The Double CUE Domain of Chromatin Remodelling Factor SMARCAD1



Philip Michael West

Nuffield Department of Clinical Medicine

Linacre College, University of Oxford

A thesis submitted for the degree of

DPhil in Clinical Medicine

Trinity Term 2012

The Double CUE Domain of Chromatin Remodelling Factor SMARCAD1

A thesis submitted for the degree of
DPhil in Clinical Medicine

Abstract

ATP-dependent chromatin remodellers represent a class of proteins that restructure chromatin through the action of a conserved helicase-like ATPase domain. Remodellers typically have several accessory binding domains alongside the ATPase. These confer target specificity and most commonly recognise histone post-translational modifications.

SMARCAD1 is a ubiquitous chromatin remodeller involved with DNA replication and repair. It binds directly to PCNA at the site of DNA replication and recruits co-repressor KAP1 in order to silence newly produced chromatin. In contrast to most other chromatin remodellers, SMARCAD1 does not contain several different types of accessory domains. Only two CUE domains have been identified in addition to the SMARCAD1 core ATPase domain. CUE domains are a type of helical ubiquitin-binding domain.

This thesis presents the findings of an investigation into the structure and function of the SMARCAD1 double CUE domain. The solution NMR structure is presented with results from NMR binding experiments mapped onto the structure. Each CUE domain was found to be an independent helix bundle connected by a dynamic flexible linker. The N-terminal CUE domain, CUE-1, binds ubiquitin and has an adjacent SUMO (a ubiquitin-like protein) binding motif on a protruding extended helix. The C-terminal CUE domain, CUE-2, has a very similar structure to several published CUE domains but does not bind ubiquitin due to a charged substitution at a highly conserved CUE consensus position. The SMARCAD1 double CUE domain binds KAP1 from nuclear extract and is likely to mediate the interaction between SMARCAD1 and KAP1. SMARCAD1 double CUE domain is not involved with PCNA binding.

Philip West
Linacre College

Trinity Term 2012

For my parents, John and Katrine West.



"Science is the great antidote to the poison of enthusiasm and superstition."

Adam Smith

The Wealth of Nations, 1776

Acknowledgements

I am indebted to both Erika Mancini and Christina Redfield for their patience and supervision. A sincere thank you to Erika for the opportunity, direction, and amazing level of understanding and support. Christina's generosity with time, problem solving, teaching and guidance has been astounding and invaluable. Thank you.

Many people have helped me along the way and I would like to acknowledge all members from the Division of Structural Biology for their help. I owe several people a special mention. Thank you to Kamel El Omari for showing me the ropes in the early months. Luke Yates for his unflappable optimism and friendship. Rosa Morra for her help and enthusiasm. Chris Anderton for being a great part II student. Geoff Sutton for knowing how everything works and not getting too mad (usually) when I broke it. Margaret Jones for holding everything together.

Thank you to Patrick Varga-Weisz for very useful discussions, generously supplying a number of reagents and inviting me to his lab to do experiments. Camilla Oxley for her kindness and willingness to help me try NMR. Nick Soffe for keeping the NMR facility running well.

Lastly, I would like to say thank you to my friends and family. My friends at Linacre and Oxford have been a great source of support to me and I have loved getting to know them all. Siân Davies in particular deserves a special mention for keeping the walls padded. My brothers, Kevin and Graham, and parents, John and Katrine, have all been a source of strength. I truly appreciate their generosity, understanding and support.

Contents

List of Figures **x**

List of Abbreviations **xv**

1 Introduction **1**

1.1 Chromatin 1

1.1.1 Chromatin Structure and Dynamics 3

1.1.2 Histone Modifications and Variants 6

1.1.3 Recognition of Histone Modification 13

1.2 ATP-Dependent Chromatin Remodelling 19

1.2.1 The ATPase Domain 22

1.2.2 Mechanism of Action 22

1.2.3 Importance of ATP-dependent Chromatin Remodelling 26

1.3 SMARCAD1 26

1.3.1 Isoforms 30

1.3.2 Function and Binding Partners 31

1.3.2.1 KAP1 32

1.3.2.2 PCNA 34

1.4 Ubiquitin Recognition and CUE Domains 36

1.4.1 Ubiquitin Structure and Function 36

1.4.2 Ubiquitin-Like Proteins 40

1.4.3 CUE Domains 40

1.4.4 Other Ubiquitin Binding Domains 44

1.5 Aims of this Thesis 47

2 Materials and Methods **50**

2.1 DNA Handling and Subcloning Techniques 50
2.1.1 Plasmid DNA Amplification 51
2.1.2 Transformation of Bacteria by Heat-Shock 52
2.1.3 Plasmid DNA Purification 52
2.1.4 DNA Sequencing 53
2.1.5 Glycerol Stocks 53
2.1.6 DNA Quantitation 54
2.1.7 Polymerase Chain Reaction 54
2.1.8 Agarose Gel Electrophoresis 56
2.1.9 Subcloning Step for PCR Products 56
2.1.10 DNA Gel Extraction and Purification from Agarose Gels 57
2.1.11 Restriction Digest 57
2.1.12 Ligation 58
2.2 Protein Production and Purification 58
2.2.1 Unlabelled Double CUE Expression 58
2.2.2 Isotopically Labelled Protein Expression and Purification 59
2.2.3 Purification of SMARCAD1 Double CUE from cell pellets 59
2.2.4 Mass Spectrometry of Purified Proteins 60
2.2.5 Protein Acrylamide Gel Electrophoresis 60
2.2.6 Protein Staining 62
2.2.7 Confirmation of Expression by Western Blots 63
2.2.8 Recombinant Un-tagged PCNA 64
2.2.8.1 Expression 64
2.2.8.2 Purification 65
2.3 Crystallisation Trials 66
2.4 Oligomeric State Determination 66
2.5 NMR Experiments 67
2.5.1 Sample Preparation 67
2.5.2 Data Collection 67
2.5.3 Residual Dipolar Coupling Measurements 68
2.5.4 NMR Titration Experiments 71
2.5.4.1 NMR Titration Analysis 72

2.5.5 Structure Calculations 74
2.6 Pull-down of Recombinant PCNA 76
2.7 Pull-down from Nuclear Extract 78

3 Overexpression and Initial Structural Characterisation of SMARCAD1 Double CUE Domain 79

3.1 Introduction 79
3.2 Sequence Analysis and Construct Design 80
3.2.1 Identification of CUE Domains 81
3.2.2 Secondary Structure Prediction 83
3.2.3 Protein Disorder Prediction 83
3.2.4 Construct Design 85
3.3 Protein Production 86
3.4 Oligomeric State Determination 86
3.5 Crystallisation 90
3.6 NMR Trials 92

4 Assignment of Backbone Resonances in the Nuclear Magnetic Resonance Spectrum of SMARCAD1 Double CUE Domain 95

4.1 Introduction 95
4.2 Strategy for Assignment of Backbone Resonances 96
4.3 Results 98
4.3.1 ^{1}H, ^{15}N-HSQC 98
4.3.2 Identification of Side-chain H^{N} in the ^{1}H, ^{15}N-HSQC 102
4.3.3 Triple Resonance, 3D Experiments 102
4.3.3.1 Sequential Connections Using Cα Chemical Shifts . . . 104
4.3.3.2 Spin System Identification and Connection using Cβ Chemical Shifts 104
4.3.3.3 CO Assignments 109
4.3.4 NH-NH 111
4.3.5 Final Backbone Assignment 113

5 Structure Determination of Double CUE Domains of SMARCAD1 114
5.1 Introduction . . . 114
5.2 Side-chain Assignments . . . 117
5.2.1 Hα, Hβ Assignments . . . 117
5.2.2 Additional Side-chain assignments . . . 117
5.2.3 Summary of Assignment Statistics . . . 121
5.3 Secondary Structure . . . 121
5.3.1 Characteristic Alpha Helical NOE Pattern . . . 123
5.3.2 Dihedral Angles . . . 125
5.3.2.1 TALOS+ . . . 125
5.3.2.2 DANGLE . . . 127
5.3.3 Helix-capping and β-Turn Prediction . . . 127
5.3.4 Helix Packing . . . 130
5.4 Backbone Dynamics . . . 130
5.4.1 Heteronuclear NOE . . . 130
5.5 Residual Dipolar Couplings . . . 132
5.6 NOE Distance Restraints . . . 138
5.7 Structure Calculations . . . 139
5.7.1 CUE-1 . . . 141
5.7.1.1 Two Families of Structures . . . 141
5.7.1.2 Energy-minimised Average Structure . . . 145
5.7.1.3 Comparison of CUE-1 Families with other CUE Domains 146
5.7.1.4 N-terminal End of α-helix 1 of CUE-1 . . . 146
5.7.2 CUE-2 . . . 153
5.7.3 Complete Double CUE-Domain . . . 155

6 Functional Studies of SMARCAD1 Double CUE Domain 165
6.1 Introduction . . . 165
6.2 Ubiquitin Binding . . . 166
6.2.1 NMR Titrations . . . 166
6.2.1.1 Binding Sites . . . 166
6.2.1.2 Dissociation Constant . . . 175
6.2.2 Docking Model . . . 179

6.2.2.1 Comparison with Other Helical Ubiquitin-binding Domains . . . 179
6.3 SUMO . . . 181
6.3.1 NMR Titrations . . . 181
6.3.1.1 Binding Site . . . 181
6.3.1.2 Dissociation Constant . . . 189
6.4 PCNA . . . 189
6.4.1 NMR Titrations . . . 192
6.4.2 Pull-Down Experiments and Western Blots . . . 194
6.5 KAP1 . . . 197
6.6 Additional Pull-down Experiments . . . 201

7 Discussion **203**
7.1 Structure of the Double CUE Domain . . . 203
7.2 Conserved CUE Sequence . . . 206
7.3 Binding Partners of the Double CUE Domain . . . 209
7.3.1 CUE-1 Ubiquitin and SUMO binding . . . 209
7.3.2 CUE-2 Doesn't Bind Ubiquitin . . . 211
7.3.3 KAP1 . . . 211
7.4 Linker . . . 215
7.5 Role of Double CUE Domain in SMARCAD1 . . . 216
7.5.1 Silencing of Newly Synthesised Chromatin . . . 216
7.5.2 DNA Repair . . . 218
7.6 Final Conclusions and Future Work . . . 218
7.6.1 Structure . . . 220
7.6.2 KAP1 Binding . . . 221
7.6.3 SUMO Binding . . . 221
7.6.4 Additional dCUE Binding . . . 221
7.6.5 dCUE Phosphorylation and KAP1 Binding . . . 222

Appendices **224**

A NMR Data **224**

A.1 Chemical Shift Tables 224

A.2 Dihedral Angles Used in Structure Calculations 229

A.3 Measured Residual Dipolar Coupling Values 233

B Principles of Protein NMR **236**

B.1 Nuclear Magnetic Resonance 236

B.1.1 Nuclear Spin 236

B.1.2 Nuclear Magnetism 237

B.1.3 Larmor Precession 237

B.2 Nuclear Magnetic Resonance Spectroscopy 239

B.2.1 Chemical Shift 242

B.3 Protein NMR 243

B.3.1 Multidimensional NMR 243

B.3.2 Protein Structure Determination 246

B.3.2.1 The Nuclear Overhauser Effect 246

B.3.2.2 $\{^1H\}$ ^{15}N Heteronuclear steady-state NOE 246

B.3.2.3 Residual Dipolar Couplings 248

B.3.2.4 Structure Calculations 252

References **257**

List of Figures

1.1 Hierarchy of chromatin compaction 2
1.2 Nucleosome 5
1.3 Histone Tail Modification 7
1.4 Comparison of Structural Features of Major Histone Proteins and their Variants 14
1.5 Recognition of Acetylated Lysine by Gc5pn Bromodomain 16
1.6 Recognition of Methylated Lysine by BPTF Zinc Finger 18
1.7 Classification of ATP-dependent Chromatin Remodellers by ATPase Sequence 20
1.8 Domain Organisation of ATPase Subunit of Selected Remodellers 21
1.9 Reconfiguration of Nucleosomes by Snf2-like Remodellers 24
1.10 Structure of CHD1 ATPase Domain and Chromodomains 25
1.11 Proposed Model for CHD1-mediated Nucleosome Sliding 27
1.12 Comparison of SMARCAD1 and homologues 28
1.13 Comparison of different isoforms of SMARCAD1 30
1.14 Schematic of KAP1 Domain Architecture 33
1.15 Ubiquitin-related PCNA Modifications 35
1.16 Structure of Ubiquitin 37
1.17 Crystal structures of three forms of Di-ubiquitin 39
1.18 Structure of Ubiquitin-like Proteins 41
1.19 Alignment of SUMO Proteins 41
1.20 CUE Domains 43
1.21 Major Ubiquitin Binding Patches 44
1.22 Ubiquitin Recognition Domains at the Hydrophobic Patch 45

1.23 Ubiquitin Recognition Domains at Additional Binding Interfaces 46

2.1 Structure Calculation Protocol 75

3.1 PROSITE Consensus Sequence for CUE Domain 81
3.2 Sequence Homology of CUE motifs 82
3.3 Secondary Structure 83
3.4 Disorder Prediction Plot 84
3.5 Domain Structure of SMARCAD1 85
3.6 Protein Purification of dCUE 87
3.7 Expression and Purification of Constructs 88
3.8 Western Blot Confirmation of His_6-expression 88
3.9 Mass spectrometry results 89
3.10 Multiple-angle light scattering (MALS) results 91
3.11 Formation of Non-Crystalline Protein 'Gel' 92
3.12 ^{1}H, ^{15}N-HSQC from dCUE-short 94

4.1 Overview of Backbone Assignment Process 97
4.2 SMARCAD1 dCUE-short Construct Sequence 98
4.3 ^{1}H, ^{15}N-HSQC Coherence Transfer Diagram 99
4.4 Assigned ^{1}H, ^{15}N-HSQC Spectrum for SMARCAD1 Double CUE Domain (dCUE). 100
4.5 ^{1}H, ^{15}N-HSQC 101
4.6 HSQC (NH_2 selective) 103
4.7 HNCA and HN(CO)CA Coherence Transfer Diagram 104
4.8 Cα Assignments and Connections 105
4.9 CBCANH and CBCA(CO)NH Coherence Transfer Diagram 106
4.10 Cβ Shift Assignments 107
4.11 Cβ Connections 108
4.12 HN(CA)CO and HNCO Coherence Transfer Diagram 109
4.13 Carbonyl Assignments 110
4.14 ^{15}N-NOESY-HSQC Coherence Transfer Diagram 111
4.15 NH-NH Connections from ^{15}N-NOESY-HSQC Experiment 112
4.16 Backbone Assignments of SMARCAD1 dCUE 113

5.1 Process of Protein Structure Determination using NMR 116
5.2 HBHA(CO)NH Coherence Transfer Diagram 118
5.3 ^{15}N-TOCSY-HSQC Coherence Transfer Diagram 118
5.4 Assignment of Hα and Hβ Shifts 119
5.5 HBHA(CO)NH and ^{15}N-Edited TOCSY 120
5.6 HCCH-TOCSY Coherence Transfer Diagram 121
5.7 HCCH-TOCSY Strips for Isoleucine 126 122
5.8 NOE Connections Within an Alpha-helix 124
5.9 TALOS+ Results Indicate Two Domains of Three Helices Separated by an Unstrucutred Linker 126
5.10 DANGLE Results Suggest Helical Structure of two CUE Domains. . . . 128
5.11 Helix Capping and β-Turn Prediction Results 129
5.12 Helical Wheels for CUE Domains 131
5.13 Heteronuclear NOE 133
5.14 Dipolar Couplings from HSQC-IPAP Spectra 135
5.15 Residual Dipolar Couplings 136
5.16 RDC Distribution for each CUE Domain 137
5.17 ^{13}C-NOESY-HSQC Coherence Transfer Diagram 139
5.18 Long Range Distance Restraints 140
5.19 Ensembles of Refined Energy Minimised Structures of CUE1 142
5.20 CUE-1 Low-energy Structural Families 143
5.21 Energy Minimised Average Structures for Both CUE-1 Conformations . 147
5.22 Root-mean-square deviation of Cα atoms in CUE-1 Ensembles 148
5.23 Overlay of CUE-1 Structures with Other CUE Domains 149
5.24 Comparison of CUE-1 Structures with Other CUE Domains ('Front' View) 150
5.25 Comparison of CUE-1 Structures with Other CUE Domains ('Side' View) 151
5.26 Chemical Shift Values for Cα Relative to Random Coil Values 153
5.27 The N-terminal end of CUE-1 helix 1 154
5.28 CUE-2 Ensemble of Structures and Energy Minimised Average Structure 156
5.29 Room-mean-square deviation of Cα atoms in CUE-2 Ensemble 158
5.30 Comparison of SMARCAD1 CUE Domains with Other CUE Domains . 159
5.31 Full-length dCUE 160
5.32 Superposition of 10 low-energy dCUE Structure 163

5.33 Superposition of 10 low-energy dCUE Structure 164

6.1 NMR Titration Experiment of ^{15}N-dCUE Shifts with Ubiquitin 167
6.2 Assigned ^{1}H, ^{15}N-HSQC of ^{15}N-labelled Ubiquitin 169
6.3 NMR Titration Experiment of ^{15}N Ubiquitin with Unlabelled dCUE. . . 170
6.4 Chemical Shift Differences Caused by Ubiquitin:dCUE Binding 172
6.5 Ubiquitin Residues Affected by dCUE Binding 173
6.6 SMARCAD1 CUE Domain 1 Residues Affected by Ubiquitin Binding . 174
6.7 Chemical Shift Changes from Double CUE and Ubiquitin Interactions . 177
6.8 SMARCAD1 CUE-1-Monoubiquitin Binding Model 180
6.9 CUE-Ubiquitin Binding 182
6.10 Comparison of Non-CUE Helical Ubiquitin Binding Domains (UBDs) . 183
6.11 NMR Titration Experiment of ^{15}N-dCUE Shifts with SUMO-1 185
6.12 Comparison of Chemical Shift Differences Caused by SUMO:dCUE Binding and Ubiquitin:dCUE 186
6.13 Potential SUMO Binding Motif Upstream of dCUE 188
6.14 Chemical Shift Change from Double CUE and SUMO Interactions . . . 190
6.15 NMR Titration Experiment of ^{15}N-dCUE with PCNA 193
6.16 At High Concentrations of PCNA, Peak Intensities Vary 195
6.17 Protein Concentration-Dependent, Non-Specific Peak Broadening Occurs in Structured Regions 196
6.18 Pull-down Experiment Using Recombinant PCNA 198
6.19 Pull-down Experiment Using Nuclear Extract and anti-PCNA Primary Antibody 199
6.20 Pull-down Experiment Using Nuclear Extract and anti-KAP1 Primary Antibody 200
6.21 Pull-down Experiments Using Nuclear Extract 202

7.1 Electrostatic Potentials of CUE Domains 205
7.2 Aligned CUE Domain Sequences from UniProt with CUE-1 Ubiquitin Interacting Residues 207
7.3 CUE Domain Sequence Homology 208
7.4 Conserved Residues bridge CUE Helices 1 and 3 210

7.5 Theoretical Model of CUE-1 Binding to Atypical Ubiquitin-SUMO Heterologous Chain 212
7.6 Positively Charged Lysine in Conserved Position Responsible for CUE-2 Inability to Bind Ubiquitin 213
7.7 Model of SMARCAD1 Recruitment and Function During DNA Replication 217
7.8 Model of SMARCAD1-KAP1 Response to DNA Damage 219

B.1 Bulk Magnetisation 238
B.2 Larmor Precession 239
B.3 NMR Detection 240
B.4 NMR Fourier Transform 241
B.5 HSQC Pulse Sequence and Spectrum 245
B.6 Long-Range NOE Distance Restraints 247
B.7 Dipolar Couplings 249
B.8 Residual Dipolar Couplings 251
B.9 Example of RDC Distribution 253
B.10 Schematic of NMR Structure Calculation 255

List of Abbreviations

Acetyl-CoA Acetyl-Coenzyme A

BSA Bovine serum albumin

C-terminus Carboxy terminus

CUE-1 The first CUE domain from SMARCAD1

CUE-2 The second CUE domain from SMARCAD1

Cue2 Ubiquitin-binding protein CUE2; a 445 amino acid long protein from *Saccharomyces cerevisiae* which contains 2 CUE domains

DANGLE Dihedral angles from Global Likelihood Estimations

dCUE SMARCAD1 shorter double CUE domain construct

dCUE-short SMARCAD1 shorter double CUE domain construct

DNA Deoxyribonucleic acid

DUBs Deubiquitinating enzymes

E. coli *Escherichia coli*

FID Free induction decay

FUN30 Function unknown protein 30. Yeast homologue of SMARCAD1

GST Glutathione S-transferase

HSQC Heteronuclear single-quantum correlation spectroscopy

HATs Histone acetyl-transferases

HDACs Histone deacetylases

HKMT Histone lysine methyltransferase

H1 Histone H1

H2A Histone H2A

H2B Histone H2B

H3 Histone H3

H4 Histone H4

INEPT Insensitive Nuclei Enhanced by Polarization Transfer

IPAP In-Phase / AntiPhase NMR experiment

IPTG Isopropyl-β-D-thiogalactopyranoside

KAP1 KRAB (Krüppel associated box)-associated protein 1

LB Lysogeny broth: a nutritionally rich media for growing bacteria

MIU motif interacting with ubiquitin - a single helical motif that binds ubiquitin

N-terminal Amino terminal

NETs Neutrophil extracellular traps

NMR Nuclear magnetic resonance

NOSEY Nuclear Overhauser effect spectroscopy.

NOE Nuclear Overhauser effect

NURF Nuclear remodeling factor; a chromatin remodelling complex

PARP1 Poly [ADP-ribose] polymerase 1

PBS Phosphate buffered saline; 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4 \cdot 2\,H_2O$, 2.0 mM KH_2PO_4, pH 7.4.

PCNA Proliferating cell nuclear antigen

PCR Polymerase chain reaction

PDB Protein data bank

PHD Plant homeo domain

PMSF phenylmethanesulfonyl fluoride; an inhibitor of serine proteases such as trypsin and chymotrypsin

ppm Parts per million

PTM Post-translational modification

RAMA The RAMA potential energy term

RDCs Residual dipolar couplings

RONN Regional order neural network

SBM SUMO binding motif.

SUMO Small ubiquitin-like modifier.

TOCSY Total correlated spectroscopy.

