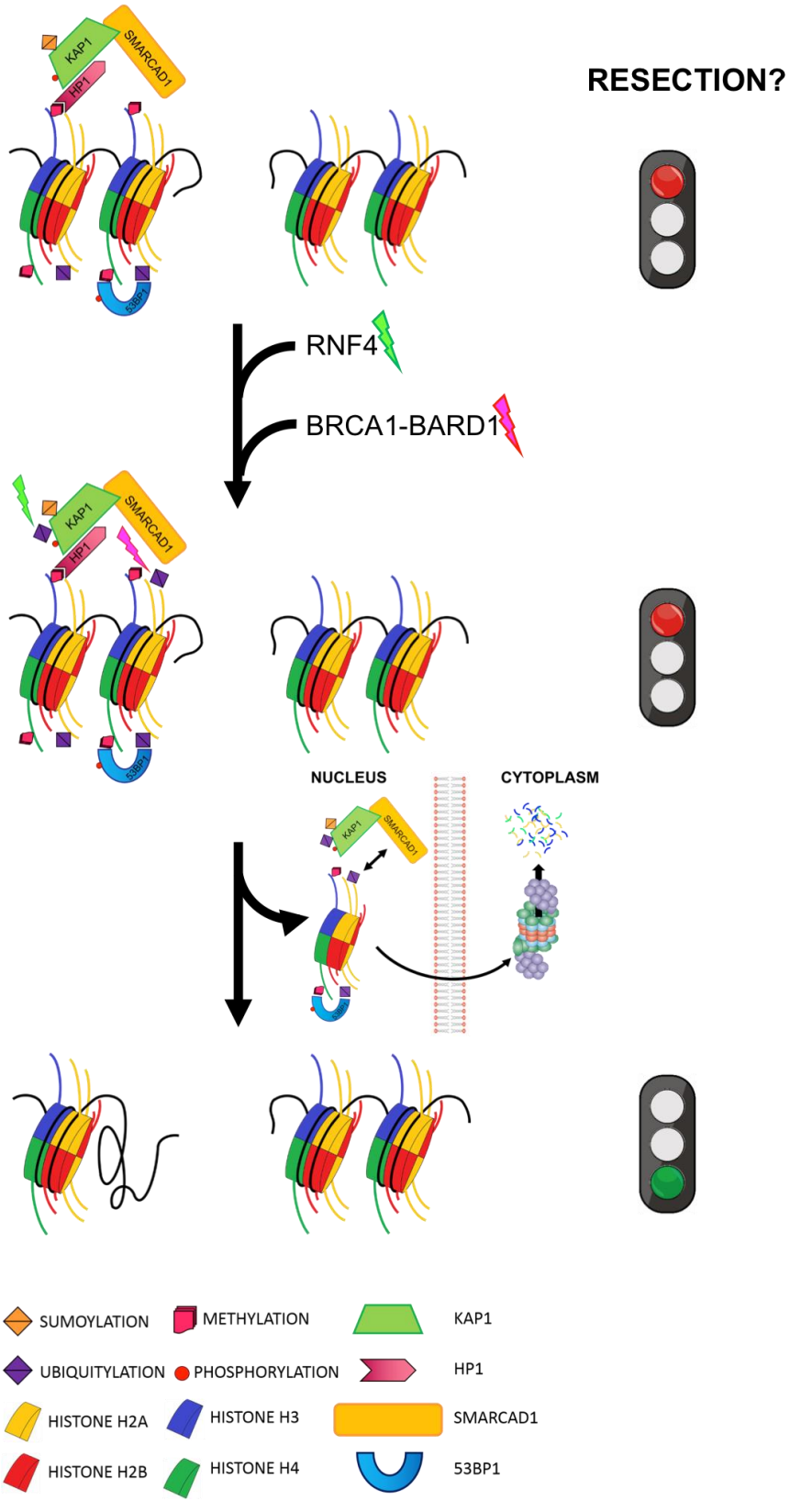



**Figure 6.5 SMARCAD1 bridges the concerted eviction of KAP1 and 53BP1 downstream of the ubiquitin-ligase activity of RNF4 and BRCA1-BARD1.**

Equally possible, and perhaps more likely given that it is a simpler model, there might be several subpopulations of SMARCAD1 which could be simultaneously recruited to diverging remodeling cascades, connected under an overarching goal, presented in Figure 6.6. In favour of this, and as previously mentioned, the differential MS profiling of domain dependent interactions indicates SMARCAD1 involvement with multiple cellular pathways influencing the DDR. Furthermore, the preliminary Proximity Ligation Assay (PLA) immunofluorescence staining results demonstrate that from the anti-SMARCAD1 antibody uniformly stained nucleus only a fraction is engaged with KAP1 and co-localises with heterochromatin (as detected from H3K9me3 co-staining). From this it may be inferred that additional SMARCAD1 complexes with alternate functions might be present in the cell, allowing for KAP1 eviction and H2A-Ub remodeling within the same pathway without requiring a grand unifying hypothesis. Further analysis of these and complementary aspects surrounding SMARCAD1 function in repair, replication and transcriptional regulation are required before a definitive answer may be reached.