TB Terrific broth

UBA Ubiquitin associated domain (structurally related to CUE domains with a shared three-helix bundle)

UIM Ubiquitin-interacting motif (specific single helical motif that binds ubiquitin)

WSTF Williams syndrome transcription factor

Chapter 1

Introduction

1.1 Chromatin

A single human cell contains approximately 1.8 metres of DNA in length, which must be compacted many thousand-fold to fit into the cell's nucleus less than 10 μm in diameter and yet remain accessible for essential cellular processes (1, 2, 3). Compacted nuclear DNA is present as a complex called chromatin, which consists of nucleic acid coated with at least an equal mass of protein (4). Chromatin is dynamic, highly regulated and maintained in such a way that the DNA is selectively available to be read by and interact with binding partners during processes such as replication, transcription, recombination and repair (3, 5). All eukaryotes share a highly conserved scheme where genomic DNA is wound around a series of disc-like core histone proteins to form a linear array of nucelosomes, which are further compacted into higher order structures, ultimately forming chromosomes (figure 1.1) (4).

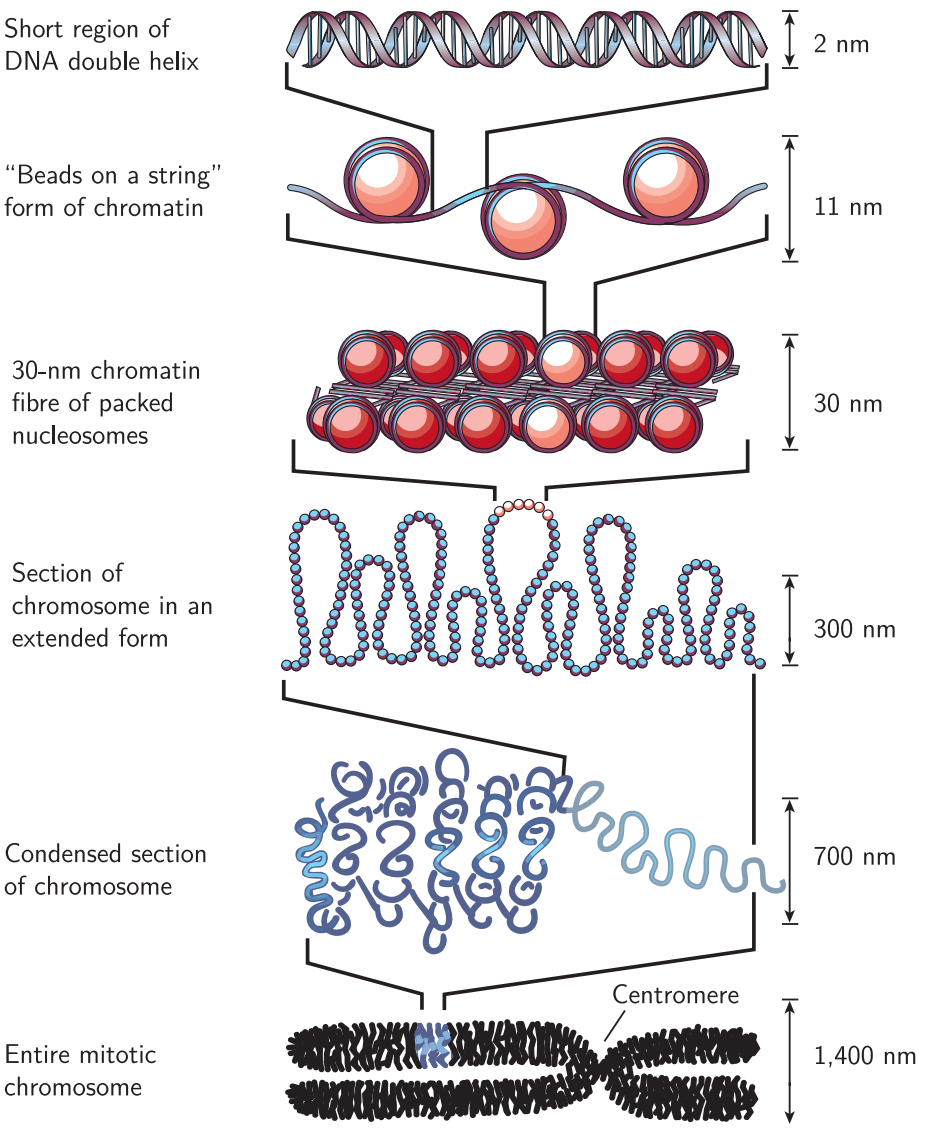


Figure 1.1: Hierarchy of chromatin compaction - DNA is wound around core histones to produce nucleosomes, which are connected by short stretches of linker DNA giving rise to the "beads on a string" form of chromatin. Linked nucleosomes are further compacted to produce a 30 nm fibre, which folds into higher-order structures and eventually forms chromosomes. Understanding of chromatin structure beyond that of the nucleosome remains unclear (4, 6). Adapted from Felsenfeld, 2003.

1.1.1 Chromatin Structure and Dynamics

Chromatin is increasingly recognised as an active regulator of gene expression and not simply a passive scaffold (7). Understanding the nature of chromatin structure and its regulation is central to fully explaining the mechanisms that control development, cancer and stem cell biology (3).

During mitotic division, chromatin can be visualised as individual, compact chromosomes, however, in non-dividing cells, chromatin is distributed throughout the nucleus. Broadly speaking, this is organised into pockets of condensed heterochromatin generally located at the edge of the nucleus and pockets of more 'open' euchromatin (4, 8). The latter is relatively relaxed, contains active genes and experiences regular changes throughout the cell cycle. In contrast, heterochromatic regions are more compact, contain mainly inactive genes and are generally protected from cyclical changes during the cell-cycle (9).

Recent research in *Drosophila* has identified five distinct states of chromatin, suggesting that this binary description of chromatin may be too simplistic (10, 11). It remains the case, however, that chromatin domains can be grouped depending on whether they contain active or inactive genes; euchromatin and heterochromatin are still useful distinctions (9). Additionally there are classically recognised sub-types of both hete-

rochromatin and euchromatin. Islands of localised modifications within euchromatin are associated with transcription and transcriptional regulation. So-called facultative heterochromatin corresponds to regions of genes that are variably expressed throughout development / differentiation and then become silenced. Constitutive heterochromatin contains permanently silenced genes and is found at centromeres and telomeres (9).

Demarcation between chromatin types is thought to be maintained by insulator proteins associated with active gene promoters, which prevent transcriptionally silent, heterochromatin from spreading to regions of active euchromatin (12). Clusters of insulator proteins are proposed to play a key role in ordering chromatin at a high level through the formation of 'insulator bodies' that spatially localise co-regulated genes by facilitating long-range interactions between and within chromosomes by forming large chromatin loops (13, 14).

At a lower level, the fundamental repeating unit of chromatin is the nucleosome, which is made from 147 base pairs of DNA wrapped almost completely twice around an octamer of core histone proteins (2 copies of each H2A, H2B, H3 and H4) (figure 1.2). Nucleosomes are separated by variable length linker DNA to give the "beads on a string" form of chromatin, which is thought to be compacted approximately 25-fold by the addition of a fifth histone protein, linker histone H1, to form a 30 nm fibre (3, 7). The structure of this fibre and mechanisms used to produce higher order

levels of compaction are poorly understood and the subject of considerable debate (8, 15, 16, 17, 18).

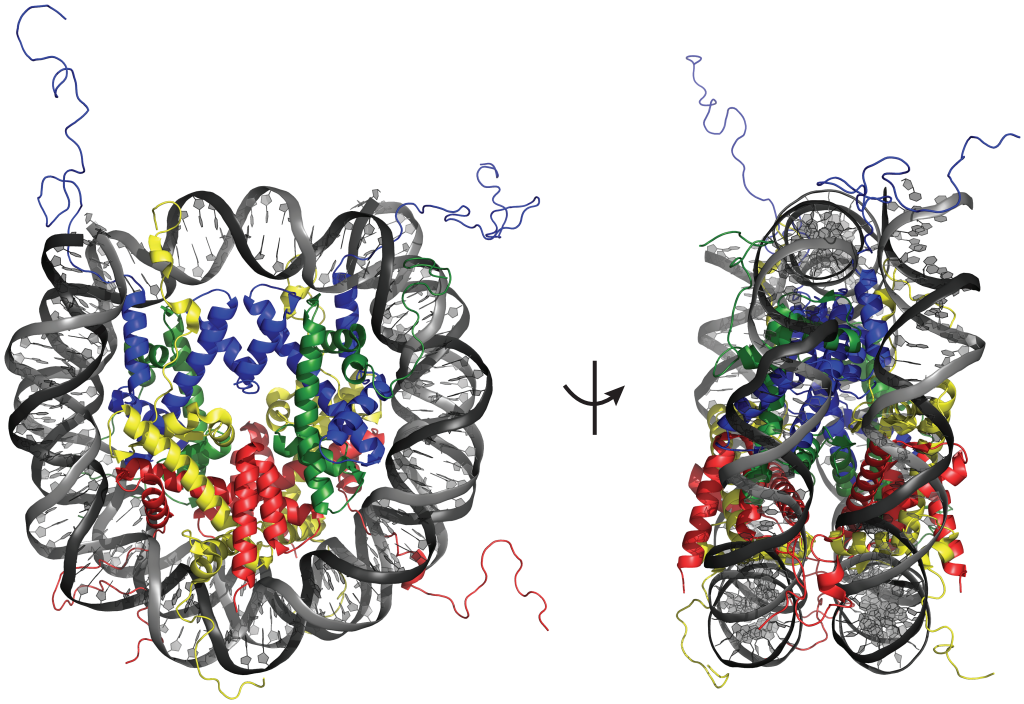


Figure 1.2: Nucleosome - The fundamental unit of DNA packaging in eukaryotes, consisting of 147 base pairs of DNA wound around the surface of a histone protein core in 1 3/4 superhelical turns (19). The core octamer has two of each core histones H2A (yellow), H2B (red) , H3 (blue) and H4 (green). DNA is drawn in grey. Accessible histone tails, which represent 25-30% of core histone mass, are thought to be unstructured and radiate outward from the core. Generally the majority of these are not visible in crystal structures due to inherent flexibility (20). PDB: 1KX5 (21), rendered in PyMOL (22).

The dynamic nature of chromatin is tightly regulated through considerable localised modification. For example, nucleosomes can contain functionally and structurally important histone variants and protruding histone tails can have numerous post-translational covalent modification marks along their length. The overall effect on

chromatin structure is a result of the complex interplay of complementary and opposing modifications. Nucleosomes can be slid along DNA or completely removed to expose DNA. Some regions of DNA are kept nucleosome-free (characteristic of eukaryotic gene promoters) (3, 9).

1.1.2 Histone Modifications and Variants

Since the early 1960s we have known that histone proteins are post-translationally modified (23). We are now beginning to understand the importance of these modifications and how they are involved with regulating most processes involving DNA. Our understanding of the types of histone modifications and their combinatorial complexity is continuing to expand (9). The central concept is that highly basic N-terminal histone tails protrude from the centre of nucleosomes and are covalently modified at specific residues (figure 1.3). These modified tails can affect chromatin structure by modulating direct interactions between nucleosomes and, importantly, by recruiting additional protein that utilise ATP to alter the position of nucleosomes (4, 9, 19). Modifications within the core domain have also been identified and shown to regulate heterochromatic silencing, DNA repair, transcription and chromatin assembly (24). Together these modifications represent a "histone-code", critically important for determining how the associated DNA sequence is maintained and expressed (25, 26).

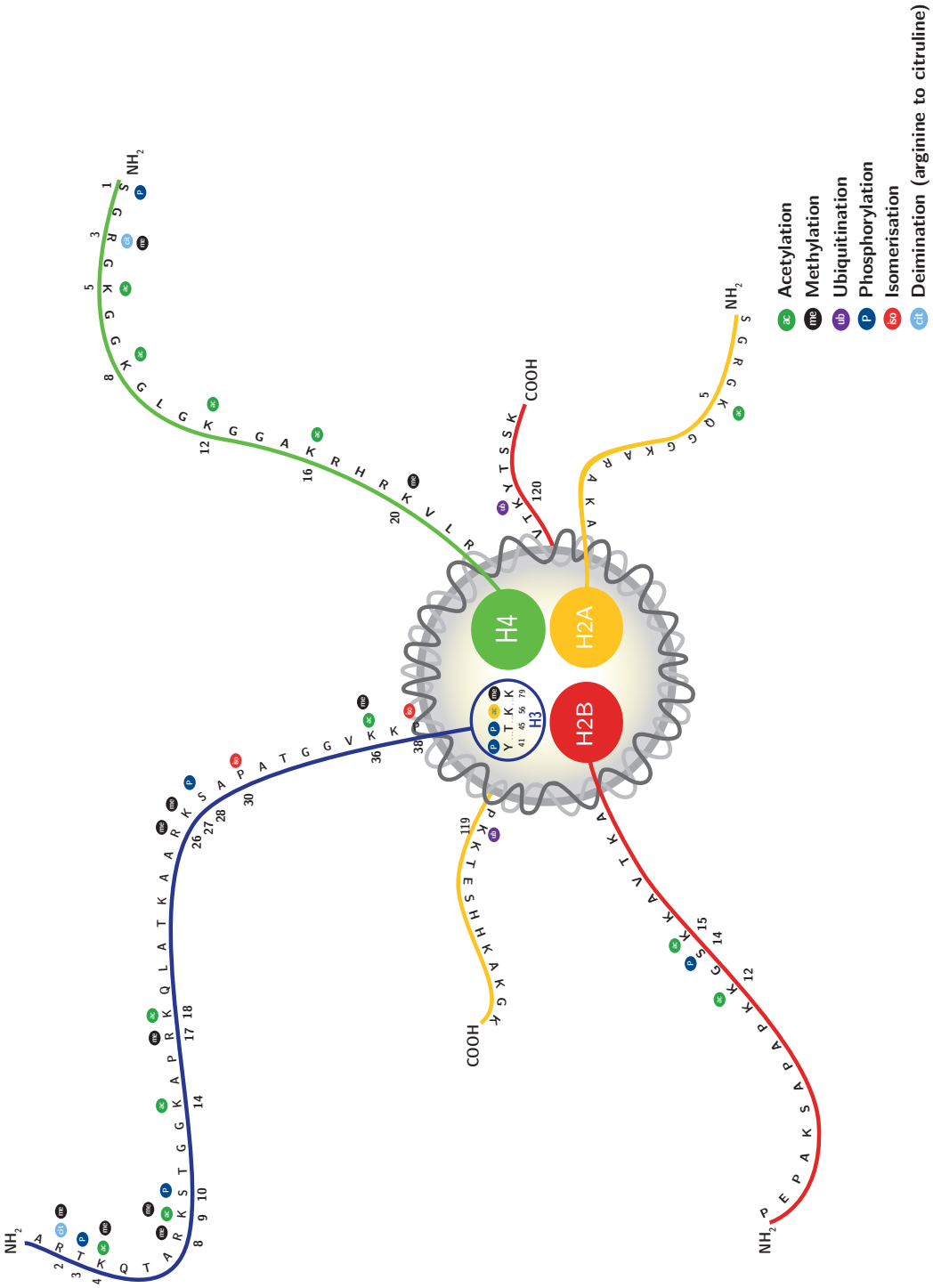


Figure 1.3: Histone Tail Modification - Schematic diagram of a nucleosome with protruding histone tails and sites of various known modifications. Many tail residues are capable of being modified to produce histone code. This does not represent a complete list (27, 28, 29).

One of the most significant post-translational modifications (PTM) is the acetylation of lysine residues, which is controlled by the actions of two families of enzymes: histone acetyl-transferases (HATs) that add acetyl marks and histone deacetylases (HDACs) that remove them (9, 30). HATs catalyse the transfer of an acetyl group from acetyl-CoA to the ϵ-amino group of lysine side chains, which neutralises the lysine's positive charge and could weaken the interactions between histones and DNA. (9, 31). In general, acetylation of histone tails is associated with active genes (32). hesPost-translational phosphorylation, like lysine acetylation, is a reversible marker that alters side chain charge. Phosphorylation normally occurs on the hydroxyl group of serine, threonines and tyrosines located in histone N-terminal tails, although sites within the histone core, such as H3Y41, can also be modified. Histone kinases transfer a phosphate group from ATP to the target hydroxyl group, introducing negative charge that affects chromatin structure (9, 33). Specific histone phosphorylation sites are associated with cell-cycle regulation and chromosome condensation. Relatively little is known about histone phosphatases but it's clear that removal of specific phosphorylation sites, working in opposition to particular kinases, occurs frequently and that this is a dynamic process (34, 35, 36).

Methylation of histone lysine and arginine side-chains is another important type of modification that, unlike phosphorylation and acetylation marks, does not alter the

charge of the histone (9). Additionally, it was long thought that methylation was a permanent modification, however both arginine and lysine demethylases have been reported (37, 38). Specificity for a given methylated residue is further complicated as both lysine and arginine side-chains can be methylated in several different ways, which has implications for protein interactions. Methylated histone lysine residues are either mono-, di- or tri-methylated on their amine group by one of numerous histone methyltransferases (HKMT). The ω-guanidino group of arginine residues can be either mono- or di-methylated, with the di-methylated form present as either symmetric or asymmetric (i.e. both methyl groups attached to the same N atom or spread symmetrically across both N atoms) (39, 40, 41, 42). Histone methylation is associated with both gene activation and repression. This is dependent upon both the specific residues involved and the extent of methylation. For example, mono-methylated H3K27, H3K9, H4K20, H3K79, and H2BK5 are linked to gene activation, whereas trimethylated H3K27, H3K9, and H3K79 are linked to gene repression (43, 44).

Ubiquitin can also be attached to several lysine residues in histones H2A and H2B and has a role in transcription regulation. Post-translational modification with ubiquitin represents a considerably larger mark than those mentioned above and is thought to affect nucleosome stability (45). Ubiquitin is a 76 amino acid long polypeptide that can be linked to histones as either mono- or poly-ubiqutin in a three step process. Ubiqui-

tin is first activated by an activating enzyme (E1), then coupled to a cysteine residue in a ubiquitin-conjugating enzyme (E2) before being transferred to the target lysine residue by a ubiquitin-protein isopeptide ligase (E3) (46). The specific lysine residue targeted for modification is determined by the enzyme complexes involved (9). Ubiquitin modifications are highly dynamic and removed by a group of isopeptidases called Deubiquitinating enzymes (DUBs) (47). Histones modified with di-ubiquitin play a role in mounting the response to DNA damage (45) and mono-ubiquitinated H2A is thought to be involved with polycomb-mediated transcriptional silencing (48, 49).

Sumoylation is a related modification to that of ubiquitination. It involves the addition of SUMO (small ubiqutin-like modifier) proteins to lysines via dedicated E1, E2 and E3 enzymes. All four core histone proteins are known to be sumoylated and data suggests a repressive role, antagonising acetylation and ubiquitination. In keeping with the idea that SUMO acts as a repressive mark, histone variant H2AZ, which is associated with active transcription, is significantly less sumoylated than other histones (9, 50).

Several other forms of histone modification exist that are less well understood, these include deimination, ADP ribosylation, proline isomerisation, β-N-acetylglucosamine, biotinylation and histone tail clipping (9, 51, 52).

Deimination is the conversion of arginine or mono-methyl arginine to citrulline,

which neutralises arginine's positive charge (and in the case of mono-methyl arginine, removes the methyl mark) (53, 54). Deimination of histones H3 and H4 by piptidyl deiminase PAD4 is required for the formation of the decondensed chromatin present in neutrophil extracellular traps (NETs); a relatively newly discovered feature of the innate immune response used by neutrophils to kill bacteria (55, 56, 57).

Both mono- and poly-ADP ribosylation of histone glutamate and arginine residues are reversible modifications that confer a negative charge, also correlated with a relaxation of the chromatin state (9). Mono-ADP ribosylation of histone H1 is associated with DNA damage response (58, 59).

Proline isomerisation is the interconversion between *cis* and *trans* dihedral angle conformations of the peptide bond. In *Saccharomyces cerevisiae*, this has been found to regulate the methylation of a neighbouring lysine residue with implications for gene regulation. Specifically, *Saccharomyces cerevisiae* histone H3P38 interconversion through the action of the proline isomerase scFpr4 is a necessary precursor to methylation of H3K36 by scSet2 methyltransferase or demethylation by scJMJD2 demethylase (60).

Like many proteins, histones H2A, H2B and H4 have been shown to be dynamically modified by β-N-acetylglucosamine (O-GlcNAc). The sugar residue is transferred from UDP-GlcNAc to serine and threonine residues by the action of the enzyme O-GlcNAc transferase working with various interacting subunits to confer specificity. β-N-

acetylglucosaminidase (O-GlcNAcase) works in opposition and can removing the mark (61). Histone modification with β-N-acetylglucosamine increases after cellular stress events such as heat shock and is correlated with increased chromatin compaction, suggesting a role in protecting DNA from damage (62).

Biotinylation of histone lysines is a poorly understood and somewhat contentious modification that has now been shown to occur on histones H3 and H4 with a good degree of certainly, although possibly only very rarely (52, 63). Biotinylated histones have been detected at repeat regions and repressed loci with a proposed role in genome maintenance (51, 52, 63).

Histone tail clipping is an unusual modification in that part of the N-terminal is permanently removed by proteolytic enzyme activity. So far this has only been shown to occur on histone H3, where the first 21 residues in both mice and yeast are removed, and in the protozoa *Tetrahymena*, where the first six residues are cleaved (64, 65, 66, 67). H3 tail clipping in yeast regulates gene expression and in mouse it has been shown to be involved with embryonic stem cell differentiation. Interestingly, H3 tail clipping appears to be both negatively and positively regulated by associated covalent modifications methylation and acetylation. (64, 65)

Instead of modifications to the canonical histone proteins discussed above, there are also a number of variant histones that replace particular histones at defined regions

of chromatin. Unlike the major histone protein, expression of histone variants is not exclusively associated with DNA replication. Variant histone proteins are expressed throughout the cell cycle and deposited both during and after S phase and often have specialised functions. (68). Centromeric protein A (CENP-A) is a centromere specific variant of histone H3 protein, which is involved with recruiting nucleosome associated complex (NAC), required for correct chromosome alignment and segregation (69). Figure 1.4 compares some of the structural features between canonical histones and variants.