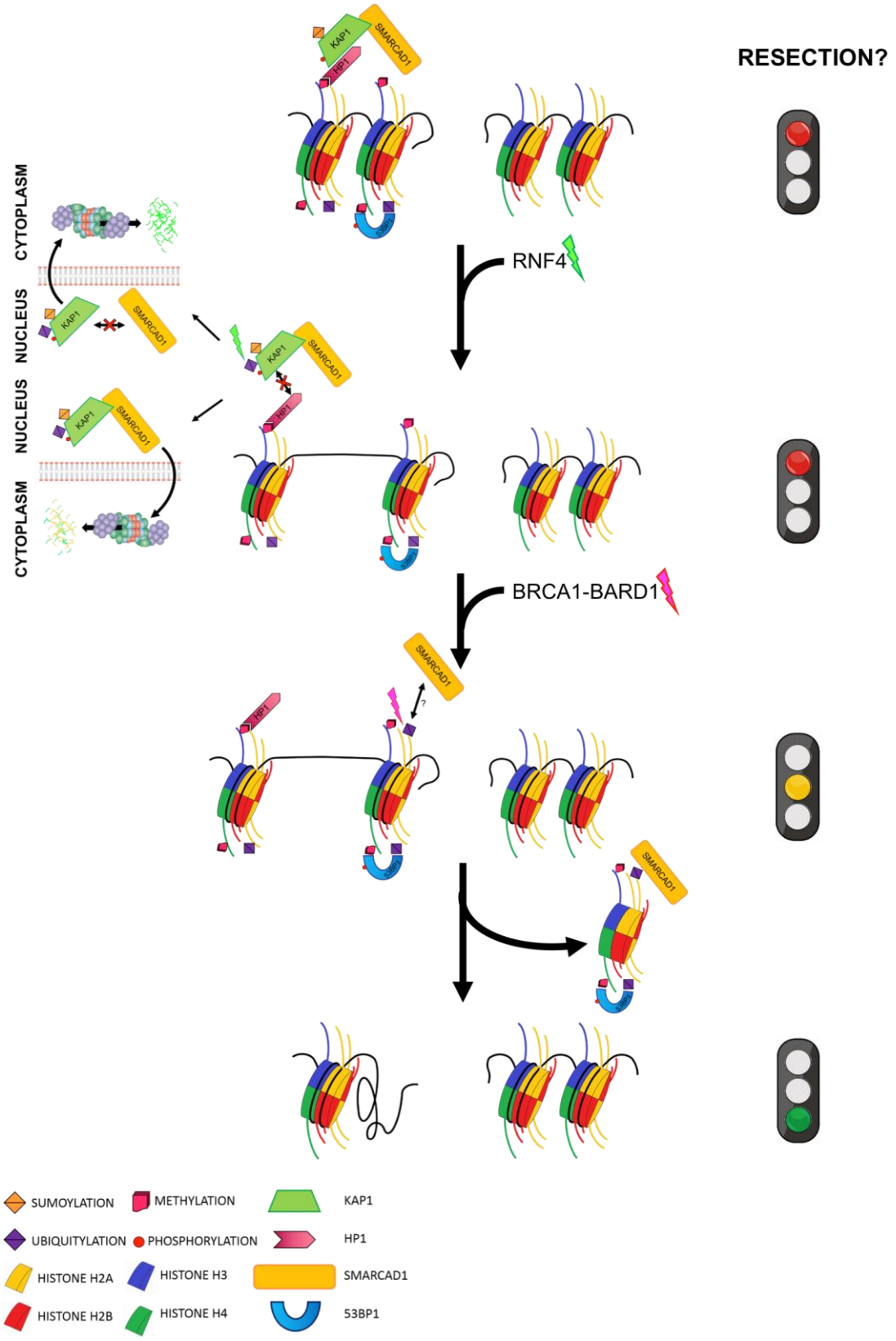



**Figure 6.6 SMARCAD1 nuclear subpopulations enforce autonomous functions associated with KAP1 interaction and 53BP1 chromatin eviction.**

# References


1. van Steensel, B., *Chromatin: constructing the big picture.* EMBO J, 2011. **30**(10): p. 1885-95.
2. Luger, K., M.L. Dechassa, and D.J. Tremethick, *New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 436-47.
3. Woodcock, C.L., *Chromatin architecture.* Curr Opin Struct Biol, 2006. **16**(2): p. 213-20.
4. Ball, P., *Portrait of a molecule.* Nature, 2003. **421**(6921): p. 421-2.
5. Belmont, A.S., *Large-scale chromatin organization: the good, the surprising, and the still perplexing.* Curr Opin Cell Biol, 2014. **26**: p. 69-78.
6. Felsenfeld, G. and M. Groudine, *Controlling the double helix.* Nature, 2003. **421**(6921): p. 448-53.
7. Woodcock, C.L. and R.P. Ghosh, *Chromatin higher-order structure and dynamics.* Cold Spring Harb Perspect Biol, 2010. **2**(5): p. a000596.
8. Politz, J.C., D. Scalzo, and M. Groudine, *Something silent this way forms: the functional organization of the repressive nuclear compartment.* Annu Rev Cell Dev Biol, 2013. **29**: p. 241-70.
9. Saksouk, N., E. Simboeck, and J. Dejardin, *Constitutive heterochromatin formation and transcription in mammals.* Epigenetics Chromatin, 2015. **8**: p. 3.
10. Chen, L. and J. Widom, *Mechanism of transcriptional silencing in yeast.* Cell, 2005. **120**(1): p. 37-48.
11. Ruthenburg, A.J., et al., *Multivalent engagement of chromatin modifications by linked binding modules.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 983-94.
12. Venkatesh, S. and J.L. Workman, *Histone exchange, chromatin structure and the regulation of transcription.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 178-89.
13. Glatt, S., C. Alfieri, and C.W. Muller, *Recognizing and remodeling the nucleosome.* Curr Opin Struct Biol, 2011. **21**(3): p. 335-41.
14. Cheema, M.S. and J. Ausio, *The Structural Determinants behind the Epigenetic Role of Histone Variants.* Genes (Basel), 2015. **6**(3): p. 685-713.
15. Soria, G., S.E. Polo, and G. Almouzni, *Prime, repair, restore: the active role of chromatin in the DNA damage response.* Mol Cell, 2012. **46**(6): p. 722-34.
16. Quenet, D. and Y. Dalal, *The CENP-A nucleosome: a dynamic structure and role at the centromere.* Chromosome Res, 2012. **20**(5): p. 465-79.
17. Sevilla, A. and O. Binda, *Post-translational modifications of the histone variant H2AZ.* Stem Cell Res, 2014. **12**(1): p. 289-95.
18. Turinetto, V. and C. Giachino, *Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions.* Nucleic Acids Res, 2015. **43**(5): p. 2489-98.
19. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

20. Zhou, V.W., A. Goren, and B.E. Bernstein, *Charting histone modifications and the functional organization of mammalian genomes.* Nat Rev Genet, 2011. **12**(1): p. 7-18.

21. Audia, J.E. and R.M. Campbell, *Histone Modifications and Cancer.* Cold Spring Harb Perspect Biol, 2016. **8**(4): p. a019521.

22. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.

23. Portela, A. and M. Esteller, *Epigenetic modifications and human disease.* Nat Biotechnol, 2010. **28**(10): p. 1057-68.

24. Waldmann, T. and R. Schneider, *Targeting histone modifications--epigenetics in cancer.* Curr Opin Cell Biol, 2013. **25**(2): p. 184-9.

25. Rando, O.J., *Combinatorial complexity in chromatin structure and function: revisiting the histone code.* Curr Opin Genet Dev, 2012. **22**(2): p. 148-55.

26. Feng, Y.L., et al., *Buried territories: heterochromatic response to DNA double-strand breaks.* Acta Biochim Biophys Sin (Shanghai), 2016. **48**(7): p. 594-602.

27. Huang, H., et al., *SnapShot: histone modifications.* Cell, 2014. **159**(2): p. 458-458 e1.

28. Smeenk, G. and H. van Attikum, *The chromatin response to DNA breaks: leaving a mark on genome integrity.* Annu Rev Biochem, 2013. **82**: p. 55-80.

29. Bartholomew, B., *Regulating the chromatin landscape: structural and mechanistic perspectives.* Annu Rev Biochem, 2014. **83**: p. 671-96.

30. Escargueil, A.E., et al., *What histone code for DNA repair?* Mutat Res, 2008. **658**(3): p. 259-70.

31. Clapier, C.R. and B.R. Cairns, *The biology of chromatin remodeling complexes.* Annu Rev Biochem, 2009. **78**: p. 273-304.

32. Seeber, A., M. Hauer, and S.M. Gasser, *Nucleosome remodelers in double-strand break repair.* Curr Opin Genet Dev, 2013. **23**(2): p. 174-84.

33. Peterson, C.L. and J. Cote, *Cellular machineries for chromosomal DNA repair.* Genes Dev, 2004. **18**(6): p. 602-16.

34. Dantuma, N.P. and H. van Attikum, *Spatiotemporal regulation of posttranslational modifications in the DNA damage response.* EMBO J, 2016. **35**(1): p. 6-23.

35. Groth, A., et al., *Regulation of replication fork progression through histone supply and demand.* Science, 2007. **318**(5858): p. 1928-31.

36. Bennett, G., M. Papamichos-Chronakis, and C.L. Peterson, *DNA repair choice defines a common pathway for recruitment of chromatin regulators.* Nat Commun, 2013. **4**: p. 2084.

37. Branzei, D. and M. Foiani, *Regulation of DNA repair throughout the cell cycle.* Nat Rev Mol Cell Biol, 2008. **9**(4): p. 297-308.

38. Goodarzi, A.A., P. Jeggo, and M. Lobrich, *The influence of heterochromatin on DNA double strand break repair: Getting the strong, silent type to relax.* DNA Repair (Amst), 2010. **9**(12): p. 1273-82.

39. Polo, S.E. and S.P. Jackson, *Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications.* Genes Dev, 2011. **25**(5): p. 409-33.

40. Gospodinov, A. and Z. Herceg, *Shaping chromatin for repair.* Mutat Res, 2013. **752**(1): p. 45-60.

41. Papamichos-Chronakis, M. and C.L. Peterson, *Chromatin and the genome integrity network.* Nat Rev Genet, 2013. **14**(1): p. 62-75.

42. Bolderson, E., et al., *Kruppel-associated Box (KRAB)-associated co-repressor (KAP-1) Ser-473 phosphorylation regulates heterochromatin protein 1beta (HP1-beta) mobilization and DNA repair in heterochromatin.* J Biol Chem, 2012. **287**(33): p. 28122-31.

43. Goodarzi, A.A., T. Kurka, and P.A. Jeggo, *KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response.* Nat Struct Mol Biol, 2011. **18**(7): p. 831-9.

44. Ziv, Y., et al., *Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway.* Nat Cell Biol, 2006. **8**(8): p. 870-6.

45. Ayoub, N., et al., *HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response.* Nature, 2008. **453**(7195): p. 682-6.

46. Luijsterburg, M.S. and H. van Attikum, *Chromatin and the DNA damage response: the cancer connection.* Mol Oncol, 2011. **5**(4): p. 349-67.

47. Smeenk, G. and N. Mailand, *Writers, Readers, and Erasers of Histone Ubiquitylation in DNA Double-Strand Break Repair.* Front Genet, 2016. **7**: p. 122.

48. Gong, L.W., E; Lin, S, *Chromatin Remodeling in DNA Damage Response and Human Aging*, in *Chromatin Remodelling*, D. Radzioch, Editor. 2013, InTech.

49. Iyama, T. and D.M. Wilson, 3rd, *DNA repair mechanisms in dividing and non-dividing cells.* DNA Repair (Amst), 2013. **12**(8): p. 620-36.

50. Mermoud, J.E., S.P. Rowbotham, and P.D. Varga-Weisz, *Keeping chromatin quiet: how nucleosome remodeling restores heterochromatin after replication.* Cell Cycle, 2011. **10**(23): p. 4017-25.

51. Rowbotham, S.P., et al., *Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1.* Mol Cell, 2011. **42**(3): p. 285-96.

52. Soininen, R., et al., *The mouse Enhancer trap locus 1 (Etl-1): a novel mammalian gene related to Drosophila and yeast transcriptional regulator genes.* Mech Dev, 1992. **39**(1-2): p. 111-23.

53. Schoor, M., et al., *Skeletal dysplasias, growth retardation, reduced postnatal survival, and impaired fertility in mice lacking the SNF2/SWI2 family member ETL1.* Mech Dev, 1999. **85**(1-2): p. 73-83.

54. Okazaki, N., et al., *The novel protein complex with SMARCAD1/KIAA1122 binds to the vicinity of TSS.* J Mol Biol, 2008. **382**(2): p. 257-65.

55. Nousbeck, J., et al., *A mutation in a skin-specific isoform of SMARCAD1 causes autosomal-dominant adermatoglyphia.* Am J Hum Genet, 2011. **89**(2): p. 302-7.

56. Marks, K.C., et al., *Analysis of two candidate genes for Basan syndrome.* Am J Med Genet A, 2014. **164A**(5): p. 1188-91.

57. Nousbeck, J., et al., *Mutations in SMARCAD1 cause autosomal dominant adermatoglyphia and perturb the expression of epidermal differentiation-associated genes.* Br J Dermatol, 2014. **171**(6): p. 1521-4.

58. Li, M., et al., *Genome-wide linkage analysis and whole-genome sequencing identify a recurrent SMARCAD1 variant in a unique Chinese family with Basan syndrome.* Eur J Hum Genet, 2016. **24**(9): p. 1367-70.