1.1.3 Recognition of Histone Modification

Histone post-translational modifications can be thought of as representing a 'histone code' that can be read by other proteins (25). Given the variety of modifications, a diverse range of recognition domains exists that are capable of recognising specific histone modifications (table 1.1) (71). Many of these domains that recognise histone modifications occur in tandem as part of a larger protein and, in some cases, alongside various other histone recognition domains. For example, CHD5 (Chromodomain-helicase-DNA-binding protein 5) is a chromatin remodelling factor that contains tandem PHD domains and tandem Chromodomains (72). This adds additional complexity in the recognition of histone marks and it remains unclear as to whether tandem do-

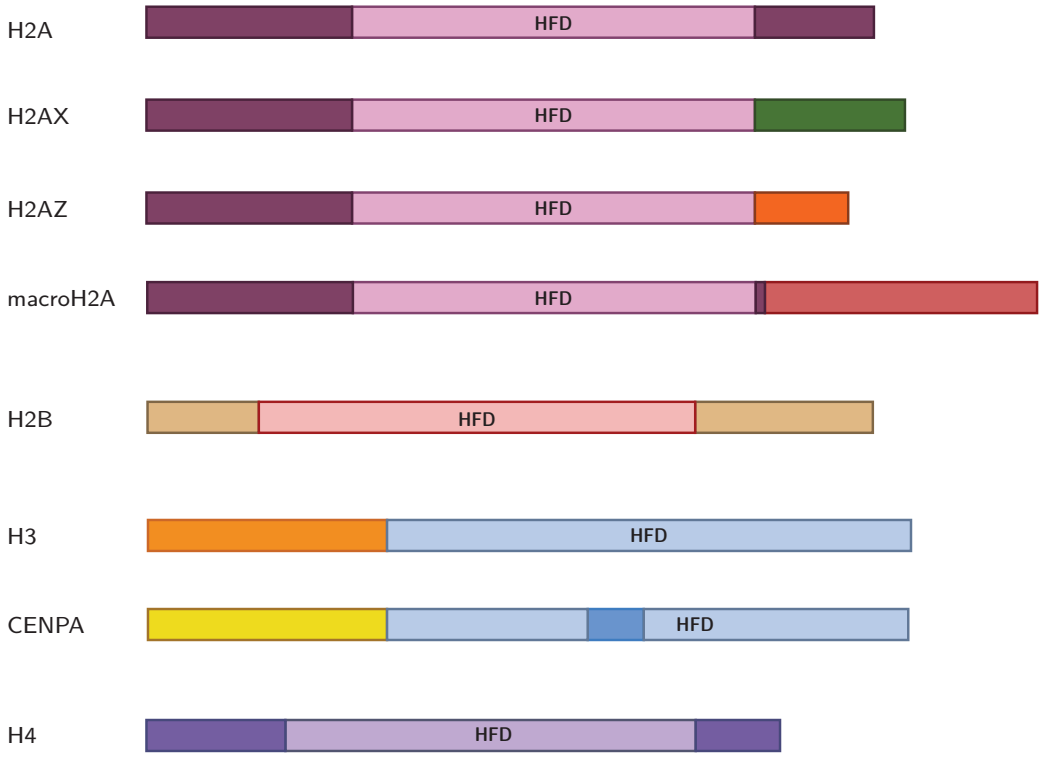


Figure 1.4: Comparison of Structural Features of Major Histone Proteins and their Variants - Histones contain a conserved histone-fold domain (HFD) with N- and C-terminal tails. Variant histones with altering amino acid sequences share structural features but have distinct differences to enable them to perform a specific function. Variants are capable of recruiting associated proteins or being covalently modified at different sites from their canonical counterparts. For example, several H2A variants have varying C terminals (shown top, green, orange and red). H2AX is involved with DNA repair, H2AZ has a role in chromosome segregation and recombination and macroH2A is associated with X chromosome inactivation. CENP-A, a variant of histone H3 present in the centromere and required for kinetochore assembly, contains an alternate N terminus (yellow instead of orange), which is highly variable between species (70). A section of the globular HFD domain (dark blue) is responsible for targeting CENP-A to the centromere (68). Adapted from Sarma, 2005.

mains typically recognise multiple modifications simultaneously or confer recognition of mutually exclusive marks (73).

Reader Module	Histone Modification Recognised
Bromodomain	Acetylated Lysine
⋆ Chromodomain	Methylated Lysine (H3K9me2/3, H3K27me2/3)
⋆ Double Chromodomain	Methylated Lysine (H3K4me1/2/3)
⋆ Chromo Barrel	Methylated Lysine (H3K36me2/3)
⋆ Tudor	Methylated Arginine (Rme2s)
⋆ Double Tudor	Methylated Lysine (H3K4me3, H4K20me3, H4K20me1/2)
⋆ MBT	Methylated Lysine (H4K20me1/2, H1K26me1/2 H3K4me1, H3K9me1/2)
PHD Finger	Methylated Lysine (H3K4me3, H3K4me0 H3K9me3, H3K36me3)
WD40 Repeat	Phosphoserine, Phosphothreonine, Methylated Arginine and Methylated Lysine, (H3R2/K4me2),
14-3-3	Phosphoserine, Phosphothreonine (H3S10ph, H3S28ph)
BRCT	Phosphoserine, Phosphothreonine (H2AX-S139ph)
UBA	Ubiqutinated Lysine
UIM	Ubiqutinated Lysine

Table 1.1: Overview of Recognition Modules for Histone Post-translational Modifications Different classes of recognition modules exist that specialise in recognising specific modifications. Examples of histone sites known to be recognised by domain family members are shown in parenthesis. ⋆ Denotes recognition domains that are members of the Royal family of protein folds (71, 74).

Bromodomains are evolutionarily conserved alpha helical protein modules that recognise acetyllysine in histone tails and other proteins (figure 1.5). They are typically about 110 amino acids long and form a bundle of four α-helices. Bromodomains are found in many chromatin-associated proteins and almost all known nuclear histone acetyltransferases (HATs) (71, 75, 76).

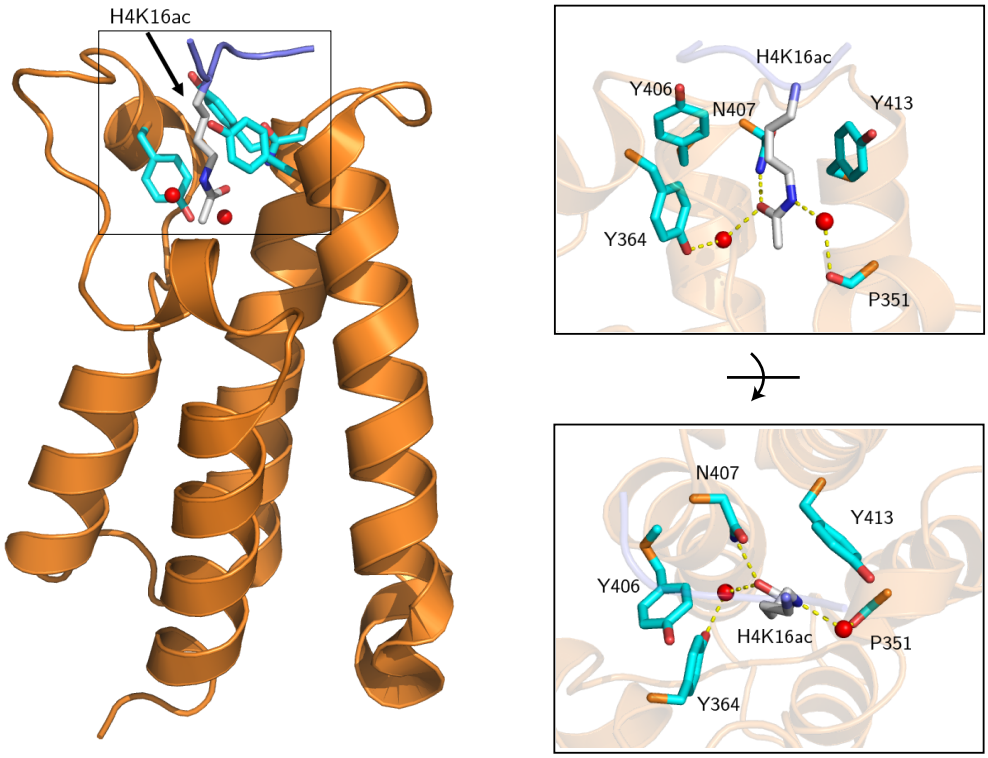


Figure 1.5: Recognition of Acetylated Lysine by Gc5pn Bromodomain - The N-terminal tail of Histone H4 (purple) bound to the bromodomain domain of the acetyltransferase Gc5pn (orange). Hydrogen bonds and water molcules shown in yellow and red respectively. Interacting residues from the bromodomain binding interface are shown in cyan. PDB:1E6I (71, 76, 77).

Methyllysine residues can be recognised by a 'cage' of aromatic residues. Two classes of protein modules are capable of binding histone methylated lysines in this manner; members of the Royal family of folds and Plant Homeo Domain (PHD) fingers (71).

The Royal family is a superfamily of structurally related protein folds derived from a common ancestor with methylated substrate binding. Members include Tudor domains, PWWP domains (from a Pro-Trp-Trp-Pro motif), MBT (malignant brain tumour) domains and chromodomains. All share a curved three strand β-sheet and short 3_{10} helix with additional strands or helices depending on the specific domain type (78).

PHD zinc fingers are approximately 50 residues in size, coordinate Zn^{2+} ions and are structurally related to RING domains, a motif common to many proteins found in the ubiquitination pathway. They are relatively common motifs, with over 150 occurrences in the human genome, predominantly in nuclear proteins (73, 79). PHD fingers typically have a two stranded β-sheet and an α-helix. The PHD domain in BPTF (the largest subunit in the NURF chromatin remodelling complex) forms a contact with histone tails containing a methylysine along its antiparallel β-sheet. This effectively extends the β-sheet of the PHD domain and brings the methyllysine residue into contact with the aromatic caging residues in the PHD finger (figure 1.6) (71, 80, 81).

Our understanding of the mechanisms and protein domains involved in recognising the full range of histone modifications continues to expand. It is known that various

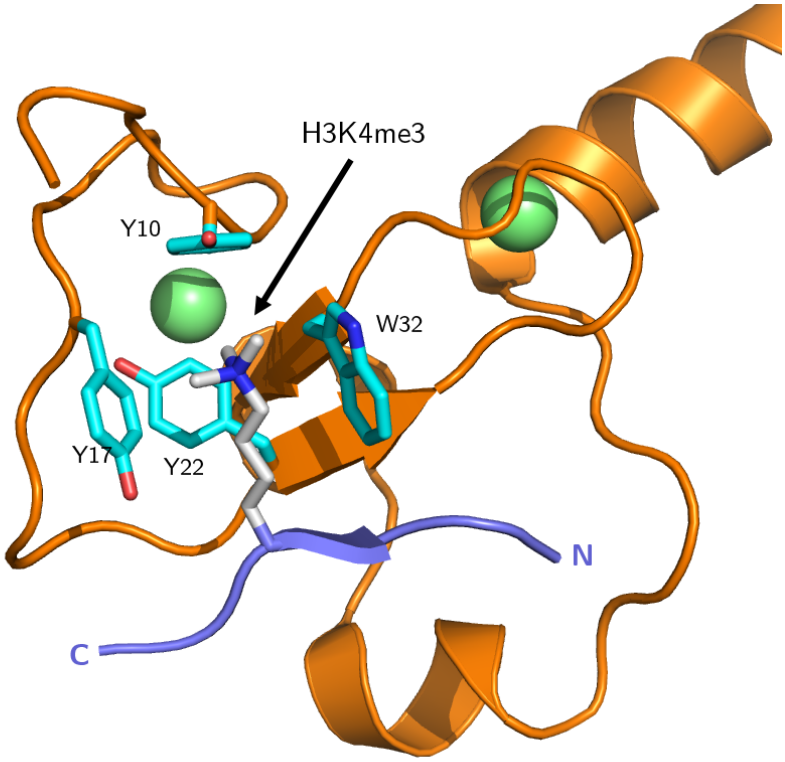


Figure 1.6: Recognition of Methylated Lysine by BPTF Zinc Finger - The N-terminal tail of Histone H3 (purple) bound to the PHD domain of BPTF (orange). H3 forms a third strand in the antiparallel β-sheet while the tri-methylated lysine interacts with aromatic caging residues from the PHD domain (cyan). Coordinated zinc ions are shown in green. PDB: 2F6J (71, 80).

protein modules, such as the WD40 Repeat, BRCT domains, BIR domains (Baculovirus Inhibitor of apoptosis protein Repeat) and 14-3-3 proteins can bind phosphorylated serine / threonine but very little is known about recognition of ubiqutinated histones or other less well studied histone modifications (71, 82, 83).

1.2 ATP-Dependent Chromatin Remodelling

ATP-dependent chromatin-remodelling factors alter interactions between DNA and histones to restructure, move and eject nucleosomes (84). Many remodelling factors have been identified, ranging in complexity from relatively simple proteins to very large megadalton-sized protein complexes with numerous, variable components. All, however, share a related ATPase with a conserved ATPase domain belonging to one of several highly conserved subfamilies of the Snf2 family of helicase-like proteins (85) (figure 1.7). The ATPase in many remodellers contain a variety of accessory domains, which, in addition to other domains from other subunits, are thought to direct their function by binding to various substrates. The accessory domains alongside the main ATPase domain provide an alternative basis for classifying ATP-dependent chromatin remodeller families (figure 1.8) (86, 87).

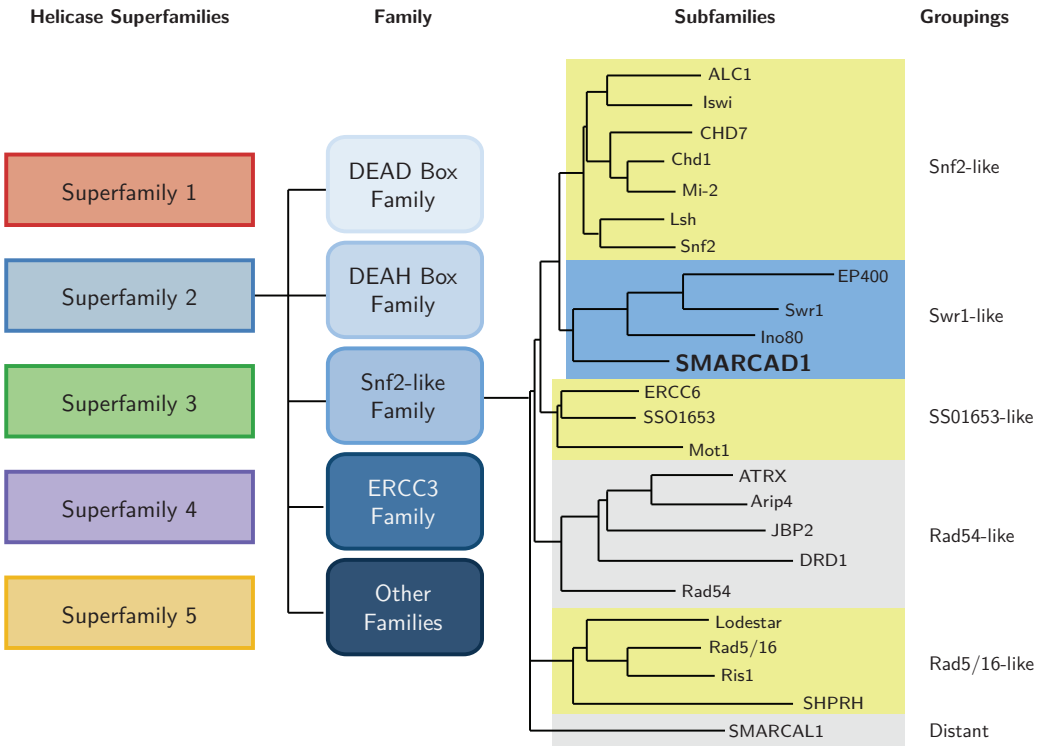


Figure 1.7: Classification of ATP-dependent Chromatin Remodellers by ATPase Sequence - ATP-dependent chromatin remodellers can be categorised according to the sequence of their ATPase domain. All share a conserved ATPase domain, which based on seven conserved helicase motifs, belong to the helicase-like superfamily 2 (SF2). 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 of over 1300 family member sequences, 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 subfamily (also known as the Etl1 subfamily), which is in the Swr1-like grouping, shown in blue (85, 88, 89). Adapted from Flaus *et al*, 2006 (85).

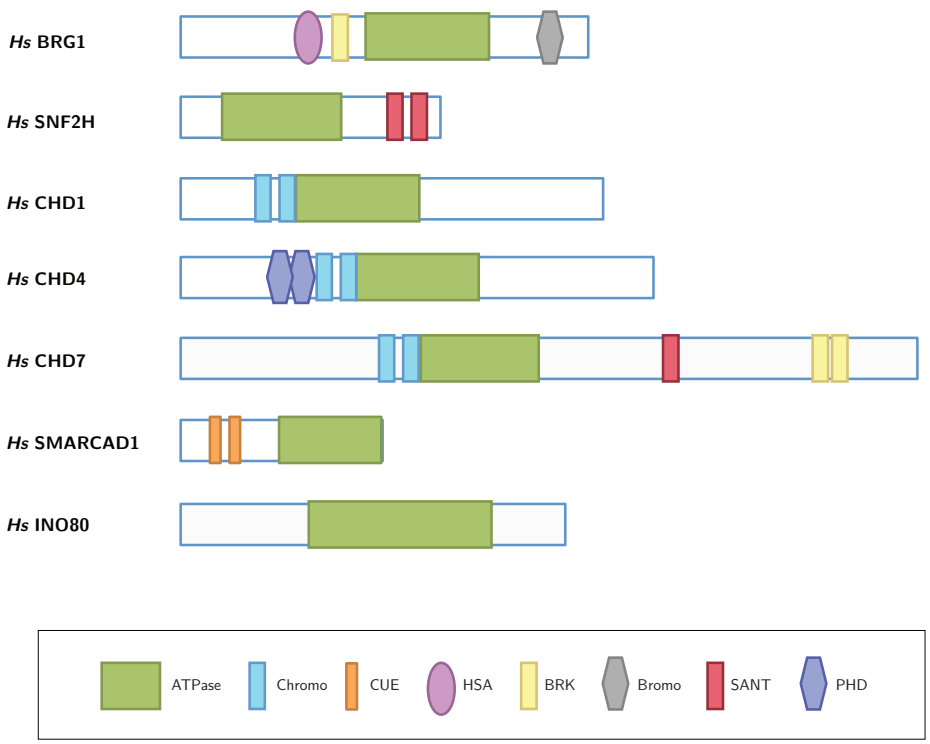


Figure 1.8: Domain Organisation of ATPase Subunit of Selected Remodellers - In addition to the conserved ATPase domain, the ATPases have a variety of different subdomains such as SANT domains (named from Swi3, Ada2, N-Cor, and TFIIIB proteins), HSA domains (helicase-SANT-associated), BRK domain (from Brahma and Kismet proteins; also known as TCH domain), PHD (plant homeo domain) and CUE domains (named after yeast protein Cue1; couples ubiquitin conjugation to ER degradation). Many of these are recognition domains and play a role in targeting the protein / complex so that the ATPase domain can facilitate remodelling. The organisation of these accessory domains can be used as an alternative way of classifying subfamilies of ATP-dependent chromatin remodellers. For example, CHD1-9 proteins can be grouped together in the CHD family based on their shared tandem-chromo domains N-terminal of the ATPase domain (3, 90).

1.2.1 The ATPase Domain

Sequence analysis of the ATPase domain in chromatin remodellers suggests they are very similar to DEAD/H helicases, families within Superfamily 2 of helicase-like proteins. The family of ATPases in chromatin remodellers, the Snf2-like family, are named after their homology to *S. cerevisiae* Snf2 protein and also belong in the Superfamily 2 of helicase-like proteins (91). Snf2-like ATPases are, however, not helicases as they do not split nucleic acid strands. Snf2-like proteins are DNA translocases, which utilise energy from ATP hydrolysis to generate torsional strain on DNA (3, 87). The Snf2-like ATPase domain consists of two tandem RecA-like lobes (87, 89, 92). RecA is an *E. coli* DNA-dependent ATPase with multiple DNA binding sites and a key role in DNA repair through homologous recombination (93, 94).

1.2.2 Mechanism of Action

Snf2-like family members remodel chromatin by their action on nucleosomes. The way in which the force generated by ATP hydrolysis is converted into remodelling remains unclear and the majority of chromatin remodellers are part of large multi-subunit complexes, which are likely to affect the action of the ATPase (3). There are a number of different reconfigurations that can occur due to ATPase activity (figure 1.9). The progress towards a particular remodelled state is thought to be dependent

on the specific remodeller. These potential outcomes are sliding of nucleosomes along DNA, either away or towards a specific position such as the end of DNA or neighbouring nucleosomes, removal of nucleosomes, swapping of nucleosomes across DNA or exchange of histone variants within a nucleosome (87).

CHD1 (Chromodomain Helicase DNA-binding protein 1) is an example of a well established ATP-dependent chromatin remodeller capable of sliding nucleosomes (95). It is known to space nucleosomes along DNA, which is thought to be important in genome stability, especially after DNA replication and recognition of H3K4me3 by the PHD domains of CHD1 has been shown to target the protein to sites of active transcription (3, 96, 97). CHD1 has a double chromodomain directly N-terminal of the ATPase and a C-terminal DNA binding region. A structure of the tandem chromodomains and ATPase region has been solved, which shows the chromodomains in direct contact with the ATPase domain (figure 1.10) (86, 87, 95, 98).