59. Chen, X., et al., *The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends.* Nature, 2012. **489**(7417): p. 576-80.

60. Eapen, V.V., et al., *The Saccharomyces cerevisiae chromatin remodeler Fun30 regulates DNA end resection and checkpoint deactivation.* Mol Cell Biol, 2012. **32**(22): p. 4727-40.

61. Costelloe, T., et al., *The yeast Fun30 and human SMARCAD1 chromatin remodellers promote DNA end resection.* Nature, 2012. **489**(7417): p. 581-4.

62. Densham, R.M., et al., *Human BRCA1-BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection.* Nat Struct Mol Biol, 2016. **23**(7): p. 647-55.

63. Brzovic, P.S., et al., *Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex.* Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5646-51.

64. Panier, S. and S.J. Boulton, *Double-strand break repair: 53BP1 comes into focus.* Nat Rev Mol Cell Biol, 2014. **15**(1): p. 7-18.

65. Panier, S. and D. Durocher, *Push back to respond better: regulatory inhibition of the DNA double-strand break response.* Nat Rev Mol Cell Biol, 2013. **14**(10): p. 661-72.

66. Matsuoka, S., et al., *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.* Science, 2007. **316**(5828): p. 1160-6.

67. Cremona, C.A., et al., *Extensive DNA damage-induced sumoylation contributes to replication and repair and acts in addition to the mec1 checkpoint.* Mol Cell, 2012. **45**(3): p. 422-32.

68. Elia, A.E., et al., *RFWD3-Dependent Ubiquitination of RPA Regulates Repair at Stalled Replication Forks.* Mol Cell, 2015. **60**(2): p. 280-93.

69. Psakhye, I. and S. Jentsch, *Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair.* Cell, 2012. **151**(4): p. 807-20.

70. Harding, S.M. and R.A. Greenberg, *Choreographing the Double Strand Break Response: Ubiquitin and SUMO Control of Nuclear Architecture.* Front Genet, 2016. **7**: p. 103.

71. Garcia-Rodriguez, N., R.P. Wong, and H.D. Ulrich, *Functions of Ubiquitin and SUMO in DNA Replication and Replication Stress.* Front Genet, 2016. **7**: p. 87.

72. Perry, J.J., J.A. Tainer, and M.N. Boddy, *A SIM-ultaneous role for SUMO and ubiquitin.* Trends Biochem Sci, 2008. **33**(5): p. 201-8.

73. van Wijk, S.J., S. Muller, and I. Dikic, *Shared and unique properties of ubiquitin and SUMO interaction networks in DNA repair.* Genes Dev, 2011. **25**(17): p. 1763-9.

74. Kerscher, O., *SUMO junction-what's your function? New insights through SUMO-interacting motifs.* EMBO Rep, 2007. **8**(6): p. 550-5.

75. Himmels, S.F. and A.A. Sartori, *Controlling DNA-End Resection: An Emerging Task for Ubiquitin and SUMO.* Front Genet, 2016. **7**: p. 152.

76. Jackson, S.P. and D. Durocher, *Regulation of DNA damage responses by ubiquitin and SUMO.* Mol Cell, 2013. **49**(5): p. 795-807.

77. Hay, R.T., *Decoding the SUMO signal.* Biochem Soc Trans, 2013. **41**(2): p. 463-73.

78. Ronau, J.A., J.F. Beckmann, and M. Hochstrasser, *Substrate specificity of the ubiquitin and Ubl proteases.* Cell Res, 2016. **26**(4): p. 441-56.

79. Nijman, S.M., et al., *A genomic and functional inventory of deubiquitinating enzymes.* Cell, 2005. **123**(5): p. 773-86.

80. Pinto-Fernandez, A. and B.M. Kessler, *DUBbing Cancer: Deubiquitylating Enzymes Involved in Epigenetics, DNA Damage and the Cell Cycle As Therapeutic Targets.* Front Genet, 2016. **7**: p. 133.

81. Geiss-Friedlander, R. and F. Melchior, *Concepts in sumoylation: a decade on.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 947-56.

82. Garza, R. and L. Pillus, *STUbLs in chromatin and genome stability.* Biopolymers, 2013. **99**(2): p. 146-54.

83. Sriramachandran, A.M. and R.J. Dohmen, *SUMO-targeted ubiquitin ligases.* Biochim Biophys Acta, 2014. **1843**(1): p. 75-85.

84. Sobhian, B., et al., *RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites.* Science, 2007. **316**(5828): p. 1198-202.

85. Nikkila, J., et al., *Familial breast cancer screening reveals an alteration in the RAP80 UIM domain that impairs DNA damage response function.* Oncogene, 2009. **28**(16): p. 1843-52.

86. Solyom, S., et al., *Breast cancer-associated Abraxas mutation disrupts nuclear localization and DNA damage response functions.* Sci Transl Med, 2012. **4**(122): p. 122ra23.

87. Raman, N., A. Nayak, and S. Muller, *The SUMO system: a master organizer of nuclear protein assemblies.* Chromosoma, 2013. **122**(6): p. 475-85.

88. Thorslund, T., et al., *Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage.* Nature, 2015. **527**(7578): p. 389-93.

89. Pinder, J.B., K.M. Attwood, and G. Dellaire, *Reading, writing, and repair: the role of ubiquitin and the ubiquitin-like proteins in DNA damage signaling and repair.* Front Genet, 2013. **4**: p. 45.

90. Anamika and L. Spyracopoulos, *Molecular Basis for Phosphorylation-dependent SUMO Recognition by the DNA Repair Protein RAP80.* J Biol Chem, 2016. **291**(9): p. 4417-28.

91. Guzzo, C.M., et al., *RNF4-dependent hybrid SUMO-ubiquitin chains are signals for RAP80 and thereby mediate the recruitment of BRCA1 to sites of DNA damage.* Sci Signal, 2012. **5**(253): p. ra88.

92. Dikic, I., S. Wakatsuki, and K.J. Walters, *Ubiquitin-binding domains - from structures to functions.* Nat Rev Mol Cell Biol, 2009. **10**(10): p. 659-71.

93. Hicke, L., H.L. Schubert, and C.P. Hill, *Ubiquitin-binding domains.* Nat Rev Mol Cell Biol, 2005. **6**(8): p. 610-21.