A model for the functioning of CHD1 based on two alternatives states of the ATPase domain has been proposed (figure 1.11). These are suggested to be regulated by the chromodomains and produce “gated” and “ungated” forms. In the gated state, the chromodomains inhibit action of the ATPase by blocking binding to double stranded DNA. In the ungated state, the ATPase is accessible for DNA binding and ATP hydrolysis. It is suggested that this gating model improves specificity of CHD1, allowing

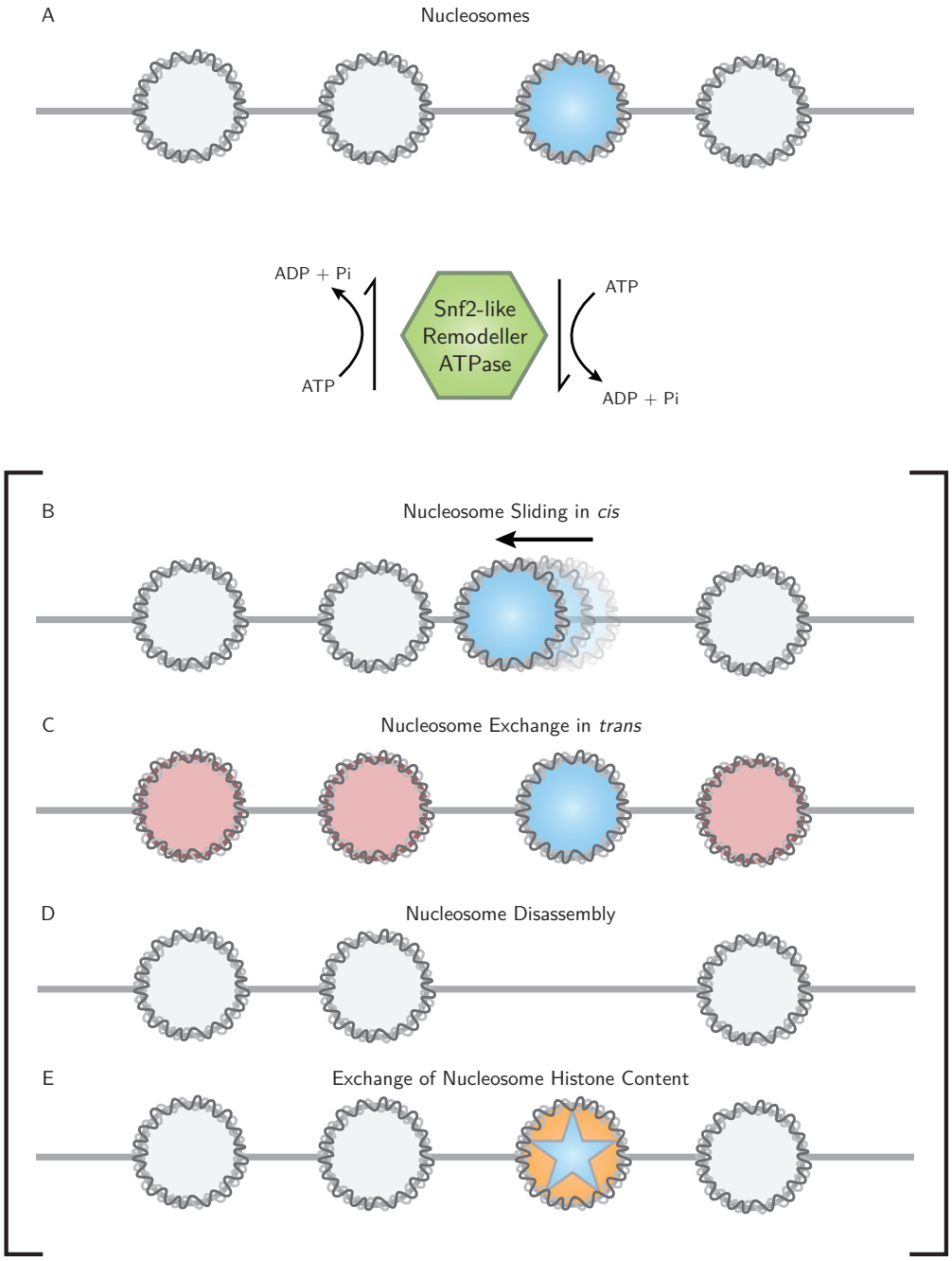


Figure 1.9: Reconfiguration of Nucleosomes by Snf2-like Remodellers - Snf2-like ATPases are capable of remodelling nucleosomes (A) to produce a variety of different outputs: (B) nucleosomes slid along DNA; (C) histones transferred between DNA molecules; (D) nucleosomes can be disassembled / assembled or (E) histone variants can be swapped in or out to alter the nucleosome composition. The precise type of remodelling is thought to be dependent on the specific ATPase involved. The reverse reactions are also possible and often catalysed by different Snf2 proteins. Adapted from Ryan *et al*, 2011 (87).

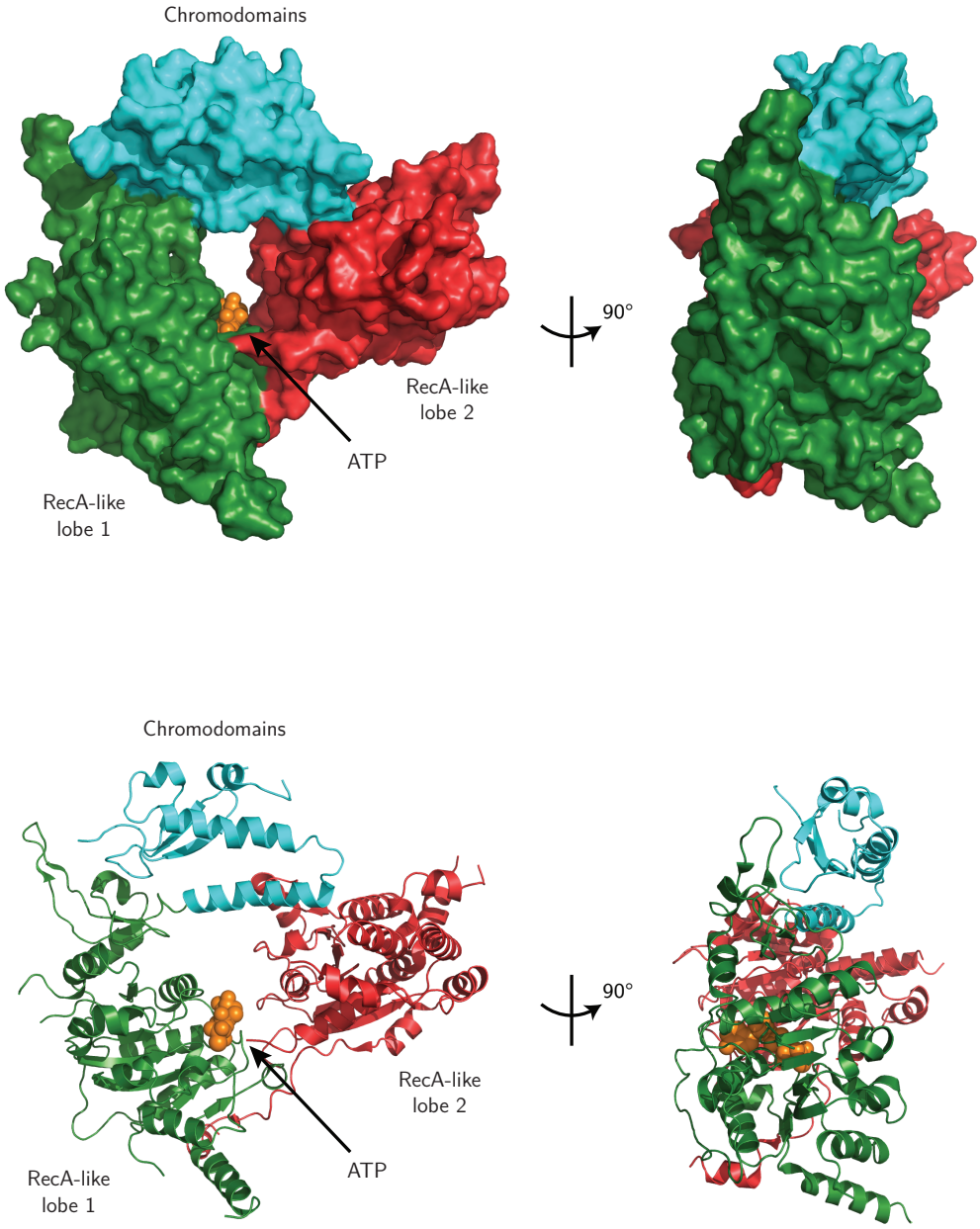


Figure 1.10: Structure of CHD1 ATPase Domain and Chromodomains - Top Surface representation of CHD1. **Bottom** Cartoon representation. The chromodomains (cyan), attached to the ATPase domain (green and red), is shown in a 'closed' conformation, preventing access to the DNA binding sites on the two RecA-like lobes of the ATPase domain. The DNA binding sites are in the middle of the structure, near the ATP binding site (orange). PDB: 3MWY (98, 99).

it to differentiate between naked DNA and nucleosomes (98).

1.2.3 Importance of ATP-dependent Chromatin Remodelling

Given the key role in genome maintenance and transcriptional regulation of chromatin remodelling complexes, it is clear that aberrant regulation could have significant impact. Various chromatin remodelling complexes, such as the BAF complex, have been implicated in cancer initiation or progression (3). CHD7, a member of the CHD family, has been shown to be mutated in CHARGE syndrome, a disease that causes deafness, blindness and affects cardiac, craniofacial, and nervous system development (100).

1.3 SMARCAD1

SMARCAD1 (SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin, subfamily A, containing DEAD/H box 1; also known as Etl1, HEL1 or KIAA1122) is a relatively poorly understood SWI/SNF-like mammalian chromatin remodelling factor (101, 102). SMARCAD1 is ubiquitously expressed and has been implicated in regulating pluripotency, self-renewal and DNA repair (103, 104). The protein is conserved from yeast to human although the exact domain arrangement varies (figure 1.12).

SMARCAD1 consists of a C-terminal helicase domain with two N-terminal CUE domains. CUE domains are traditionally regarded as ubiquitin binding domains (section 1.4.3) (106). Based on ATPase domain similarity, SMARCAD1 and its homologues,

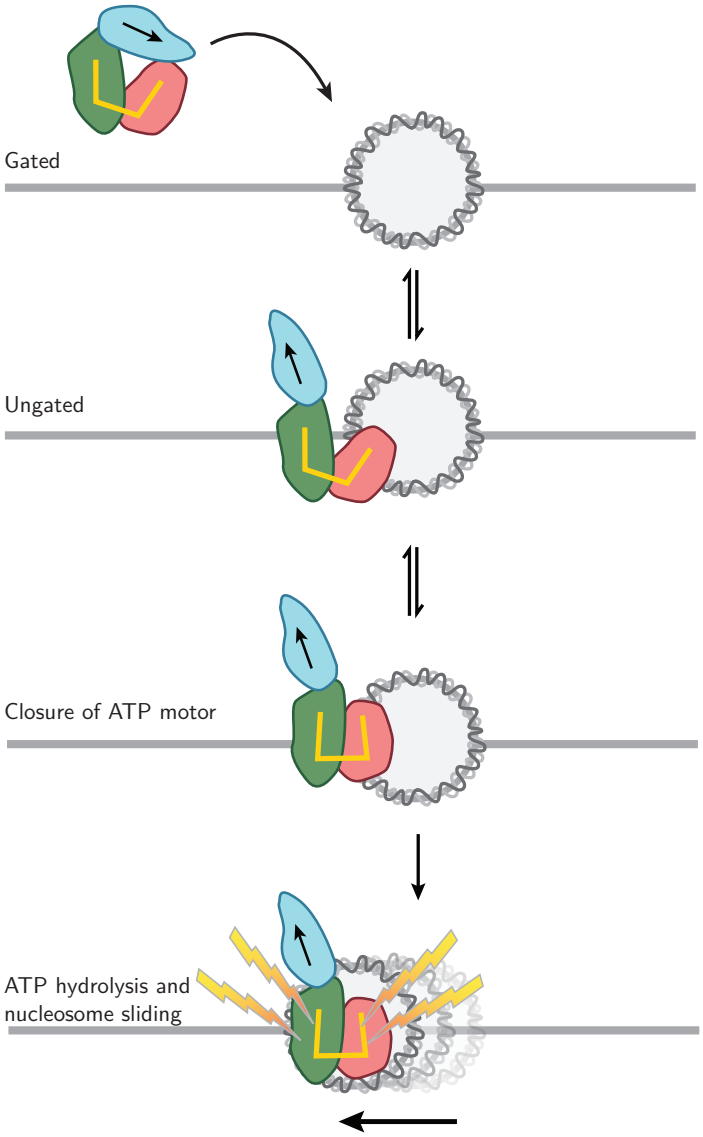


Figure 1.11: Proposed Model for CHD1-mediated Nucleosome Sliding - CHD1 can be either closed (gated) or open (ungated) depending on the position of the chromodomains (cyan). In the ungated form, the DNA and ATP-binding sites are accessible. Adapted from Hauk *et al*, 2010 (98)

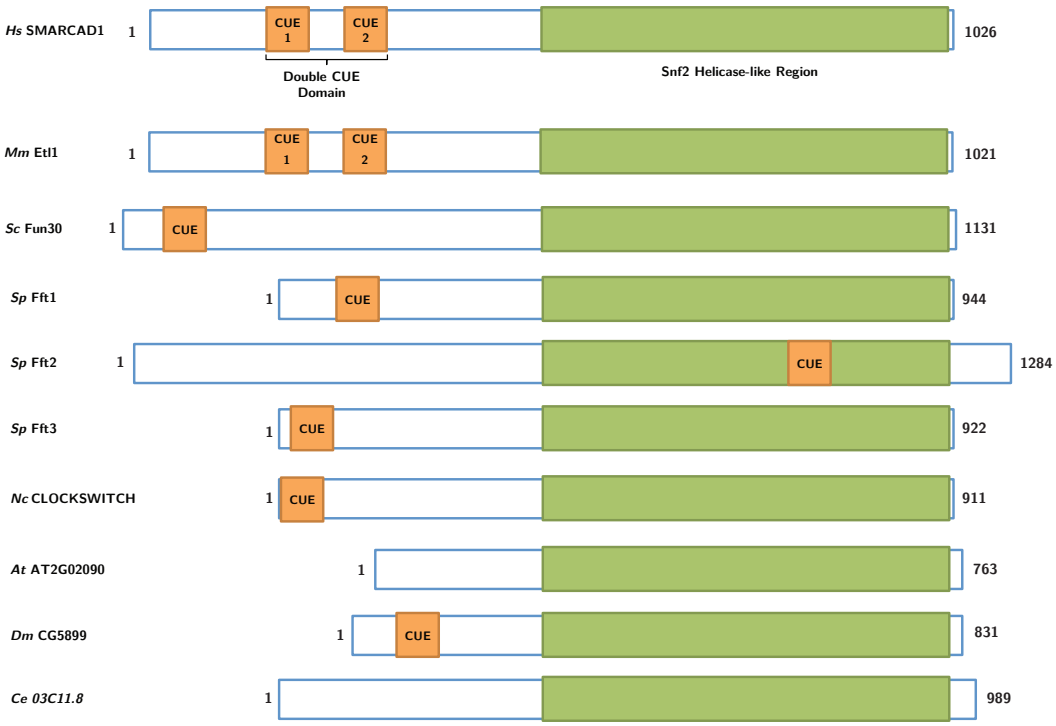


Figure 1.12: Comparison of SMARCAD1 and homologues - Relative positions of CUE domains (orange) and helicase domain (green) in SMARCAD1 and homologous proteins. Human (*Hs*) SMARCAD1, mouse (*Mm*) Etl1, *Saccharomyces cerevisiae* (*Sc*) Fun30, *Schizosaccharomyces pombe* (*Sp*) Fft1, Fft2, Fft3, *Neurospora crassa* (*Nc*) CLOCKSWITCH, *Arabadopsis thaliana* (*At*) AT2G02090, *Drosophila melanogaster* (*Dm*) CG5899 and *Caenorhabditis elegans* (*Ce*) M03C11.8. Adapted from Neves-Costa, 2009 (105).

such as Fun30 (budding yeast) and Fft1, Fft2, Fft3 (Fission yeast Fun Thirty), mouse ETL1, *Neurospora crassa* CLOCKSWITCH (107), *Arabadopsis thaliana* AT2G02090, *Drosophila melanogaster* CG5899 and *Caenorhabditis elegans* M03C11.8, constitute a conserved subfamily of Snf2 helicase proteins (85, 105). The subfamily is associated with gene silencing and is characterised by a common Snf2 ATPase domain, generally found in the presence of putative-CUE motifs. Subfamily members are the only known Snf2 family proteins with a CUE domain.

FUN30 is involved with telomeric silencing in budding yeast and is capable of ATP-dependent nucleosome sliding (105). Similarly, fission yeast homologue Fft3 has a known role in both telomeric and centromeric silencing and is linked to histone deacetylation (105).

SMARCAD1 is known to interact with PCNA, a DNA replication fork factor, and is targeted to sites of replication (108). SMARCAD1 also directly binds KAP1, a co-repressor involved with transcriptional control (108). Functionally, SMARCAD1 regulates global histone acetylation with a suggested role in histone deacetylation after DNA replication by interacting with class I HDACs (108).

Unlike many other classes of chromatin remodelers, SMARCAD1 does not contain multiple different types of accessory domains. Instead, only two CUE domains have been identified alongside the helicase domain. This suggests a key role for the CUE

domains in determining how SMARCAD1 carries out its helicase activity and how it is targeted.

1.3.1 Isoforms

One recent study has identified a skin-specific isoform of SMARCAD1 (109). While the longer, full-length, isoform of SMARCAD1 is expressed ubiquitously at a low level, a shorter isoform is expressed mainly in skin fibroblasts, keratinocytes and the oesophagus. This shorter isoform is a splicing variant that yields a protein product consisting of approximately the 600 C-terminal residues, largely corresponding to the helicase domain and conspicuously lacking CUE domains (figure 1.13) (109). Interestingly, this shortened isoform was discovered by a team investigating the genetic causes of a adermatoglyphia (absence of fingerprints) and suggests a role for short isoform SMARCAD1 in dematoglyph / fingerprint development (109).

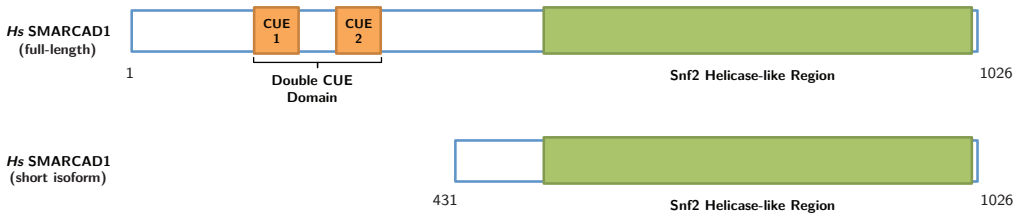


Figure 1.13: Comparison of different isoforms of SMARCAD1 - Two different isoforms of SMARCAD1 exist. The short isoform lacks the double CUE domains and is expressed mainly in the skin. A mutation causing aberrant splicing of the short isoform and a resultant decrease in RNA stability has been linked to the absence of fingerprints (109).

1.3.2 Function and Binding Partners

Knock-down studies show an increase in global acetylation of histones H3 and H4, generally associated with 'open chromatin', and a reduction of methyllysine at H3K9me, a heterochromatic marker, when SMARCAD1 is depleted in HeLa cells. Association between chromatin and key transcriptional silencing factors, heterochromatin protein 1 (HP1), histone deacetylase 1 (HDAC1) and the co-repressor KAP1 is diminished with reduced SMARCAD1 levels (110). Additionally, SMARCAD1 has been shown to be tightly associated with DNA replication sites during S-phase of the cell cycle (DNA synthesis phase) (108, 110). These observations indicate a role for SMARCAD1 in silencing by maintaining / promoting heterochromatin formation after DNA replication. SMARCAD1 promotes histone deacetylation of newly deposited histones, which allows for H3K9 methylation and recruitment of HP1. HP1 encourages heterochromatin formation by linking nearby nucleosomes together, recruiting KAP1 and H3K9 methyltransferase, which propagates the H3K9me heterochromatic mark (108, 110).

SMARCAD1 predominantly interacts with proteins involved with gene regulation, heterochromatin formation, DNA replication or DNA repair. Major binding partners HDAC1, HDAC2 and histone H3K9 methyltransferase G9a all catalyse modifications that are depleted in knock-down SMARCAD1, which implies that SMARCAD1 is in-

volved with the recruitment of these factors in order for them to catalyse the formation of particular histone post-translational modifications and propagate heterochromatin (110). Recently, involvement with coordinating a response to DNA damage has become more fully understood. SMARCAD1 is recruited to double-stranded DNA breaks and yeast homologue FUN30 utilises ATP activity to promote protein-mediated resection of double-strand DNA breaks (104).

1.3.2.1 KAP1

KAP1 (KRAB (Krüppel associated box)-associated protein 1, also known as TRIM28 and TIF1-β) is a component of the SMARCAD1 complex, which, when depleted, also appears to prevent global histone deacetylation and H3K9 trimethylation. This suggests that KAP1 and SMARCAD1 work together to propagate post-translational modifications of newly incorporated histones with a scaffold role proposed for KAP1 due to its many binding partners and protein binding domains (figure 1.14) (108, 111, 112). KAP1, has been associated with all areas of SMARCAD1 function; DNA replication, DNA repair and heterochromatin formation, which, when considered alongside the discovery that it pulls-down in stoichiometric amounts with SMARCAD1 identifies it as the major binding partner of a dynamic SMARCAD1 functional complex (110).

In addition to an array of binding partners, KAP1 can be covalently modified. Five lysine residues (K554, K676, K750, K779 and K804), four of which are located

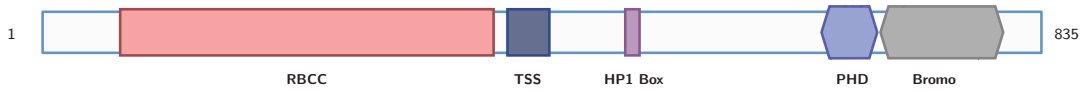


Figure 1.14: Schematic of KAP1 Domain Architecture - KAP1, also known as TIF1β and TRIM28, is a member of the TIF1 family, which all share a similar domain structure; an N-terminal tripartite motif (TRIM), for high affinity protein-protein interactions and oligomerisation, which contains an RBCC (Ring (really interesting new gene) finger, two B-box zinc fingers and a coiled coil) domain, a TIF1 signature sequence (TSS) domain, which is a 25-amino acid tryptophan and phenylalanine-rich sequence, and a C-terminal PHD and bromodomain. KAP1, also has an HP1-binding domain, which is only present in two other TIF1 proteins (TIF1α and TIF1δ). Human KAP1 is *ca* 90 kDa in mass and 835 residues long (113).

in the C-terminal PHD/Bromo tandem domain, have been identified as SUMOylation targets, a ubiquitin-like modification (section 1.4.2) (114, 115, 116). KAP1 (and possibly SMARCAD1 (104, 117)) undergo ATM (ataxia telangiectasia mutated)-mediated phosphorylation, which causes chromatin relaxation in order to allow heterochromatin repair (110, 113, 118). The kinase action of ATM is a response to DNA damage causing phosphorylation of KAP1 and a resultant loss of SUMO modification from KAP1. These markers may be important for switching between the heterochromatic maintenance and DNA repair functions of KAP1. SUMOylated KAP1 is involved with transcriptional repression, whereas phosphorylated KAP1 is involved with DNA repair of double-stranded DNA breaks (110, 113).

1.3.2.2 PCNA

From FRET analysis and GST-pulldown results, SMARCAD1 has been shown to physically interact with PCNA (proliferating cell nuclear antigen), an essential DNA clamp at the replication fork (108). Mutagenesis studies of SMARCAD1 show that the SMARCAD1-PCNA interaction does not depend on conserved PCNA interacting protein (PIP) boxes (110, 119), which is the mode of binding for other chromatin remodeller subunits such as WSTF (Williams syndrome transcription factor) and many other PCNA-binding proteins (110, 120). This leaves a question as to what part(s) of SMARCAD1 interacts with PCNA.

A recently published study shows that flanking regions either side of the SMARCAD1 double CUE domain can individually co-localise to some extent with PCNA in HeLa cells. Full-length SMARCAD1 (aa1-1028) is tightly associated with the location of PCNA in the nucleus whereas expressed protein constructs of SMARCAD1 sequence aa1-157 and aa293-512 both with Venus tag, a variant of yellow fluorescent protein, are slightly enriched at sites of PCNA although also present throughout the rest of the nucleus suggesting that neither stretch of SMARCAD1 sequence is capable of co-localising SMARCAD1 with PCNA alone (note: the double CUE domain is approximately aa150-295) (110).

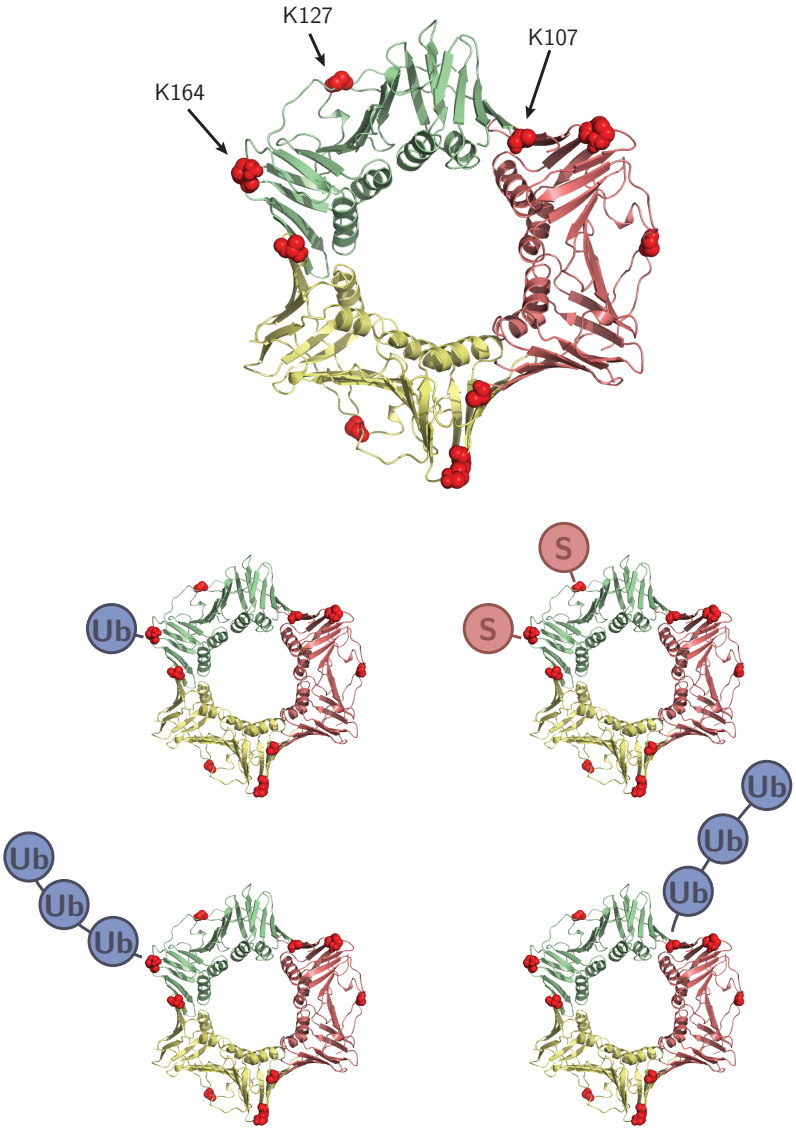


Figure 1.15: Ubiquitin-related PCNA Modifications - Three lysine residues, K107, K127 and K164 (shown as red spheres) on each PCNA monomer can be ligated with either monoubiquitin (Ub), polyubiquitin or SUMO (S) marks. The position, extent of ubiquitination / sumoylation and linkages between ubiquitin monomer are associated with different functions. PCNA K164 monoubiquitination and K164 polyubiquitination are involved with promoting different responses to DNA repair. PCNA K107 ubiquitination is involved with recognition of damage to nascent DNA strands during replication and PCNA SUMOylation at either K164, K127 or is thought to be involved in regulating normal DNA replication (121, 122)

PCNA is a homotrimer of three *ca* 30 kDa subunits (123). Multiple lysine residues on each PCNA subunit can be post-translationally modified with ubiquitin markers and, as with KAP1, SUMOylated. These are associated with different functional outcomes depending on the modifications (figure 1.15) (121, 124). Given that SMARCAD1 and PCNA demonstrate a direct interaction and that PCNA is significantly modified with ubiquitin or a ubiquitin-like modifier, it may be that the CUE domains of SMARCAD1 are involved recognising these marks.

1.4 Ubiquitin Recognition and CUE Domains

1.4.1 Ubiquitin Structure and Function

Ubiquitin (figure 1.16) is a well-studied small (8.6 kDa) regulatory protein found in almost every eukaryotic cell. It is 76-amino acids long and either present as free protein or covalently linked to a large variety of proteins (125, 126). Ubiquitin was first shown to be covalently linked to a protein, histone H2A, in 1977 (127). In addition to free protein and monoubiquitin tags, ubiquitin can be linked together to produce multiubiquitin chains, which can branch off from the lysine side-chain of a substrate protein through an isopeptide bond (128, 129, 130). Chains of ubiquitin can be linked together by different lysine residues, generally either Lys-48 or Lys-63 or, less commonly, via one of the other five ubiquitin Lys residues or linearly head-to-head (figure 1.17). Monoubiquitin and

these different forms of polyubiquitin are associated with numerous different functions such as destruction of damaged proteins, gene transcription, cell-cycle progression, virus budding, apoptosis and DNA repair (131, 132, 133).

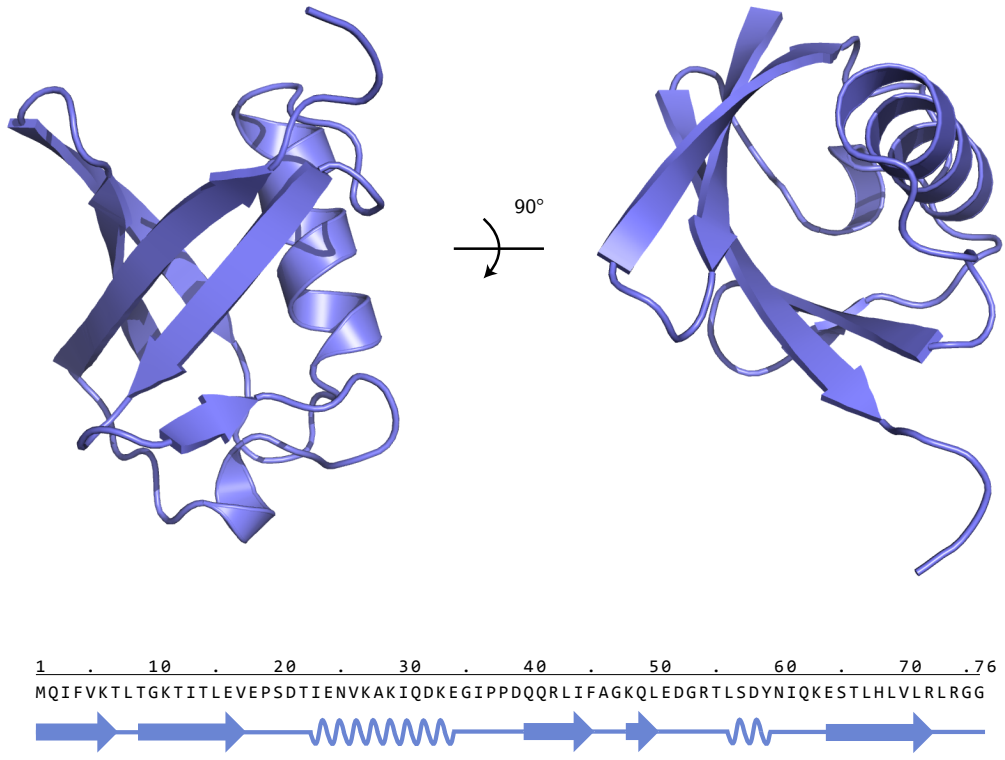


Figure 1.16: Structure of Ubiquitin - 1.8 Å crystal structure of ubiquitin (PDB 1UBQ; rendered in Pymol (22)) with complete sequence and secondary structure below (134). Ubiquitin has seven Lys residues (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 and Lys63), which allow for a range of different linkages.

Covalent modification of a protein with ubiquitin is a multi-step process involving ATP-dependent ubiquitin-activating enzyme (E1), transfer of the activated ubiquitin to a ubiquitin-conjugating enzyme (E2) and subsequent transfer to a specific substrate

with the help of ubiquitin ligases / accessory proteins (E3). There are hundreds of related but different E2/E3 complexes, which are responsible for conferring selectivity of the ubiquitinated proteins (129, 135). Ubiquitin is a dynamic mark that can be removed from substrate proteins by deubiquitinating enzymes (DUBs), of which almost 100 have been identified in the human genome (136, 137, 138).

One long-established function of ubiquitin is its role in mediating protein destruction by the proteasome (139). Tagged proteins, with at least tetra-ubiquitin Lys48-linked chains can be targeted to the proteasome and recognised by a range of ubiquitin binding motifs in various receptor proteins located in the cap subunit (19S) of the proteasome (129, 140, 141, 142, 143, 144, 145). More recently, the importance of ubiquitin in other cellular processes has become increasingly apparent (136, 146).

Monoubiquitin is known to regulate chromatin remodeling, DNA repair, endocytosis and meiosis (133, 149, 150). There is a wide range of different types of polyubiqutin and their function varies. As already discussed, Lys48-linked polyubiquitin is normally associated with the proteasome pathway for protein destruction. Lysine 63 (K63-G76) linked polyubiquitin is associated with more diverse range of functions such as DNA damage repair, protein trafficking, activation of kinases / phosphatases and regulation of tumour formation (136, 137, 151, 152). Linear (head-to-tail) polyubiquitin chains, a relatively recent discovery that don't use Lysine isopeptide bonds, play and essential

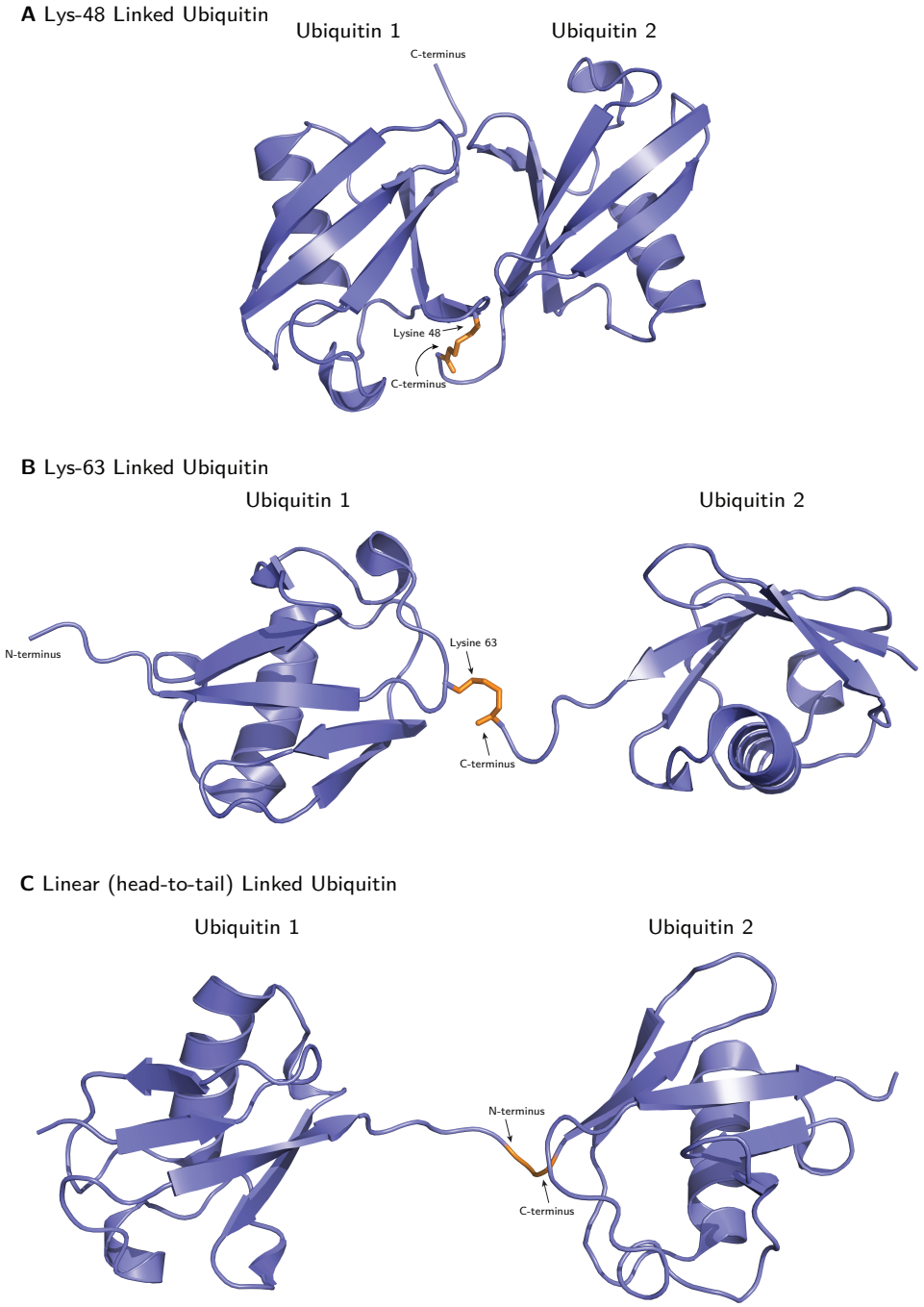


Figure 1.17: Crystal structures of three forms of Di-ubiquitin - **A** Lys48-Linked Di-ubiquitin (PDB: 1AAR) (147). **B** Lys63-Linked Di-ubiquitin (PDB: 2J5F) (148). **C** Linear (head-to-tail) linked Di-ubiquitin (PDB: 2W9N) (148). In all, 'Ubiquitin 2' is the proximal ubiquitin and the isopeptide / peptide bond between ubiquitin subunits is shown in orange.

role in the signalling of nuclear factor-κB (NF-κB), a complex involved in regulating immune response (149, 153, 154). These latter two forms of polyubiquitin chains are structurally more extended conformations, with no contact between sequential ubiquitin moieties (148, 149).

1.4.2 Ubiquitin-Like Proteins

In addition to protein modification with ubiquitin, there is growing appreciation of the role played by structurally related ubiquitin-like proteins (UBLs). SUMO (Small Ubiquitin-like MOdifier) is a family of proteins that are related to ubiquitin. SUMO proteins share the same β-grasp fold as ubiquitin (figure 1.18), despite only having 16 - 20% sequence identity (figure 1.19), and can be also be ligated to lysine residues in various proteins (155). Most of the identified modification substrate proteins for SUMO proteins are involved with chromatin organisation, transcription and RNA metabolism (156). Nedd8 (neuronal-precursor-cell-expressed developmentally down-regulated protein-8) is another ubiquitin-like protein with very similar structure and is involved with regulating a subset of ubiquitin E3 ligases (157).

1.4.3 CUE Domains

CUE motifs, named after yeast protein Cue1 (couples ubiquitin conjugation to ER degradation), are one of at least twenty different types of ubiquitin binding domains

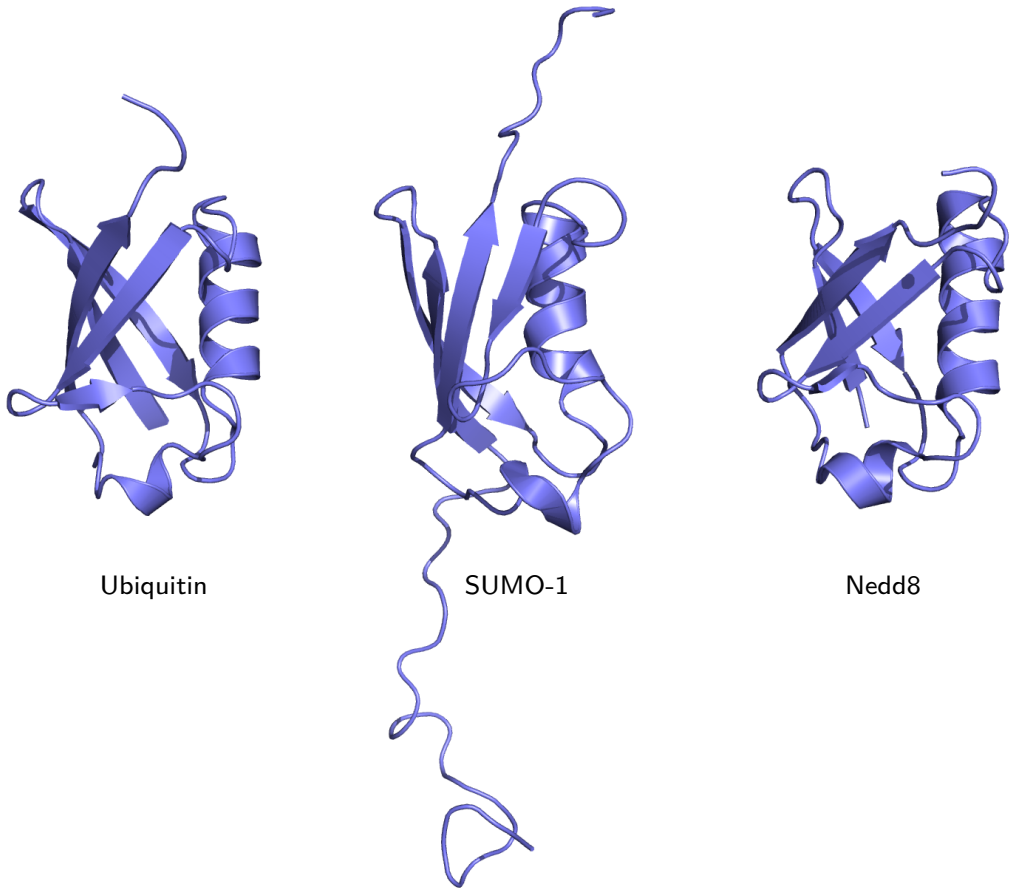


Figure 1.18: Structure of Ubiquitin-like Proteins - UBLs share a β-grasp fold. Ubiquitin PDB: 1UBQ (158), SUMO-1 PDB: 1A5R (159) and Nedd8 PDB: 2KO3 (160).

```
Ubiquitin          1 ----------MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDK 33
SUMO1 (P63165)     1 MSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQR 54
SUMO2 (P61956)     1 MADE--KP-KEGVKTENN-DHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCER 50
SUMO3 (P55854)     1 MSEE--KP-KEGVKTEN--DHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCER 49
SUMO4 (Q6EEV6)     1 MANE--KP-TEEVKTENN-NHINLKVAGQDGSVVQFKIKRQTPLSKLMKAYCEP 50

Ubiquitin         34 EGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG---------- 76
SUMO1 (P63165)    55 QGVPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGGHSTV------- 101
SUMO2 (P61956)    51 QGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGVY---------- 95
SUMO3 (P55854)    50 QGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGVPESSLAGHSF 103
SUMO4 (Q6EEV6)    51 RGLSMKQIRFRFGGQPISGTDKPAQLEMEDEDTIDVFQQPTGGVY---------- 95
```

Figure 1.19: Alignment of SUMO Proteins - Sequence alignment of ubiquitin and human SUMO proteins using ClustalW pairwise alignment. Uniprot codes shown in parenthesis and colouring based on amino acid similarity (Jalview2; ClustalX colour scheme). (161, 162, 163, 164).

(125, 149, 165). CUE motifs are structurally related to ubiquitin associated domains (UBA) and are small helical motifs approximately 50 amino acids long, generally forming a three-helical bundle with conserved (M/L)FP and LL amino acid sequences at the ends of helices that interact with ubiquitin (166). The affinities and modes of ubiquitin binding of CUE motifs from different proteins vary (figure 1.20) although affinity is generally low, as with other classes of ubiquitin binding domains (167). Some CUE motifs are capable of binding polyubiquitin chains in addition to monoubiquitin although the functional significance of this is yet to be determined (125, 168).