94. Hurley, J.H., S. Lee, and G. Prag, *Ubiquitin-binding domains.* Biochem J, 2006. **399**(3): p. 361-72.

95. Kang, R.S., et al., *Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding.* Cell, 2003. **113**(5): p. 621-30.

96. Shih, S.C., et al., *A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain.* EMBO J, 2003. **22**(6): p. 1273-81.

97. Bergink, S. and S. Jentsch, *Principles of ubiquitin and SUMO modifications in DNA repair.* Nature, 2009. **458**(7237): p. 461-7.

98. Dantuma, N.P., C. Heinen, and D. Hoogstraten, *The ubiquitin receptor Rad23: at the crossroads of nucleotide excision repair and proteasomal degradation.* DNA Repair (Amst), 2009. **8**(4): p. 449-60.

99. Grabbe, C. and I. Dikic, *Functional roles of ubiquitin-like domain (ULD) and ubiquitin-binding domain (UBD) containing proteins.* Chem Rev, 2009. **109**(4): p. 1481-94.

100. Hecker, C.M., et al., *Specification of SUMO1- and SUMO2-interacting motifs.* J Biol Chem, 2006. **281**(23): p. 16117-27.

101. Cappadocia, L., et al., *Structural and functional characterization of the phosphorylation-dependent interaction between PML and SUMO1.* Structure, 2015. **23**(1): p. 126-38.

102. Xu, Y., et al., *Structural insight into SUMO chain recognition and manipulation by the ubiquitin ligase RNF4.* Nat Commun, 2014. **5**: p. 4217.

103. Chang, C.C., et al., *Structural and functional roles of Daxx SIM phosphorylation in SUMO paralog-selective binding and apoptosis modulation.* Mol Cell, 2011. **42**(1): p. 62-74.

104. Escobar-Cabrera, E., et al., *Characterizing the N- and C-terminal Small ubiquitin-like modifier (SUMO)-interacting motifs of the scaffold protein DAXX.* J Biol Chem, 2011. **286**(22): p. 19816-29.

105. Sambrook, J. and D.W. Russell, *Molecular cloning : a laboratory manual*. 3rd ed. 2001, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

106. Berrow, N.S., et al., *A versatile ligation-independent cloning method suitable for high-throughput expression screening applications.* Nucleic Acids Res, 2007. **35**(6): p. e45.

107. Oxford Protein Production Facility. *Opiner - Infusion cloning designer*. 2016; Available from: https://www.oppf.rc-harwell.ac.uk/Opiner/.

108. Yachdav, G., et al., *PredictProtein--an open resource for online prediction of protein structural and functional features.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W337-43.

109. Mooij, W.T., E. Mitsiki, and A. Perrakis, *ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W402-5.

110. Yang, Z.R., et al., *RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins.* Bioinformatics, 2005. **21**(16): p. 3369-76.

111. Agilent Technologies. *QuikChange Primer Design*. 2016 [cited 2013 July]; Available from: http://www.agilent.com/genomics/qcpd.

112. Alexy Novoradovsky, V.Z., Madhushree Ghosh, Holly Hogrefe, Joseph A Sorge, Terry Gasterland. *Computational Principles of Primer Design for Site Directed Mutagenesis*. in *2005 NSTI Nanotechnology Conference and Trade Show*. 2005. Anaheim, USA.

113. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-34.

114. Yates, L.A. and R.J. Gilbert, *Efficient production and purification of recombinant murine kindlin-3 from insect cells for biophysical studies.* J Vis Exp, 2014(85).

115. Zhao, Y., D.A. Chapman, and I.M. Jones, *Improving baculovirus recombination.* Nucleic Acids Res, 2003. **31**(2): p. E6-6.

116. Nettleship, J.E., et al., *Recent advances in the production of proteins in insect and mammalian cells for structural biology.* J Struct Biol, 2010. **172**(1): p. 55-65.

117. Aricescu, A.R., et al., *Eukaryotic expression: developments for structural proteomics.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 10): p. 1114-24.

118. Pickart, C.M. and S. Raasi, *Controlled synthesis of polyubiquitin chains.* Methods Enzymol, 2005. **399**: p. 21-36.

119. Wessel, D. and U.I. Flugge, *A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids.* Anal Biochem, 1984. **138**(1): p. 141-3.

120. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.

121. Vranken, W.F., et al., *The CCPN data model for NMR spectroscopy: development of a software pipeline.* Proteins, 2005. **59**(4): p. 687-96.

122. Ottiger, M., F. Delaglio, and A. Bax, *Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra.* J Magn Reson, 1998. **131**(2): p. 373-8.

123. Otting, G., et al., *NMR experiments for the sign determination of homonuclear scalar and residual dipolar couplings.* J Biomol NMR, 2000. **16**(4): p. 343-6.

124. Rieping, W., et al., *ARIA2: automated NOE assignment and data integration in NMR structure calculation.* Bioinformatics, 2007. **23**(3): p. 381-2.

125. Bardiaux, B., *Structure calculation of proteins from solution and solid-state NMR data: application to monomers and symmetric aggregates*. 2009. p. VIII, 129 S.

126. Nilges, M., *Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities.* J Mol Biol, 1995. **245**(5): p. 645-60.

127. Wider, G., *Structure determination of biological macromolecules in solution using nuclear magnetic resonance spectroscopy.* Biotechniques, 2000. **29**(6): p. 1278-82, 1284-90, 1292 passim.

128. Walter, T.S., et al., *A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 6): p. 651-7.

129. Mayo, C.J., et al., *Benefits of automated crystallization plate tracking, imaging, and analysis.* Structure, 2005. **13**(2): p. 175-82.

130. Walter, T.S., et al., *Semi-automated microseeding of nanolitre crystallization experiments.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2008. **64**(Pt 1): p. 14-8.

131. Kaiser, P. and C. Tagwerker, *Is this protein ubiquitinated?* Methods Enzymol, 2005. **399**: p. 243-8.

132. Maine, G.N., et al., *A bimolecular affinity purification method under denaturing conditions for rapid isolation of a ubiquitinated protein for mass spectrometry analysis.* Nat Protoc, 2010. **5**(8): p. 1447-59.

133. Zimm, B.H., *The dependence of the scattering of light on angle and concentration in linear polymer solutions.* J Phys Colloid Chem, 1948. **52**(1): p. 260-7.