As discussed above, accessory domains such as PHD and bromodomains are often found in tandem as part of histone associated proteins involved with gene silencing (171). Mammalian SMARCAD1 is unique as the only protein with tandem CUE motifs that has been identified as a chromatin remodelling factor. The yeast protein Cue2 is another example of a protein with tandem CUE motifs. Pull-down experiments demonstrate that each CUE motif of Cue2 is capable of binding ubiquitin on its own and there is no evidence for cooperative binding when the complete double CUE domain from Cue2 is used (167). In contrast, Vps9p (vacuolar protein sorting) protein only contains a single CUE motif but forms a domain-swapped dimer to bind ubiquitin with relatively high affinity (169, 170). Prior to this investigation, no structure was known for a complete tandem CUE domain or whether the CUE domains of SMARCAD1

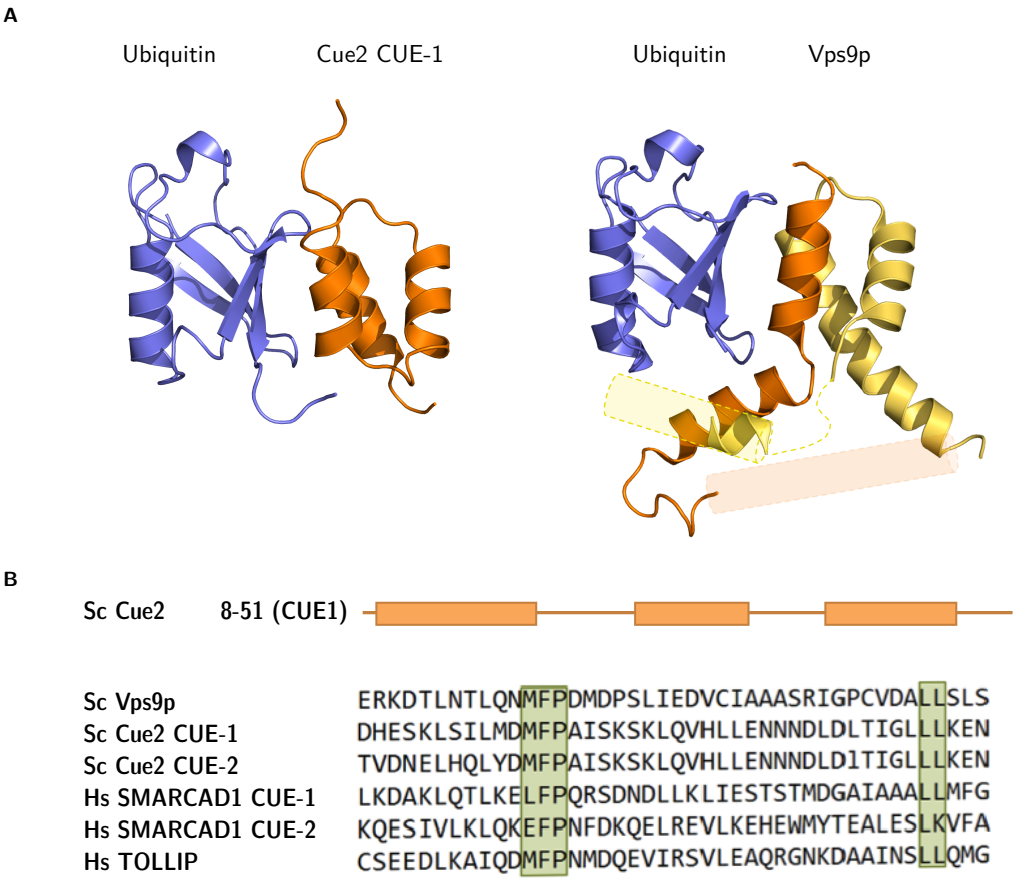


Figure 1.20: CUE Domains - **A** Examples of two different ubiquitin binding modes by CUE domains. The left image is the first CUE motif from the yeast protein Cue2, which adopts a more common CUE conformation as a helical bundle and binds monoubiquitin as a monomer (PDB: 1OTR) with a KD of 155µM (167). The right image is a model based on a solved structure of a domain-swapped dimer of the CUE motif from the yeast protein Vps9p (169) (PDB: 1P3Q). Unresolved helices drawn based on the apo structure. Vps9p binds monoubiquitin with a KD of 20µM (170). **B** Sequence alignment of selected CUE motifs (170). Highly conserved M/LFP and LL amino acid motifs, which are characteristic of CUE domains and known to be involved with ubiquitin binding (166), are highlighted in green and located at the end of the first and third helices of each CUE motif. Notably, the SMARCAD1 CUE-2 motif has modified versions of these short motifs; EFP and LK instead of more conserved MFP and LL. This may be significant for whether it is a canonical ubiquitin binder.

interacted with one another, bound ubiquitin or acted cooperatively.

1.4.4 Other Ubiquitin Binding Domains

In addition to CUE domains, more than 20 types of ubiquitin binding domains (UBD) with varying binding specificities have been identified. UBDs typically interact with ubiquitin's hydrophobic β-sheet surface although there are examples of UBDs using alternative binding surfaces on ubiquitin (172, 173, 174). Figure 1.21 describes the different binding surfaces on ubiquitin and figures 1.22 and 1.23 illustrate these with several examples.

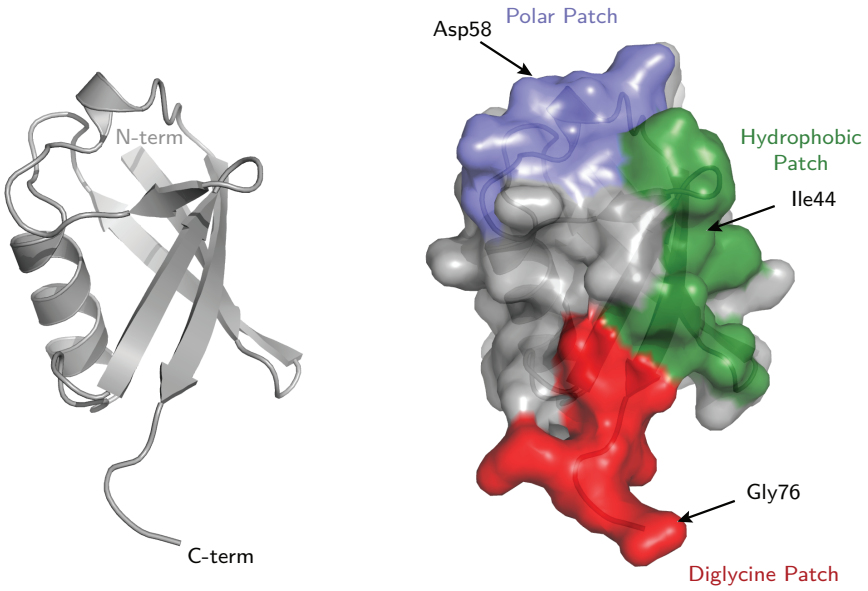


Figure 1.21: Major Ubiquitin Binding Patches - Surface representation of ubiquitin (PDB: 1UBQ) shown on the right, orientated as the cartoon representation on the left. Colours indicate the major binding patches. The majority of UBDs interact with the hydrophobic patch, green, centred around ubiquitin Ile 44. Other significant binding sites include the polar patch, blue, around Asp 58 and the diglycine C-terminal motif, red. Figure based on Hurley *et al* 2005 (175).

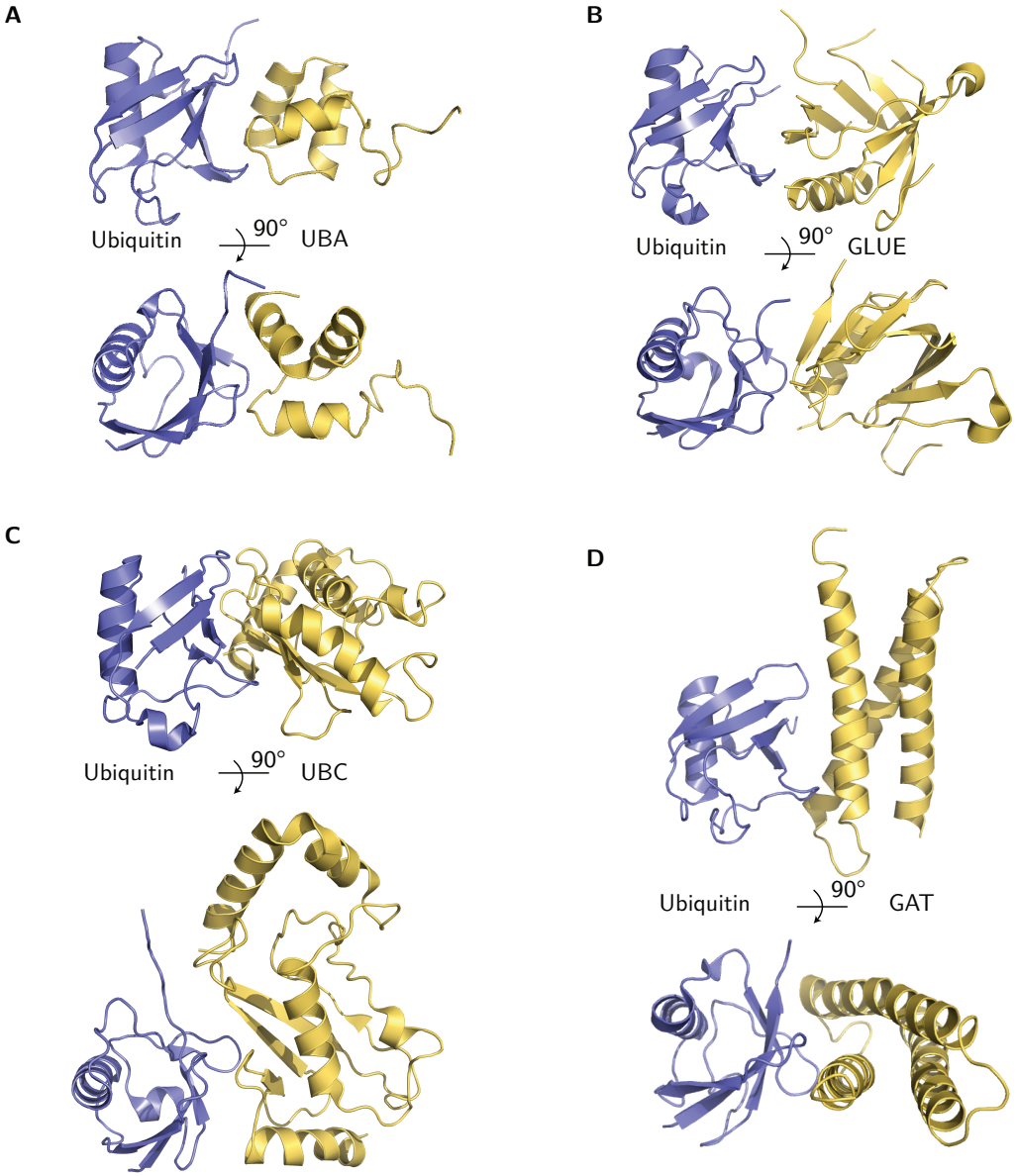


Figure 1.22: Ubiquitin Recognition Domains at the Hydrophobic Patch - A range of different folds are capable of binding ubiquitin. All the examples shown here bind along ubiquitin on a common β-strand surface but other modes of binding have been identified (172, 176). **A** Ubiquitin (purple) complexed with ubiquitin-associated (UBA) domain from protein linking IAP with cytoskeleton 1 (PLIC1; yellow). PDB: 2JY6 (149, 177). **B** Ubiquitin (purple) complexed with Eap45 GLUE domain (yellow) - a Pleckstrin homology (PH) fold. PDB: 2DX5 (178). **C** Ubiquitin (purple) complexed with UBC domain from E2 ubiquitin-conjugating enzyme UBCH5C (yellow). PDB: 2FUH (149, 179). **D** Ubiquitin (purple) complexed with GGA3 GAT domain. PDB: 1YD8 (175).

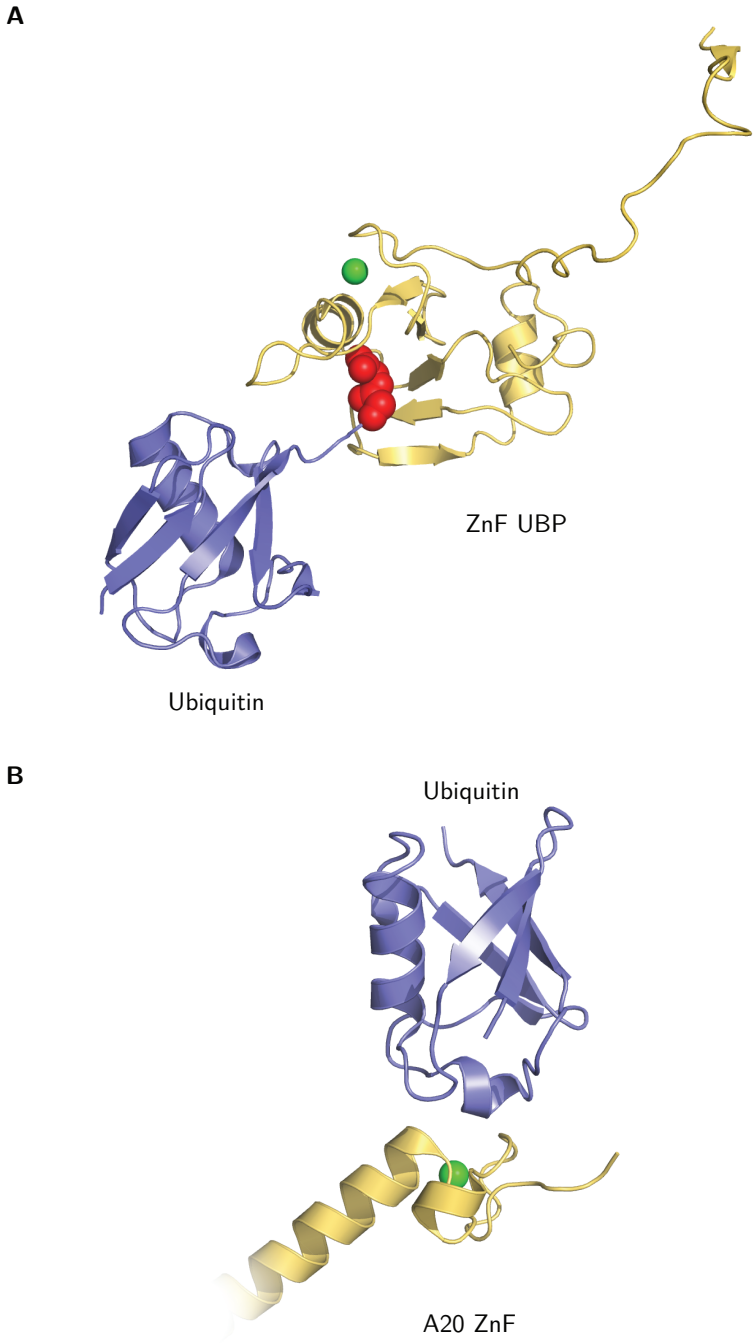


Figure 1.23: Ubiquitin Recognition Domains at Additional Binding Interfaces - **A** Zinc-finger ubiquitin binding domain (ZnF UBP) from deubiquitinating enzyme isopeptidase T (IsoT / USP5) binds tightly to the diglycine C-terminal tail on ubiquitin (red spheres). PDB: 2G45 (180).**B** A20 ZnF domain from Rabex 5 interacts with the mostly polar patch of ubiquitin. Rabex-5 (a mammalian homologue of Vps9) binds ubiquitin via an A20 ZnF domain connected to a MIU (motif interacting with ubiquitin - a single helix which also binds ubiquitin) instead of a CUE domain, as with Vps9 (181). PDB: 2FIF. Zinc is shown as a green sphere for both structures.

Four different groups of UBDs have been identified. The majority of UBDs belong to a group of α-helical domains. This includes UBA (ubiquitin associated) domains, which are structurally related to CUE domains with a shared three-helix bundle. GAT (GGA (Golgi-localized, gamma-ear-containing, ADP-ribosylation-factor-binding protein) and TOM (target of Myb)) are also helical UBDs (125, 125, 149, 168). The second most prevalent group of UBDs are those containing zinc fingers (ZnFs), such as A20 ZnF domains and ZnF UBP (ubiquitin-specific processing protease). PH (Plekstrin homology) domains and UBC (ubiquitin-conjugating)-like domains have also been reported as ubiquitin binding domains (figure 1.22 **B** and **C**) (173).

In addition to the variety of folds and binding patch targeted by UBDs a variety of different modes of binding have been identified. For example, individual NZF (zinc finger) domains from both TAB2 and TAB3 bind preferentially to Lys63-linked ubiquitin chains (182, 183). Examples of cooperativity have also been proposed in the case of Rap80, where its two UIM (ubiquitin-interacting motif) domains bind to Lys63-linked diubiquitin with relatively high affinity (173, 184).

1.5 Aims of this Thesis

SMARCAD1 has emerged as a poorly understood yet central actor in the key cellular processes of DNA replication, DNA repair and gene regulation / heterochromatic main-

tenance. Developing a fuller picture of SMARCAD1's involvement with these functions is important for our complete understanding of the mechanisms.

SMARCAD1 has an unusual composition compared to other ATP-dependent chromatin remodellers. Most contain several modular binding domains, particularly those targeting histone modifications, whereas the only recognised binding module in SMARCAD1 alongside the ATPase subunit is the double CUE domain. This presents the prospect that the double CUE domain in SMARCAD1 is key for directing the function of SMARCAD1 by binding to target proteins.

Structural characterisation of the double CUE domain

Tandem CUE domains are relatively rare and no structure of a double CUE domain has been determined. The oligomeric characterisation, protein structure and nature of the linker region will be discussed with reference to the differing examples in the literature.

What are the binding targets for the SMARCAD1 double CUE domain?

CUE domains are classified by their affinity for ubiquitin and several of SMARCAD1's binding partners are known to be subject to ubiquitination / SUMOylation. As ubiquitin and SUMO are closely related, it was considered that many of SMARCAD1's binding interactions may occur partly through ubiquitin or SUMO

marks with specific recognition by the double CUE domain. Insights into the binding of SMARCAD1 to major binding partners KAP1 and PCNA will also be presented.

Chapter 2

Materials and Methods

Experiments and protocols were carried out at 20°C using chemicals obtained from Sigma-Aldrich unless otherwise stated.

2.1 DNA Handling and Subcloning Techniques

Broadly, the principles of the molecular biology procedures used are as described in *Molecular Cloning: A Laboratory Manual* (185). Specific protocols have been optimised and described as below.

Two different length constructs (dCUE-long and dCUE-short) were designed, and cloned into *E. coli* expression vector pET-DUET-1 (Novagen) so as to produce two His_6-tagged SMARCAD1 Double CUE domains proteins of different lengths. Plasmid DNA of glutathione S-transferase (GST)-tagged N-terminal region of SMARCAD1 (amino acid sequence 113-367) was used as a template for polymerase chain reaction (PCR) amplification with specific primers (section 2.1.7). The PCR products were

then ligated into sub-cloning vector pCR 2.1 (section 2.1.9), amplified in XL1-Blue cells (section 2.1.1), purified and the insert ligated into the first multiple cloning site (MCS1) pET-DUET-1 (sections 2.1.11 and 2.1.12). These were transformed into the protein expression strain of *Escherichia coli* Rosetta (Merck Biosciences).

2.1.1 Plasmid DNA Amplification

Plasmid DNA was amplified by either transforming purified plasmid DNA into *E. coli* and growing bacterial cultures for DNA purification, or, simply, growing glycerol stocks of already transformed cells.

In order to produce plasmid DNA (e.g. prepare additional template DNA for PCR cloning reactions), the plasmid DNA was transformed into XL1-Blue competent cells (Stratagene) by heat-shock (section 2.1.2). The plasmid used for DNA template was a pGEX-4T1 (Novagen) based construct 'GST-SMARCAD1 113-367' donated by Patrick Varga-Weisz (Babraham Institute)(108). Heat-shocked cells were then plated out on an lysogeny broth (LB) -agar plate with suitable antibiotic and left overnight at 37°C. A single colony was picked and grown in LB with antibiotic (5 ml; 100 mg/ml ampicillin) for *ca.* 8 hours. Plasmid DNA was purified by Miniprep (Qiagen) (section 2.1.3). Plasmid DNA constructs were verified by sequencing (section 2.1.4) as necessary.

For constructs where stored glycerol stocks of cells were available, additional plasmid

DNA was grown from the glycerol stock. These stocks samples were produced from single colonies (section 2.1.5) and were grown up by picking a small swab of cell stock and growing it in an 5 ml LB culture overnight at 37°C with appropriate antibiotic.

2.1.2 Transformation of Bacteria by Heat-Shock

Depending on whether the bacteria were being used for producing DNA or expressing protein, different strains of *E. coli* were used. XL1-Blue competent cells (Stratagene) were selected as a host strain during cloning and amplification of plasmid DNA. Rosetta cells (Novagen) are derivatives of BL21 *E. coli* and were selected as they offer enhanced expression of eukaryote proteins by supplying tRNAs for rare bacterial codons (186).

100 µl aliquots of chemically-competent cells were thawed on ice. Approximately 50 ng pure plasmid DNA (1 µl), or 10-20 µl from a ligation reaction, was incubated with 100 µl cells on ice for 20 minutes, 42°C water bath for 30 seconds, ice for 2 minutes then 900 µl of LB (no antibiotic) was added and the mixture further incubated at 37°C for 30 minutes.

2.1.3 Plasmid DNA Purification

Successfully transformed *E. coli* colonies were cultured overnight in 5 ml LB media containing an appropriate selective antibiotic (e.g. ampicillin). The *E. coli* cultures were pelleted (4000 x g for 30 minutes) and plasmid DNA purified from the cell pellets

using the standard procedures associated with the QIAprep Miniprep system (Qiagen). This involved lysing the cells under alkaline conditions; removal of genomic DNA, proteins and lipids through precipitation under neutral conditions; adsorption of plasmid DNA from the remaining supernatant to a silica-gel membrane under high salt conditions while other contaminants were washed away. Salt was then removed by washing the column with 80% (v/v) ethanol. Purified plasmid DNA was eluted with distilled, deionised water and quantified using nanodrop (section 2.1.6).

2.1.4 DNA Sequencing

The DNA sequence of plasmids was confirmed commercially by Geneservice, Oxford using an automated Sanger process (187). Sequencing results were analysed by eye with the aid of *Vector NTI* (188) to identify any cloning errors or mutations. Primers pETUpstream and DuetDOWN1 were used for sequence verification of the final dCUE-long and dCUE-short plasmids.

2.1.5 Glycerol Stocks

Transformed *E. coli* cultures were stored as glycerol stocks in order to simply revive them and grow fresh cultures. Glycerol stocks were prepared from the culture of a single colony of transformed *E. coli*. A single colony was picked from a LB-agar plate and grown overnight in LB media with appropriate antibiotics. 0.5 ml of the overnight

culture was added to 0.5 ml of 80% sterile glycerol in a sterile screw cap freezer tube and stored at -80°C.

2.1.6 DNA Quantitation

DNA concentration and purity was determined using a Nanodrop 1000 spectrophotometer (Thermo Scientific) and corresponding software package. The absorbance of the DNA solution was measured at 260 and 280 nm. Ratios for 260/280 and 230/280 were calculated to give an indication of purity as DNA absorbs UV light at 260 and 280 nm, whereas aromatic proteins (contaminants) absorb significantly at 280nm. A ratio greater than 1.8 was considered pure. Nucleic acid concentration (c) was calculated automatically from the absorbance at 260 nm using the Beer-Lambert law: $A = \varepsilon lc$ (A: absorbance, ε is the extinction coefficient and l is the pathlength).

2.1.7 Polymerase Chain Reaction

In order to subclone the coding sequences for both the long and short SMARCAD1 Double CUE constructs, the coding regions were amplified from template DNA (plasmid of GST-SMARCAD1 113-367) using the Polymerase Chain Reaction (PCR).

This involved the use of custom forward and reverse primers, template DNA and a master mix of active Taq DNA polymerase, free dNTPs and appropriate buffer, which was then subjected to automated PCR thermo cycling (MJ Research Tetrad PTC-225).

Thermo cycling was carried out as in table 2.1.

	Cycles	Time	Temperature
Denaturation	1	2 minutes	94 °C
Denaturation	25	10 seconds	94 °C
Annealing		60 seconds	50 °C
Elongation		2 minutes	72 °C
Elongation	1	10 minutes	72 °C

Table 2.1: Thermo cycling conditions used for PCR

Typically, reaction mixtures of 25 µl were prepared in multiple from a master mix pipetted into PCR tubes (Thermo Scientific) with a range of primer concentrations to optimise reaction product. Each 25 µl reaction mixture was made in nuclease free water and contained 2.5 µl 10x Standard Taq Reaction Buffer (New England Biolabs), 0.5 µl 10 mM dNTPs (New England Biolabs), 0.5 µl forward primer (10-100 µM), 0.5 µl reverse primer (10-100 µM), template plsmid DNA (*ca.* 0.5 µg 'GST-SMARCAD1 113-367'; section 2.1.1), 0.125 µl Taq DNA polymerase (5000 units / ml; New England Bio). PCR primer oligonucleotides were supplied by Eurogentec as in table 2.2.

Primer Name	Sequence (5'-3')
dCUE Long Forward	CGGGATCCGGAGTCCCAAGGCCTTCCTAC
dCUE Long Reverse	CCTTAAGGCTAACCGACATCAGAACCTGAAT
dCUE Short Forward	CGGGATCCGGATATTTCAGAACTGGAAGA
dCUE Short Reverse	CCTTAAGGCTAGTCTTCTGCAAACACTTTTA

Table 2.2: SMARCAD1 Double CUE Primers - Primers used to clone Human His_6-tagged SMARCAD1 Double CUE domain expression constructs.

2.1.8 Agarose Gel Electrophoresis

DNA samples were mixed with DNA loading buffer (30% glycerol, 0.25% bromophenol blue) and separated on an agarose gel (1 % agarose (w/v) in TAE buffer (40 mM Tris, 20 mM acetate, 1 mM EDTA, pH 8.5) containing 'Sybr Safe' (1X final concentration; Invitrogen). A DNA ladder which ranged from 100 - 12,000bp ('1 Kb Plus DNA Ladder' from Invitrogen) was used to provide size markers. DNA was visualized using an UV transilluminator and, as required, bands of interest were excised for further purification (section: 2.1.10).

2.1.9 Subcloning Step for PCR Products

In order to improve cloning results of successful PCR products, an additional subcloning step was used. PCR products from *Taq* polymerase contain an 3' deoxyadenosine (A) regardless of template. This corresponds with 3' deoxythymidine (T) residues on linearised vector (pCR 2.1) and can be ligated readily. This approach removes the need to enzymatically modify PCR products and was found to achieve successful cloning.

The protocol used was as supplied by the manufacturers with the TA Cloning Kit (Invitrogen) except that ligation was allowed to continue overnight at 4 °C. Typically, 1 µl fresh PCR product, was added to 9 µl buffer containing 50 ng linearised pCR 2.1

subcloning vector and T4 DNA Ligase (4.0 Weiss units).

After ligation, products were amplified by transforming XL1-Blue competent cells, plating them out and growing individual colonies (section 2.1.1). Purified plasmids were digested (section 2.1.11) and the insert ligated into the destination vector.

2.1.10 DNA Gel Extraction and Purification from Agarose Gels

DNA bands in agarose gel were purified using 'QIAquick Gel Extraction Kit' (Qiagen) in accordance with the manufacturers instructions. This involves solubilising the DNA band, adsorbing it to a silica-gel membrane in a column under high salt conditions, washing away contaminants and then eluting purified DNA in distilled, deionised water.

2.1.11 Restriction Digest

DNA digestion was typically conducted in a volume of 10-20 µl as a double digest, with conditions as recommended by the manufacturer (New England Biolabs). Specifically, for a double digest reaction using restriction enzymes EcoRI and BamHI, reactions were carried out in *NEBuffer EcoRI*, with 1X BSA, 1 U enzyme / µg DNA at 37 °C for 2 hours. Successful digestion was confirmed by analysing the resulting fragments by agarose gel electrophoresis.

2.1.12 Ligation

Prior to ligation, linearised vectors were first dephosphorylated using phosphatase to prevent re-circularisation of empty vector. Antartic Phosphatase (New England Biolabs) was used with supplied buffer (50 mM Bis-Tris-Propane-HCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, pH 6.0) and incubated with DNA (10 U phosphatase, 1 µg DNA) for 30 minutes at 37 °C.

Up to 1 µg linearised, dephosphorylated vector was diluted in a 5x DNA dilution buffer (Roche), insert DNA was added to a molar ratio of 1 vector : 3 insert. T4 DNA Ligase (Roche) was added to a final concentration of 5 U together with supplied ligation buffer. The reaction was incubated at room temperature for 15 minutes. Ligated DNA was either used immediately to directly transform competent cells or frozen at 20 °C.

2.2 Protein Production and Purification

2.2.1 Unlabelled Double CUE Expression

All expression was carried out using *E.coli* Rosetta cells (Merck Biosciences) in shaking incubators. Glycerol stocks or single colonies were swabbed and grown in an initial starter culture (LB, 100 µg/ml ampicillin) at 37 °C overnight. Starter cultures were then split across 2l flasks (10 ml starter culture in 1l) of either TB (Terrific Broth) or TB OnEX autoinduction media (Novagen) with ampicillin (100 g/ml) and

grown up at 37 °C. Expression in TB cultures was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) (Melford) when culture OD_{600} = 0.8 and left overnight at 25 °C. TB OnEx cultures were grown for 4 hours and then incubated at 25 °C for 20 hours before harvesting. After expression, cells were pelleted (4000 x g, 4°C, 30 minutes) and either frozen or immediately lysed and protein purified.

2.2.2 Isotopically Labelled Protein Expression and Purification

Uniformly labelled protein was expressed in transformed *E. Coli* as for unlabelled protein (section 2.2.1) except that cultures were grown in M9 minimal media containing ^{15}N-enriched NH_4Cl (CK Gas Products Ltd.) and unenriched or ^{13}C-enriched D-glucose (CortecNet) as sole nitrogen and carbon sources, respectively. Labelled cultures were grown in 500 ml volume in 2l baffled flask to improve aeration and optimise growth. Expression of minimal cultures was induced when OD_{600} = 0.6 with 1 mM IPTG (Melford) at 20 °C and left for 20 hours. Purification was carried out as for unlabelled protein except the final gel filtration buffer was 20 mM sodium phosphate pH 7.0 and 100 mM NaCl.

2.2.3 Purification of SMARCAD1 Double CUE from cell pellets

Both isotopically labelled and unlabelled SMARCAD1 Double CUE (dCUE) were purified in exactly the same manner. Cell pellets were resuspended in lysis buffer (20mM

Tris pH 7.5, 1M NaCl, 0.05% NP-40, 10% glycerol, 2 mM phenylmethanesulfonyl fluoride (PMSF)) with protease inhibitor cocktail tablets (Roche). Cells were lysed by sonication and centrifuged for 45 mins, 48,000 x g, 4°C. His_6-tagged protein was purified from the supernatant using Talon (Co^{2+}) beads (Clontech), washed (lysis buffer with 150 mM NaCl and 5 mM imidazole) and eluted (wash buffer with 200 mM imidizole). Protein was finally purified by size exclusion chromatography using a Superdex 75 26/60 column (GE Healthcare) with gel filtration buffer (20 mM Tris pH 7.5, 150 mM NaCl). Identity of purified proteins was confirmed by acrylamide gel electrophoresis (section 2.2.5), mass spectroscopy (section 2.2.4) and western blots against the His_6-tag (section 2.2.7).

2.2.4 Mass Spectrometry of Purified Proteins

Accurate masses of purified proteins were determined externally using liquid chromatography-electrospray ionisation-mass spectrometry (LC-ESI-MS) by the Oxford Protein Production Facility (OPPF-UK).

2.2.5 Protein Acrylamide Gel Electrophoresis

Sodium dodecyl-sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) separates denatured proteins by molecular weight. Two different electrophoresis systems were generally used: self-poured Bio-Rad Mini-Protean gels (12%) were typically used to

monitor protein-expression and purification, whereas Invitrogen pre-cast 4-12% Bis-Tris NuPage gels were used for pull-down experiments and some western blots, especially when a gradient of acrylamide concentration was desired for improved separation. NuPage gels were used with 1X MOPs buffer (NuPage).

Self-poured gels were prepared by diluting 30% (v/v) 37.5 : 1 acrylamide : bis-acrylamide in 400 mM Tris-HCl pH 8.8, 0.1 % (w/v) SDS, 0.1 % (w/v) ammonium persulphate and 0.1 % (v/v) Tetramethylethylenediamine (TEMED). Poured resolving gels were overlaid with 70% ethanol and allowed to polymerise for at least 15 minutes. The ethanol was poured off and stacking gel added, which comprised 4 % (v/v) acrylamide, 125 mM Tris-HCl pH 6.8, 0.1 % (w/v) SDS, 0.1 % (w/v) ammonium persulphate and 0.1 % (v/v) TEMED, into which a comb (1 mm thick) was placed to generate wells, which was left to polymerise for at least 15 minutes.

Three types of protein marker standards were used. Sigma Aldrich wide range markers (6.5-200 kDa) were used for gels with regular coomassie staining. Invitrogen BenchMark pre-stained (6-180 kDa) or Bio-Rad kaleidoscope standards (10-250 kDa) for western blots.

Protein samples (typically 10 µl, unless otherwise stated) were prepared in Laemmli buffer (62.5 mM Tris-HCl, 2% (v/v) SDS, 10% (v/v) glycerol, 0.02% (w/v) bromophenol blue; pH 6.8) and reduced by the addition of 2% (v/v) β-mercaptoethanol. Samples

were briefly heated at 98 °C and spun before loading. Electrophoresis was performed at 180 V in running buffer (25 mM Tris-base, 192 mM glycine and 0.1 % (w/v) SDS; pH 8). As desired, gels were stained with either Coomassie dye or silver-staining techniques (section 2.2.6) or used for western blot transfer.

2.2.6 Protein Staining

To ubiquitously stain all proteins within a gel, the gel was first washed 3 times for 5 minutes in RO water. The gel was then treated with a Coomassie dye or, for improved sensitivity, stained by silver precipitation.

To stain using Coomassie dye, the gel was submerged in a Coomassie dye solution (H_2O, methanol, acetic acid and Brilliant Blue R-250 in a ratio of 50/10/0.05/39.05 (v/v/v/v)) until fully stained; approximately 15 minutes at room temperature with agitation. The gel was then washed with RO water and immersed in destain solution (H_2O, methanol, and acetic acid in a ratio of 50/40/10 (v/v/v)) until bands could be resolved.

Silver staining was carried out as described in the manufacturers protocol (Bio-Rad Silver Stain Plus). This involved a fixative step to fix the proteins, where the gel was submerged in a fixative enhancer solution for 20 minutes (diluted with water, methanol and ethanol) and a rinsing step prior to final staining with included staining

solutions. The staining step caused visible precipitation of silver in the presence of protein. Development was stopped by quenching acetic acid (5%) once bands were resolved.

2.2.7 Confirmation of Expression by Western Blots

Expression of His-tagged dCUE protein was confirmed by western blotting and probing with an anti-penta-His antibody. Protein samples were prepared and run on a acrylamide gel as described in section 2.2.5 using a His-tagged protein marker (Sigma Aldrich wide range marker 6.5-200 kDa) and pre-stained protein ladder (Invitrogen Benchmark Pre-strained).

Semi-dry transfer with a Bio-Rad Trans-blot SD transfer cell was used to transfer protein from the acrylamide gel onto a nitrocellulose membrane. The stacking gel was removed from the separating gel and discarded. A sandwich of three pieces of Whatman 3 mm chromatography paper soaked in 1x transfer buffer (Tris-base 5.8 g/l, glycine 2.9 g/l, methanol 20% (v/v)) was assembled with a nitrocellulose membrane (GE Healthcare) on top, then the gel stacked on top of that with three more pieces of soaked chromatography paper. The 'sandwich' was placed with the membrane closest to the base of the transfer cell so that current travels through the gel and moves protein with it onto the membrane. Air bubbles were removed prior to transfer by applying

gentle pressure to the stack with a roller. Transfer was carried out at 15 V, 270 mA for 45 minutes.

After transfer was complete, membranes were washed for 30 minutes at room temperature in phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4 \cdot 2\ H_2O$, 2.0 mM KH_2PO_4, pH 7.4) with 0.3% Tween-20, washed twice briefly with milli-Q-water and blocked with blocking buffer (PBS with 0.1% Tween-20 and 5% w/v semifat dry milk powder (Marvel, Premier Foods)) for 1 hr (or overnight at 4°C). The membrane was incubated on a rocker with primary antibody (200 µg / µl Qiagen anti-penta-His, mouse, diluted 1:1000 in blocking buffer) for an hour, washed with PBS-0.1% Tween-20 three times for 5-15 min each (rocking at room temperature). The membrane was then incubated on a rocker for an hour with the secondary antibody (anti-mouse IgG coupled to Horseradish peroxidase, Sigma Aldrich; 1:2000 in blocking buffer), then twice washed briefly with PBS-0.1% Tween-20 before being developed with ECL western blotting detection kit (GE Healthcare) and Amersham Hyperfilm ECL (GE Healthcare) in accordance with instructions.

2.2.8 Recombinant Un-tagged PCNA

2.2.8.1 Expression

An untagged human PCNA expression construct in *E.coli* BL21 cells was obtained from Dr D. Bubeck (University of Oxford) (189). A starter culture (LB, 100 µg/ml

ampicillin) was grown at 37 °C overnight and used to to inoculate a 2l flask (5 ml starter culture in 1l LB) with ampicillin (100 μg/ml) and grown up at 37 °C. Expression was induced with 1.5 mM IPTG when culture OD_{600} = 0.8 and left for 5 hours at 25 °C before harvesting. After expression, cells were pelleted (4000 x g, 4 °C, 30 minutes) and frozen prior to lysis and protein purification.

2.2.8.2 Purification

Cell pellets were resuspended in lysis buffer (20 mM Na PO_4, pH 7.0, 50 mM NaCl, 0.02% Tween 20 with protease inhibitor cocktail tablets (Roche)). Cells were lysed by sonication and centrifuged for 45 mins (45,000 x g, 4 °C). Protein was separated using a HiTrap Q FF column (5ml; GE Healthcare) equilibrated in buffer A (0 mM NaCl, 10mM Na PO_4 pH 7.0). Supernatant was loaded onto the column at 5 ml/min and washed for a further 40-45 ml. Protein was eluted slowly over a linear gradient from buffer A to high salt buffer B (1.5M NaCl, 10mM Na PO_4 pH 7.0). Fractions were collected and analysed by acrylamide gel electrophoresis (section 2.2.5) to identify which fractions contained PCNA. Fractions containing PCNA corresponded to the *ca.* 40% buffer B fractions and were pooled prior to size exclusion chromatography using a Superdex 200 26/60 column (GE Healthcare) with gel filtration buffer (20 mM Na PO_4 pH 7.0, 100 mM NaCl). Fractions containing pure PCNA were identified by SDS-PAGE and pooled and concentrated as needed.

2.3 Crystallisation Trials

Nanolitre-scale crystallisation trials were set-up using the Oxford protein production facility (OPPF) high-throughput crystallisation pipeline with a standard screen of approximately 700 standard reservoir solutions (190, 191) and a range of protein concentrations.

Typically, trials were prepared for the sitting drop vapour diffusion method by mixing 100 nl of protein solution (10 mg/ml - 158 mg/ml) with 100 nl precipitant (190) in a 96-well crystallisation plate. Plates were prepared using an automated pipetting system (Cartesian Technologies). Trials were conducted at both 4°C and 20°C (the latter using a temperature controlled vault and automated imaging system).

Protein surface lysine methylation was carried out as described (192). Co-crystallisation trials were set-up by mixing unlabelled dCUE-long / dCUE-short with ubiquitin (Sigma) at 1:1, 2:1 and 1:2 molar ratios.

2.4 Oligomeric State Determination

Dynamic light scattering (DLS) of SMARCAD1 dCUE was carried out using a DynaPro (Wyatt Technology) with 200 µl sample (1 mg/ml) at both 4 °C and 20 °C. Multiple angle light scattering (MALS) experiments were performed by separating the sample (100 µl protein; 4 mg/ml) over a S75 5/150 column (GE Healthcare) then de-

tecting using a Dawn Heleos II (Wyatt Technology) at 20 °C. Gel filtration results were calibrated using a gel filtration calibration kit (GE Healthcare) for analytical gel filtration.

2.5 NMR Experiments

2.5.1 Sample Preparation

Isotropically-labelled protein samples of the dCUE-short construct were produced as described in section 2.2.2. All nuclear magnetic resonance (NMR) experiments were performed at 20°C with the dCUE-short construct at concentrations between 0.6 and 1.3 mM (except for NMR titration experiments) with a buffer of 20 mM sodium phosphate pH 7.0 and 100 mM NaCl. 5% D_2O was added to the samples for deuterium locking. A minimum of 270 µl sample volume was used with Shigemi NMR micro-tubes (Shigemi Inc.).

2.5.2 Data Collection

Home-built NMR spectrometers located in the Department of Biochemistry, University of Oxford operating at ^1H frequencies of approximately 600 or 750 MHz were used for data collection in addition to a Bruker Avance 500 MHz spectrometer. Magnets for home-built spectrometers were supplied by Oxford Instruments, UK, with custom home-built triple-resonance pulsed-field gradient probe-heads controlled by

GE/OMEGA software.

Acquisition parameters for triple-resonance experiments used to assign the backbone resonances are shown in table 2.3. Table 2.4 details parameters used in ^{15}N-edited experiments, such as those used to determine dynamics or residual dipolar couplings. Parameters used in titration experiments are shown in table 2.5 and table 2.6 details parameters for ^{13}C-edited spectra, which were used to assign side-chain resonances. Raw data was processed using NMRPipe (193) and plotted and analysed using Analysis (CCPN) (194).

2.5.3 Residual Dipolar Coupling Measurements

Residual dipolar couplings (RDCs) were measured from the change in $^{1}J_{\mathrm{NH}}$ scalar coupling values seen in an IPAP (in-phase / antiphase) experiment of a partially aligned sample relative to an isotropic sample (appendix B.3.2.3). An anisotropic sample of 2% w/v $C_{12}E_6$/hexanol was found to give good, stable splitting based on the method by Rückert and Otting (195). ^{15}N-labelled dCUE-short concentration was 690µM in both samples.

Experiment	Spectrometer	Sweep Width			Number of Complex Points			Spectrometer Frequency (MHz)	Centre Reference Chemical Shift (ppm)			Number of Scans
		^{1}H	^{15}N	^{13}C [^{1}H]	^{1}H	^{15}N	^{13}C [^{1}H]		^{1}H	^{15}N	^{13}C [^{1}H]	
HNCA	Bruker 500	5882.353	1241.44	3772.028	512	28	50	500.012	4.8	117.7	53	16
HN(CO)CA	Bruker 500	5882.353	1241.44	3772.028	512	28	50	500.012	4.8	117.7	53	16
CBCANH	Bruker 500	5882.353	1241.44	7795.436	512	28	50	500.012	4.8	117.7	41.5	24
CBCA(CO)NH	Bruker 500	5882.353	1241.44	7795.436	512	28	50	500.012	4.8	117.7	41.5	16
(H)CC(CO)NH	Omega 750	8849.56	1736.11	11363.64	512	30	46	749.954	4.8	118.15	41.77	32
HNCO	Bruker 500	5882.353	1241.44	1886.081	512	28	50	500.012	4.8	117.7	172.5	8
HN(CA)CO	Bruker 500	5882.353	1241.44	1886.081	512	28	50	500.012	4.8	117.7	172.5	16
HBHA(CO)NH	Bruker 500	5882.353	1241.44	[4761.905]	512	28	[89]	500.012	4.8	117.7	4.8	8
$^{1}H,^{15}N$-HSQC	Bruker 500	5882.353	1241.44	-	1024	128	-	500.012	4.8	117.7	-	8

Table 2.3: Acquisition Parameters for Triple Resonance NMR Experiments used in Backbone Assignment

Experiment	Sweep Width			Number of Complex Points			Spectrometer Frequency (MHz)	Centre Reference Chemical Shift (ppm)			Number of Scans	Mixing Time (ms)	Saturation Time (s)
	^{1}H	^{15}N	^{1}H	^{1}H	^{15}N	^{1}H		^{1}H	^{15}N	^{1}H			
$^{1}H,^{15}N$-HSQC	8849.56	1736.11	-	1024	128	-	749.954	4.8	117.7	-	16	-	-
^{15}N-NOESY-HSQC	8849.56	1736.11	8130.08	512	32	128	749.954	4.8	118	4.8	8	150	-
^{15}N-TOCSY-HSQC	8849.56	1736.11	8130.08	512	32	128	749.954	4.8	118	4.8	8	35	-
Heteronuclear NOE	8849.56	1736.11	-	1024	128	-	749.954	4.8	118	-	160	-	4
HSQC (NH_2 selective)	8849.56	1736.11	-	1024	128	-	749.954	4.8	117.7	-	16	-	-
$^{1}H,^{15}N$-HSQC (RDC: ISO)	7042.25	1529.05	-	1024	128	-	598.484	4.8	117.7	-	96	-	-
$^{1}H,^{15}N$-HSQC (RDC: ANISO)	7042.25	1529.05	-	1024	128	-	598.484	4.8	117.7	-	256	-	-

Table 2.4: Acquisition Parameters for ^{15}N-Edited NMR Experiments - RDC ISO and ANISO refer to isotropic or anisotropic (partially aligned) samples based on contents of an alignment medium.

Experiment	Spectrometer	Sweep Width		Number of Complex Points		Spectrometer Frequency (MHz)	Centre Reference Chemical Shift (ppm)		Number of Scans
		^{1}H	^{15}N	^{1}H	^{15}N		^{1}H	^{15}N	
^{1}H,^{15}N-HSQC: ^{15}N dCUE w/ Ubiquitin	Bruker 500	5882.353	1241.44	1024	128	500.012	4.8	118.3	16
^{1}H,^{15}N-HSQC: ^{15}N Ubiquitin w/ dCUE	Bruker 500	5882.353	1520.147	1024	128	500.012	4.8	118.3	16-160
^{1}H,^{15}N-HSQC: ^{15}N dCUE w/ SUMO-1	Bruker 500	5882.353	1241.44	1024	128	500.012	4.8	118.3	16
^{1}H,^{15}N-HSQC: ^{15}N dCUE w/ PCNA	Bruker 500	5882.353	1241.44	1024	128	500.012	4.8	118.3	16

Table 2.5: Acquisition Parameters for Titration NMR Experiments

Experiment	Spectrometer	Sweep Width			Number of Complex Points			Spectrometer Frequency (MHz)	Centre Reference Chemical Shift (ppm)			Number of Scans	Mixing Time (ms)
		^{1}H	^{13}C	^{1}H	^{1}H	^{13}C	^{1}H		^{1}H	^{13}C	^{1}H		
HCCH-TOCSY	Omega 750	8849.56	11627.91	8196.72	512	50	128	749.954	4.8	42.2	4.8	4	15.2
^{13}C-NOESY-HSQC	Omega 750	8849.56	11627.91	8849.56	512	50	128	749.954	4.8	42.2	4.8	8	75
^{13}C-HSQC	Omega 750	8849.56	11627.91	-	1024	128	-	749.954	4.8	41.9	-	8	-

Table 2.6: Acquisition Parameters for ^{13}C-Edited NMR Experiments

2.5.4 NMR Titration Experiments

NMR titration experiments were used to investigate interactions between two protein molecules by studying the affects of adding an unlabelled 'ligand' protein to an isotopically labelled protein of interest. Changes in chemical shift and peak intensity of an ^{1}H, ^{15}N-HSQC were recorded. Titration experiments were carried out using fixed concentration ^{15}N-labelled protein (100 μM) with incremental steps of ligand concentration (table 2.7). ^{1}H, ^{15}N-HSQC acquisition parameters were used as in table 2.5.