134. Thymiakou, E. and V. Episkopou, *Detection of signaling effector-complexes downstream of bmp4 using PLA, a proximity ligation assay.* J Vis Exp, 2011(49).

135. West, P.M., University of Oxford. Medical Sciences Division., and University of Oxford, *The double CUE domain of chromatin remodelling factor SMARCAD1*. 2012. xvi, 269 leaves.

136. Neves-Costa, A., et al., *The SNF2-family member Fun30 promotes gene silencing in heterochromatic loci.* PLoS One, 2009. **4**(12): p. e8111.

137. Sigrist, C.J., et al., *New and continuing developments at PROSITE.* Nucleic Acids Res, 2013. **41**(Database issue): p. D344-7.

138. Winget, J.M. and T. Mayor, *The diversity of ubiquitin recognition: hot spots and varied specificity.* Mol Cell, 2010. **38**(5): p. 627-35.

139. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.

140. Redfield, C., *Assignment of Protein NMR Spectra Using Heteronuclear NMR – A Tutorial*, in *Protein NMR, Modern Techniques and Biomedical Applications*, L. Berliner, Editor. 2015, Springer US.

141. Higman, V.A. *Protein NMR - A Practical Guide*. 2012 4 January 2016 [cited 2016 October]; Available from: http://www.protein-nmr.org.uk.

142. Marion, D., et al., *Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta.* Biochemistry, 1989. **28**(15): p. 6150-6.

143. Cavanagh, J., *Protein NMR spectroscopy : principles and practice*. 2nd ed. 2007, Amsterdam ; Boston: Academic Press. xxv, 885 p.

144. Grzesiek, S.B., A., *Improved 3D Triple-Resonance NMR Techniques Applied to a 31 kDa Protein.* Journal of Magnetic Resonance, 1992. **96**(2): p. 432-440.

145. Grzesiek, S.I., M.; Clore, G. M.; Gronenborn, A. M.; Bax, A., *A 3D triple-resonance NMR technique for qualitative measurement of carbonyl-Hβ J couplings in isotopically enriched proteins.* Journal of Magnetic Resonance, 1992. **96**(1): p. 215-221.

146. Kay, L.E., et al., *Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution.* Science, 1990. **249**(4967): p. 411-4.

147. Clubb, R.T.T., V.; Wagner, G., *A Constant-Time Three-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue 1HN, 15N, and 13C' Chemical Shifts in 15N- 13C-Labeled Proteins.* Journal of Magnetic Resonance, 1992. **97**(1): p. 213-217.

148. Kleckner, I.R. and M.P. Foster, *An introduction to NMR-based approaches for measuring protein dynamics.* Biochim Biophys Acta, 2011. **1814**(8): p. 942-68.

149. Markwick, P.R., T. Malliavin, and M. Nilges, *Structural biology by NMR: structure, dynamics, and interactions.* PLoS Comput Biol, 2008. **4**(9): p. e1000168.

150. Renner, C., et al., *Practical aspects of the 2D 15N-[1h]-NOE experiment.* J Biomol NMR, 2002. **23**(1): p. 23-33.

151. Shen, Y., et al., *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* J Biomol NMR, 2009. **44**(4): p. 213-23.

152. Eberstadt, M.G., G.; Mierke, D. F.; Kessler, H., *Scalar Coupling Constants—Their Analysis and Their Application for the Elucidation of Structures.* Angewandte Chemie, 1995. **34**(16): p. 1671-1695.

153. Wang, A.C.B., A., *Determination of the Backbone Dihedral Angles φ in Human Ubiquitin from Reparametrized Empirical Karplus Equations.* Journal of the American Chemical Society, 1996. **118**(10): p. 2483-2494.

154. Vuister, G.W.B., A., *Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins.* Journal of the American Chemical Society, 1993. **115**(17): p. 7772-7777.

155. Vuister, G.W.T., M.; Karimi-Nejad, Y.; Whitehead, B., *Pulse Sequences for Measuring Coupling Constants*, in *Modern Techniques in Protein NMR*, N.R.B. Krishna, L. J., Editor. 1999, Springer US. p. 195-257.

156. Williamson, M.P., *Applications of the NOE in Molecular Biology*, in *Annual Reports on NMR Spectroscopy*, G. Webb, Editor. 2009, Academic Press. p. 77-109.

157. Jensen, P.R., et al., *Improvement of hydrogen bond geometry in protein NMR structures by residual dipolar couplings--an assessment of the interrelation of NMR restraints.* J Biomol NMR, 2004. **28**(1): p. 31-41.

158. Mareuil, F., et al., *Improved reliability, accuracy and quality in automated NMR structure calculation with ARIA.* J Biomol NMR, 2015. **62**(4): p. 425-38.

159. Nilges, M., et al., *Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin.* J Mol Biol, 1997. **269**(3): p. 408-22.

160. Bax, A., G. Kontaxis, and N. Tjandra, *Dipolar couplings in macromolecular structure determination.* Methods Enzymol, 2001. **339**: p. 127-74.

161. Prestegard, J.H., H.M. al-Hashimi, and J.R. Tolman, *NMR structures of biomolecules using field oriented media and residual dipolar couplings.* Q Rev Biophys, 2000. **33**(4): p. 371-424.

162. Tjandra, N. and A. Bax, *Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium.* Science, 1997. **278**(5340): p. 1111-4.

163. Sørensen, M.D.M., A.; Sørensen, O. W., *Spin-state-selective coherence transfer via intermediate states of two-spin coherence in IS spin systems: Application to E.COSY-type measurement of J coupling constants.* Journal of Biomolecular NMR, 1997. **10**(2): p. 181-186.

164. Rückert, M.O., G., *Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments.* Journal of the American Chemical Society, 2000. **122**(32): p. 7793-7797.

165. Guerry, P. and T. Herrmann, *Advances in automated NMR protein structure determination.* Q Rev Biophys, 2011. **44**(3): p. 257-309.

166. Linge, J.P., et al., *ARIA: automated NOE assignment and NMR structure calculation.* Bioinformatics, 2003. **19**(2): p. 315-6.

167. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.

168. Fossi, M., et al., *Quantitative study of the effects of chemical shift tolerances and rates of SA cooling on structure calculation from automatically assigned NOE data.* J Magn Reson, 2005. **175**(1): p. 92-102.

169. Nilges, M.O.D., S. I., *Ambigous NOEs and Automated NOE Assingment.* Prog Nucl Magn Reson Spectrosc, 1998. **32**: p. 107-139.

170. Linge, J.P., et al., *Refinement of protein structures in explicit solvent.* Proteins, 2003. **50**(3): p. 496-506.

171. Vranken, W.F.V., G. W.; Bonvin, A. M. J. J. , *NMR-Based Modeling and Refinement of Protein 3D Structures*, in *Molecular Modeling of Proteins*, A. Kukol, Editor. 2014. p. 351-380.

172. Doreleijers, J.F., et al., *CING: an integrated residue-based structure validation program suite.* J Biomol NMR, 2012. **54**(3): p. 267-83.

173. Vuister, G.W., et al., *An overview of tools for the validation of protein NMR structures.* J Biomol NMR, 2014. **58**(4): p. 259-85.

174. Montelione, G.T., et al., *Recommendations of the wwPDB NMR Validation Task Force.* Structure, 2013. **21**(9): p. 1563-70.

175. Flotho, A. and F. Melchior, *Sumoylation: a regulatory protein modification in health and disease.* Annu Rev Biochem, 2013. **82**: p. 357-85.

176. Stehmeier, P. and S. Muller, *Phospho-regulated SUMO interaction modules connect the SUMO system to CK2 signaling.* Mol Cell, 2009. **33**(3): p. 400-9.

177. Sung, K.S., et al., *Role of the SUMO-interacting motif in HIPK2 targeting to the PML nuclear bodies and regulation of p53.* Exp Cell Res, 2011. **317**(7): p. 1060-70.

178. Ullmann, R., et al., *An acetylation switch regulates SUMO-dependent protein interaction networks.* Mol Cell, 2012. **46**(6): p. 759-70.

179. Bakkenist, C.J. and M.B. Kastan, *Initiating cellular stress responses.* Cell, 2004. **118**(1): p. 9-17.

180. Traven, A. and J. Heierhorst, *SQ/TQ cluster domains: concentrated ATM/ATR kinase phosphorylation site regions in DNA-damage-response proteins.* Bioessays, 2005. **27**(4): p. 397-407.

181. St-Denis, N., et al., *Systematic investigation of hierarchical phosphorylation by protein kinase CK2.* J Proteomics, 2015. **118**: p. 49-62.

182. Namanja, A.T., et al., *Insights into high affinity small ubiquitin-like modifier (SUMO) recognition by SUMO-interacting motifs (SIMs) revealed by a combination of NMR and peptide array analysis.* J Biol Chem, 2012. **287**(5): p. 3231-40.

183. Steen, H., et al., *Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements.* Mol Cell Proteomics, 2006. **5**(1): p. 172-81.

184. Dissmeyer, N. and A. Schnittger, *Use of phospho-site substitutions to analyze the biological relevance of phosphorylation events in regulatory networks.* Methods Mol Biol, 2011. **779**: p. 93-138.

185. Li, X., et al., *SUMOylation of the transcriptional co-repressor KAP1 is regulated by the serine and threonine phosphatase PP1.* Sci Signal, 2010. **3**(119): p. ra32.

186. Cheng, C.T., C.Y. Kuo, and D.K. Ann, *KAPtain in charge of multiple missions: Emerging roles of KAP1.* World J Biol Chem, 2014. **5**(3): p. 308-20.

187. Addison, J.B., et al., *KAP1 promotes proliferation and metastatic progression of breast cancer cells.* Cancer Res, 2015. **75**(2): p. 344-55.

188. Bunch, H. and S.K. Calderwood, *TRIM28 as a novel transcriptional elongation factor.* BMC Mol Biol, 2015. **16**: p. 14.

189. Peng, H., et al., *Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain.* J Biol Chem, 2000. **275**(24): p. 18000-10.

190. Iyengar, S., et al., *Functional analysis of KAP1 genomic recruitment.* Mol Cell Biol, 2011. **31**(9): p. 1833-47.

191. Bunch, H., et al., *Transcriptional elongation requires DNA break-induced signalling.* Nat Commun, 2015. **6**: p. 10191.

192. Kuo, C.Y., et al., *RNF4 regulates DNA double-strand break repair in a cell cycle-dependent manner.* Cell Cycle, 2016. **15**(6): p. 787-98.

193. Zeng, L., et al., *Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing.* Nat Struct Mol Biol, 2008. **15**(6): p. 626-33.

194. Ivanov, A.V., et al., *PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing.* Mol Cell, 2007. **28**(5): p. 823-37.

195. Kuo, C.Y., et al., *An arginine-rich motif of ring finger protein 4 (RNF4) oversees the recruitment and degradation of the phosphorylated and SUMOylated Kruppel-associated box domain-associated protein 1 (KAP1)/TRIM28 protein during genotoxic stress.* J Biol Chem, 2014. **289**(30): p. 20757-72.

196. Redfield, C., *NMR studies of partially folded molten-globule states.* Methods Mol Biol, 2004. **278**: p. 233-54.

197. Vassilenko, K.S. and V.N. Uversky, *Native-like secondary structure of molten globules.* Biochim Biophys Acta, 2002. **1594**(1): p. 168-77.

198. Kjaergaard, M., K. Teilum, and F.M. Poulsen, *Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP.* Proc Natl Acad Sci U S A, 2010. **107**(28): p. 12535-40.

199. Kalderon, D., et al., *A short amino acid sequence able to specify nuclear location.* Cell, 1984. **39**(3 Pt 2): p. 499-509.

200. Peng, H., et al., *Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions.* J Mol Biol, 2000. **295**(5): p. 1139-62.

201. Wang, H.K., S.; Murayama, K.; Takemoto, C.; Terada, T. Shirouzu, M; RIKEN Structural Genomics/Proteomics Initiative (RSGI), ;, *PDB: 2YVR, Crystal Structure of MS1043*. 2008.

202. Kumar, P. and C. Wolberger, *Structure of the yeast Bre1 RING domain.* Proteins, 2015. **83**(6): p. 1185-90.

203. Tao, H., et al., *Structure of the MID1 tandem B-boxes reveals an interaction reminiscent of intermolecular ring heterodimers.* Biochemistry, 2008. **47**(8): p. 2450-7.

204. Sanchez, J.G., et al., *The tripartite motif coiled-coil is an elongated antiparallel hairpin dimer.* Proc Natl Acad Sci U S A, 2014. **111**(7): p. 2494-9.

205. Peng, H., I. Feldman, and F.J. Rauscher, 3rd, *Hetero-oligomerization among the TIF family of RBCC/TRIM domain-containing nuclear cofactors: a potential mechanism for regulating the switch between coactivation and corepression.* J Mol Biol, 2002. **320**(3): p. 629-44.

206. Herquel, B., et al., *Transcription cofactors TRIM24, TRIM28, and TRIM33 associate to form regulatory complexes that suppress murine hepatocellular carcinoma.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8212-7.

207. Napolitano, L.M. and G. Meroni, *TRIM family: Pleiotropy and diversification through homomultimer and heteromultimer formation.* IUBMB Life, 2012. **64**(1): p. 64-71.

208. Koliopoulos, M.G., et al., *Functional role of TRIM E3 ligase oligomerization and regulation of catalytic activity.* EMBO J, 2016. **35**(11): p. 1204-18.

209. Hong, P., S. Koza, and E.S. Bouvier, *Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates.* J Liq Chromatogr Relat Technol, 2012. **35**(20): p. 2923-2950.

210. Oliva, A., M. Llabres, and J.B. Farina, *Comparative study of protein molecular weights by size-exclusion chromatography and laser-light scattering.* J Pharm Biomed Anal, 2001. **25**(5-6): p. 833-41.

211. Philo, J.S., *A critical review of methods for size characterization of non-particulate protein aggregates.* Curr Pharm Biotechnol, 2009. **10**(4): p. 359-72.

212. Goldstone, D.C., et al., *Structural studies of postentry restriction factors reveal antiparallel dimers that enable avid binding to the HIV-1 capsid lattice.* Proc Natl Acad Sci U S A, 2014. **111**(26): p. 9609-14.

213. Li, Y., et al., *Structural insights into the TRIM family of ubiquitin E3 ligases.* Cell Res, 2014. **24**(6): p. 762-5.

214. Keown, J.R., et al., *Characterisation of assembly and ubiquitylation by the RBCC motif of Trim5alpha.* Sci Rep, 2016. **6**: p. 26837.

215. Wagner, J.M., et al., *Mechanism of B-box 2 domain-mediated higher-order assembly of the retroviral restriction factor TRIM5alpha.* Elife, 2016. **5**.

216. Keown, J.R. and D.C. Goldstone, *Crystal structure of the Trim5alpha Bbox2 domain from rhesus macaques describes a plastic oligomerisation interface.* J Struct Biol, 2016. **195**(3): p. 282-5.

217. Sardiello, M., et al., *Genomic analysis of the TRIM family reveals two groups of genes with distinct evolutionary properties.* BMC Evol Biol, 2008. **8**: p. 225.

218. Li, Y.L., et al., *Primate TRIM5 proteins form hexagonal nets on HIV-1 capsids.* Elife, 2016. **5**.

219. Cole, J.L., et al., *Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium.* Methods Cell Biol, 2008. **84**: p. 143-79.

220. Freyer, M.W. and E.A. Lewis, *Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions.* Methods Cell Biol, 2008. **84**: p. 79-113.

221. Wang, C., et al., *MDM2 interaction with nuclear corepressor KAP1 contributes to p53 inactivation.* EMBO J, 2005. **24**(18): p. 3279-90.

222. Doyle, J.M., et al., *MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases.* Mol Cell, 2010. **39**(6): p. 963-74.

223. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Prog Nucl Magn Reson Spectrosc, 2013. **73**: p. 1-16.

224. Kantardjieff, K.A. and B. Rupp, *Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals.* Protein Sci, 2003. **12**(9): p. 1865-71.

225. Hennig, J., et al., *Structural organization and Zn2+-dependent subdomain interactions involving autoantigenic epitopes in the Ring-B-box-coiled-coil (RBCC) region of Ro52.* J Biol Chem, 2005. **280**(39): p. 33250-61.

226. Tu, W.Z., et al., *gammaH2AX foci formation in the absence of DNA damage: mitotic H2AX phosphorylation is mediated by the DNA-PKcs/CHK2 pathway.* FEBS Lett, 2013. **587**(21): p. 3437-43.

227. Soderberg, O., et al., *Direct observation of individual endogenous protein complexes in situ by proximity ligation.* Nat Methods, 2006. **3**(12): p. 995-1000.

228. Soderberg, O., et al., *Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay.* Methods, 2008. **45**(3): p. 227-32.

229. Al Kubaisy, E., et al., *SMARCAD1 knockdown uncovers its role in breast cancer cell migration, invasion, and metastasis.* Expert Opin Ther Targets, 2016. **20**(9): p. 1035-43.

230. Schwertman, P., S. Bekker-Jensen, and N. Mailand, *Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers.* Nat Rev Mol Cell Biol, 2016. **17**(6): p. 379-94.

231. Vissers, J.H., et al., *The many faces of ubiquitinated histone H2A: insights from the DUBs.* Cell Div, 2008. **3**: p. 8.

232. Luban, J., *Closing the net on retroviruses.* Elife, 2016. **5**.

233. Hendriks, I.A., et al., *System-wide identification of wild-type SUMO-2 conjugation sites.* Nat Commun, 2015. **6**: p. 7289.

234. Hendriks, I.A., et al., *Uncovering global SUMOylation signaling networks in a site-specific manner.* Nat Struct Mol Biol, 2014. **21**(10): p. 927-36.

235. Impens, F., et al., *Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli.* Proc Natl Acad Sci U S A, 2014. **111**(34): p. 12432-7.

# Statement of Authorship

This thesis is approximately 36,000 words long, excluding of references, figures and tables. The research presented here and the thesis itself are my own work, except where explicitly stated.

Mass spectrometry sample digestion, processing and data collection were carried out by the Kessler group at the Target Discovery Institute, University of Oxford.

Multi-angle laser light scattering (MALLS) analysis was carried out by Dr David Staunton at the Biochemistry Department, University of Oxford.

Many NMR experiments were carried out with assistance of Prof Christina Redfield, who helped set up experiments and process the raw data.

Crystallisation trials were performed with guidance from Dr Tom Walter at the Division of Structural Biology, University of Oxford.

X-ray crystallography sample collection, crystal freezing and diffraction data acquisition was performed under close supervision of Dr. Erika Mancini.

Antonio J Biasutto

October 2016