For each titration experiment, two samples were prepared: one with only the observed ^{15}N-labelled protein (100 μM), representing the initial titration point, and one with the final concentration of unlabelled ligand and ^{15}N-labelled protein (100 μM). These samples were then mixed together in appropriate proportions to give the full ligand concentration range with a fixed 'observed' protein concentration. Separate ^{1}H, ^{15}N-HSQC spectra were collected for each concentration point.

As with all other NMR experiments, the buffer used was 20 mM sodium phosphate pH 7.0 and 100 mM NaCl, dCUE was the His_6-tagged dCUE-short construct (^{15}N-labelled and unlabelled) described in section 3.2. Recombinant unlabelled PCNA was prepared as in section 2.2.8. ^{15}N-labelled ubiquitin (VLI Research Inc., 98% ^{15}N), unlabelled ubiquitin (Enzo Life Sciences) and SUMO-1 (Enzo Life Sciences, His_6-tagged

human, recombinant) were all resuspended in buffer.

2.5.4.1 NMR Titration Analysis

Spectra were assigned based on previous assignments (chapter 4) and published assignments for ubiquitin (196). Analysis (CCPN) (194), Microsoft Excel and Grace graphing software (197) were used to analyse secondary chemical shifts as described in chapter 6. Ubiquitin-dCUE docking was carried out using HADDOCK (high ambiguity driven protein-protein docking) (198) running on the HADDOCK webserver (*http://haddock.science.uu.nl/*) (199). The lowest energy docking models from the top clusters, based on HADDOCK reliability score, were considered (200).

Experiment	^{15}N-Labelled Protein		Unlabelled Ligand	
	Name	Conc. (μM)	Name	Titration Point Concentration (μM)
dCUE w/ Ubiquitin	dCUE	100	Ubiquitin	0, 100, 350, 700, 1050, 1640, 2269, 3070, 3500
Ubiquitin w/ dCUE	Ubiquitin	100	dCUE	0, 100, 200, 400, 600, 800, 1266
dCUE w/ SUMO-1	dCUE	100	SUMO-1	0, 200, 400, 600, 819
dCUE w/ PCNA	dCUE	100	PCNA	0, 50, 150, 300, 500, 800, 1500, 3050

Table 2.7: List of NMR Titration Experiments - The ^{15}N-labelled protein was kept at a fixed concentration throughout the experiment.

2.5.5 Structure Calculations

The computer program Xplor-NIH 2.30 (201, 202) was used to calculate protein structures based on experimentally-determined restraints, non-bonded contacts and covalent geometry (appendix B.3.2.4). A two-step process of Xplor structure calculations was employed for all structure determinations: an initial *ab initio* simulated annealing step followed by refinement of the low-energy structures with additional RDC restraints and RAMA energy potential (203). The lowest energy refined structures from different starting structures were grouped together as an ensemble and validated. Figure 2.1 outlines this overall approach.

For each of the separate CUE domains, 500 structures were calculated in an initial simulated annealing step (*sa_l.inp* script; M. Nilges) using dihedral angle restraints (predicted from chemical shift values using TALOS+ (204)), NOE distance restraints exported from CCPN Analysis (based on peak NOE intensity), hydrogen-bond distance restraints for ordered, helical residues (205), covalent geometry and non-bonded contacts. The 100 structures with lowest overall energy were subsequently refined (*nmr/refine.inp* script; M. Nilges, J. Kuszewski, and A. Brunger) 20 times each to generate a total of 2000 structures. The refinement script included RDC and RAMA restraints and was modified with appropriate D_a and R values (section 5.5). Refined

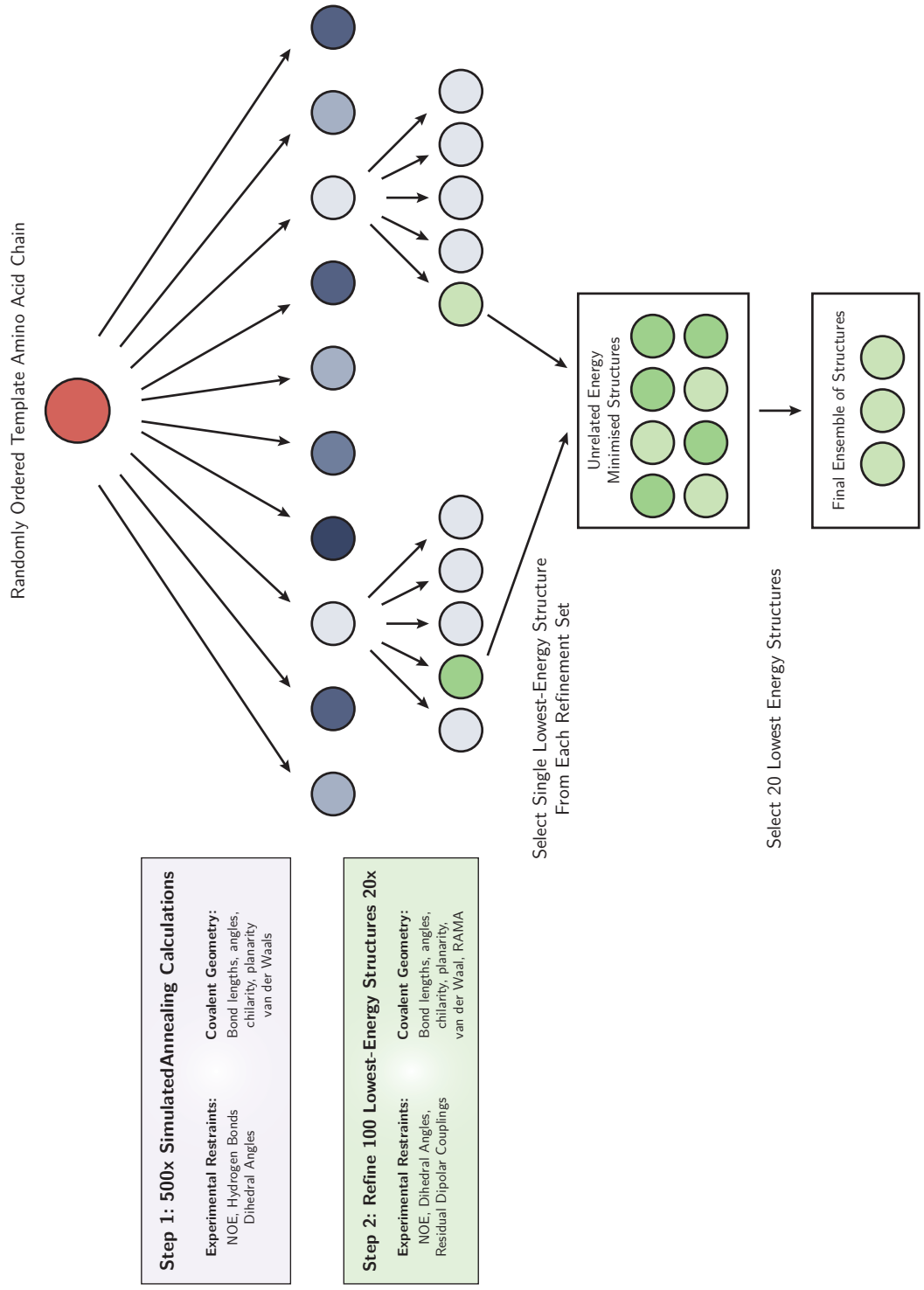


Figure 2.1: Structure Calculation Protocol - Steps involved with structure calculations: restrained simulated annealing, refinement of low-energy structures (light blue) and selection of lowest-energy structures (green). The final ensemble contains unrelated low-energy structures (i.e. those that do not share a parent structure). This ensemble of structures was then validated.

structures with the lowest potential energy from each set were compared with one another and the 20 lowest energy structures used to produce an ensemble of structures. Calculation of the complete dCUE structure (both CUE domains with linker) was instead done with an initial round of 100 simulated annealing runs and 20 refinement runs of the 20 lowest energy structures. An ensemble of 10 lowest energy structures was used for comparison.

Ensembles were superimposed using Theseus, a maximum likelihood method of macromolecular superimposition (206, 207), and validated with the help of the Protein Structure Validation Suite (PSVS) (208), which incorporates various tools including PROCHECK (209), MolProbity (210), Verify 3D (211) and Prosall (212).

Minimised average structures were generated from a superimposed ensemble PDB (protein data bank) file of structures for each CUE domain. The position of each Cα was averaged across all structures in the ensemble. This single, average structure was then energy minimised using Xplor-NIH 2.30.

2.6 Pull-down of Recombinant PCNA

The ability of both full length SMARCAD1 and the dCUE-short constructs to bind recombinant unlabelled PCNA was investigated using a pull-down experiment. His-tagged full-length SMARCAD1 was generously provided by Dr. Jacqueline Mermoud

and Dr. Patrick Varga-Weisz (Babraham Institute, Cambridgeshire, UK) (108). Full-length SMARCAD1 and dCUE (section 2.2.1) were immobilised on Ni-NTA beads, mixed with PCNA, washed (either high or low salt wash) and probed by western blot to determine whether either full-length SMARCAD1 or dCUE retain PCNA relative to a control of blank beads with no 'bait' protein.

Ni-NTA beads (Qiagen, 20 µl) were placed in a microcentrifuge spin-column, washed with 0.5 ml H_2O, spun for 3 min at 150 ×g and washed with wash buffer (10 mM imidazole in PBS, 0.1% Tween-20, 2 mg/ml BSA (bovine serum albumin)). Either His_6-tagged dCUE (50 µg in 20 µl wash buffer), His_6-tagged full-length SMARCAD1 (5 µg in 20 µl wash buffer) or a control (20 µl wash buffer) was applied to the beads. Samples were incubated on ice for 10 min with mixing every minute. Columns washed 2× 0.5 ml wash buffer, beads resuspended (20 µl wash buffer) and recombinant untagged PCNA (32 µg, section 2.2.8) added. Samples were incubated on ice for 30 min with mixing. The samples were washed thoroughly (7× 0.5 ml wash buffer with either low salt (50 mM NaCl) or high salt (500 mM NaCl)) and eluted (100 µl 500 mM imidazole, 500 mM NaCl in PBS). Eluted samples were separated by SDS-PAGE using a pre-cast 4-12% pre-cast gel (Invitrogen) (section 2.2.5) and western blotted with anti-PCNA primary antibody (mouse, Santa Cruz IgG_2) and Bio-Rad kaleidoscope marker (section 2.2.7).

2.7 Pull-down from Nuclear Extract

Pull-down experiments of proteins in nuclear extract by full-length SMARCAD1 and dCUE were carried out as for recombinant PCNA (section 2.6) with the following modifications. Instead of resuspending dCUE / SMARCAD1 coated beads in wash buffer and adding PCNA, the beads were resuspended in HeLa cell nuclear extract (250 µl supernatant; centrifuged for 15 min, 30,000 ×g at 4 °C) . The nuclear extract was generously provided by Dr. Patrick Varga-Weisz and prepared as described (213). Incubation with nuclear extract was for 60 minutes on ice and 8 thorough wash steps were performed. Eluted samples were western blotted with primary antibodies (table 2.8) and appropriate secondary antibodies (anti-mouse as in section 2.2.7 or anti-rabbit IgG, HRP (horseradish peroxidase)-linked).

Primary Antibodies

Name	Supplier	Species	Concentration
anti-PCNA	Santa Cruz IgG_2	mouse	1:500
anti-KAP1	Abcam (Ab20C1)	mouse	1:500
anti-SMARCAD1	P. Varga-Weisz (108)	rabbit	1:5000
anti-Ubiquitin (P4D1)	Santa Cruz (Sc-8017)	mouse	1:500
anti-SUMO-1	Abcam (Ab11671)	rabbit	1:500
anti-SUMO-2/3	Abcam (Ab3742)	rabbit	1:500
anti-HP1γ	Upstate (07-332)	rabbit	1:500
anti-Histone H1	Upstate (05-457)	mouse	1:500
anti-H2A-Ub	Upstate (E6C5)	mouse	1:500
anti-H3; polyclonal	Abcam (Ab179)1	rabbit	1:10000
anti-H4K31-Ub	Abcam (Ab15585)	rabbit	1:500

Table 2.8: List of Primary Antibodies Used in Pull-down Experiments

Chapter 3

Overexpression and Initial Structural Characterisation of SMARCAD1 Double CUE Domain

3.1 Introduction

Chromatin remodeling factors typically consist of an ATP-dependent helicase-like domain in combination with various accessory domains. The domain strucutres vary between remodeling proteins with trends across families. For example, SWI/SNF family members have a bromodomain for recognition of acetylated lysine residues, ISWI remodelers contain a HAND-SANT-SLIDE domain that is involved in DNA interaction and CHD members contain a double chromodomain for recognition of methylated lysine residues (214, 215, 216). In addition to common accessory domains across a family, chromatin remodelers can also share some of the many other types of DNA, histone

and other protein interaction domains (examples include Zinc fingers, PHD domains and SH3 domains respectively). These accessory domains help target and define the biological function of the remodelers (217).

Although CUE domains occur in a range of proteins with different functions (125, 218), SMARCAD1 is the only known chromatin remodeler with a CUE domain. Furthermore, sequence analysis of SMARCAD1 (section 3.2) has only robustly identified CUE domains in addition to the ATP-dependent Snf2 region. Together these observations suggest a previously unknown role for CUE domains in chromatin remodeling and especially a key role in determining the biological function of SMARCAD1.

In SMARCAD1, the CUE domains are located relatively close together, in tandem, forming a double CUE domain. This raised the possibility that the CUE motifs functioned together as a single structural unit for binding purposes. To understand this aspect of the double CUE domain, structural biology techniques were used.

3.2 Sequence Analysis and Construct Design

Full-length human SMARCAD1 is 117.4 kDa in size and 1026 amino acids long. In order to design protein constructs, sequence analysis was carried out in order to identify CUE domains and structured regions.

3.2.1 Identification of CUE Domains

Pfam, InterPro and PROSITE are searchable databases containing descriptions of known protein families and domains (219, 220, 221, 222). Protein domains, such as CUE domains, share conserved sequence profiles (figure 3.1), which means that any amino acid sequence can be compared against database profiles to predict the occurrence of such domains.

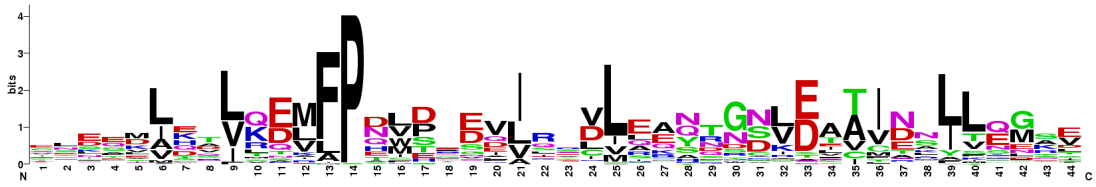


Figure 3.1: PROSITE Consensus Sequence for CUE Domain - CUE domains are identified by a 42-44 amino acid long consensus profile. More highly conserved residues are represented by a higher letter. Structurally, the CUE domain is spread over three helices. Two motifs are highly conserved: an (M/L)FP sequence at the C-terminal end of the first helix, and a di-leucine motif in the third helix. Sequence logo generated by WebLogo (223).

The sequence for SMARCAD1 was searched using Pfam, InterPro and PROSITE, resulting in two CUE domains being identified in the N-terminal half of the protein (figures 3.2). The regions identified as CUE domains gave a strong match and were found to be separated by approximately 50 amino acid sequence, raising the possibility that the two CUE domains functioned independently or together.

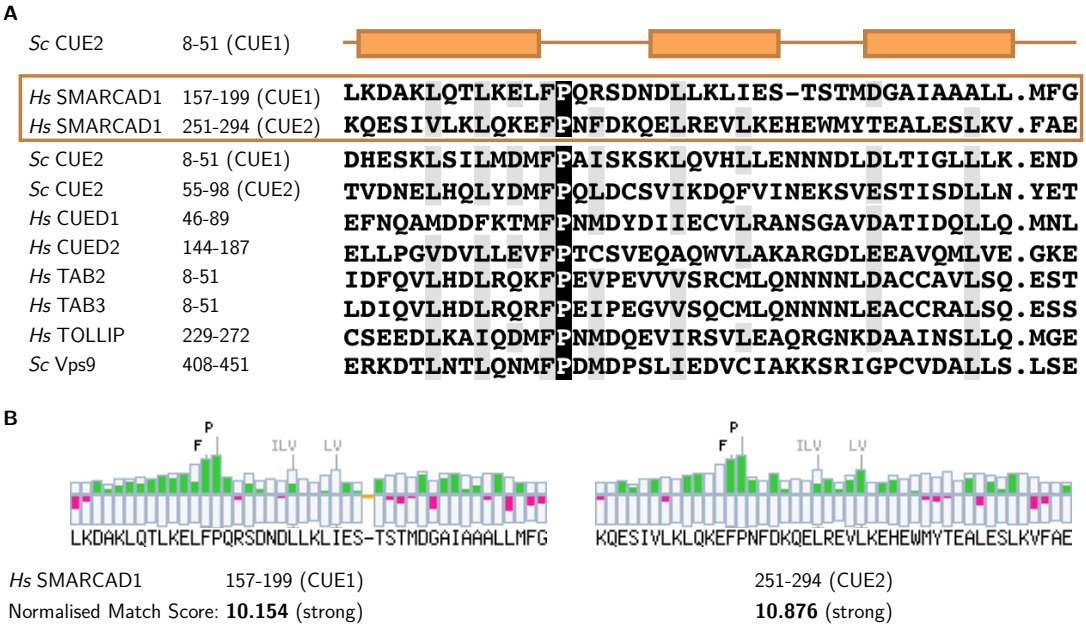


Figure 3.2: Sequence Homology of CUE motifs - **A** Selection of CUE domains identified using the PROSITE CUE domain definition. Highly conserved CUE domains occur in a range of proteins with different functions and across species (including Humans, rats, Fruit flies and yeast). The secondary structure of the N-terminal CUE domain from *Saccharomyces cerevisiae* (*Sc*) protein CUE2 is shown above the sequences for reference (167). Orange boxes represent an alpha helix. The highly conserved proline residue at the end of the first alpha helix occurs in all CUE domains. Protein names and species - *Sc* (*Saccharomyces cerevisiae*) and *Hs* (Human) - are listed along with the sequence position of the CUE domain. Both CUE domains identified from SMARCAD1 are highlighted at the top showing similar conserved residues. **B** Comparison of PROSITE CUE Hidden Markov Model (i.e. consensus profile) with actual sequence of both CUE domains from SMARCAD1. Underlying grey bars represent maximum and minimum scores from the consensus profile; importance of residues at particular positions. Green / magenta bars reflect how well the sequence correlated with the best / worst match. PROSITE normalised scores for both sequences are approximately 10, confirming a strong significant match with the CUE profile (as compared with a random sequence). Results generated using MyHits (224).

3.2.2 Secondary Structure Prediction

Secondary structure prediction algorithms PSIPRED and Jpred were used to determine the limits of helices and identify structured units (225, 226, 227) (figure 3.3). Like all CUE domains, the CUE domains in SMARCAD1 consist of three alpha helices. Very little secondary structure was identified between the two CUE domains, suggesting that the region between the two CUE domains is a flexible linker.

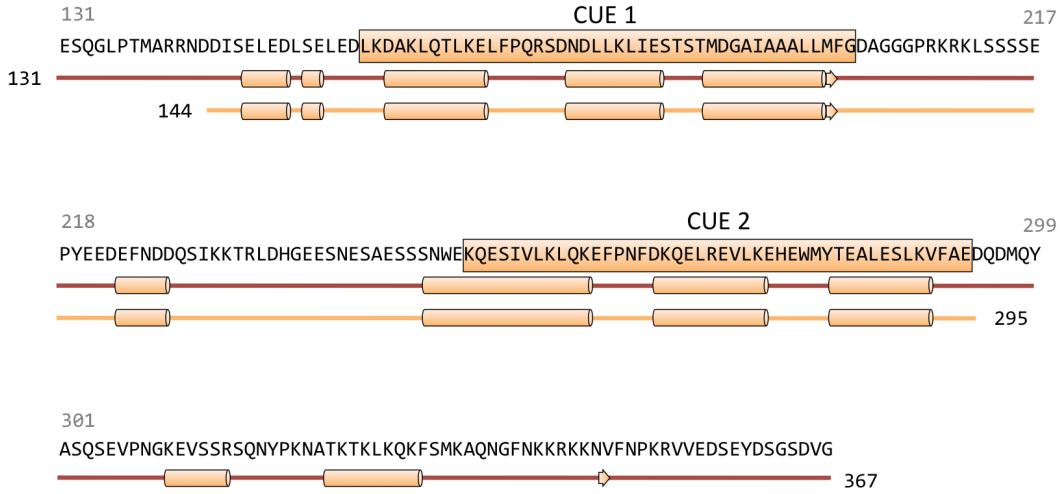


Figure 3.3: Secondary Structure - Secondary structure prediction for region of SMARCAD1 double CUE domain using JPred algorithim (227). Sequences corresponding to two constructs, dCUE-long and dCUE-short, are shown in red and orange respectively. Secondary structure prediction shows a classic three-helix structure for each CUE domain with a long coiled linker region with little secondary structure. Another secondary structure prediction algorithm, PSIPRED, gave near identical results (225, 226). Predicted CUE motifs as in section 3.2.1 are shown in an orange box.

3.2.3 Protein Disorder Prediction

Disordered regions of proteins often represent a bottleneck in structural biology. They can be difficult to purify and can be difficult to crystallise or solve using nuclear mag-

netic resonance (NMR) (228). Prediction of protein disorder from sequence was carried out using several online services; RONN, DIOPRED2 and PrDOS (229, 230, 231). As with secondary structure prediction (section 3.2.2), disorder prediction identified a long, disordered linker region between two structured CUE domains (figure 3.4).

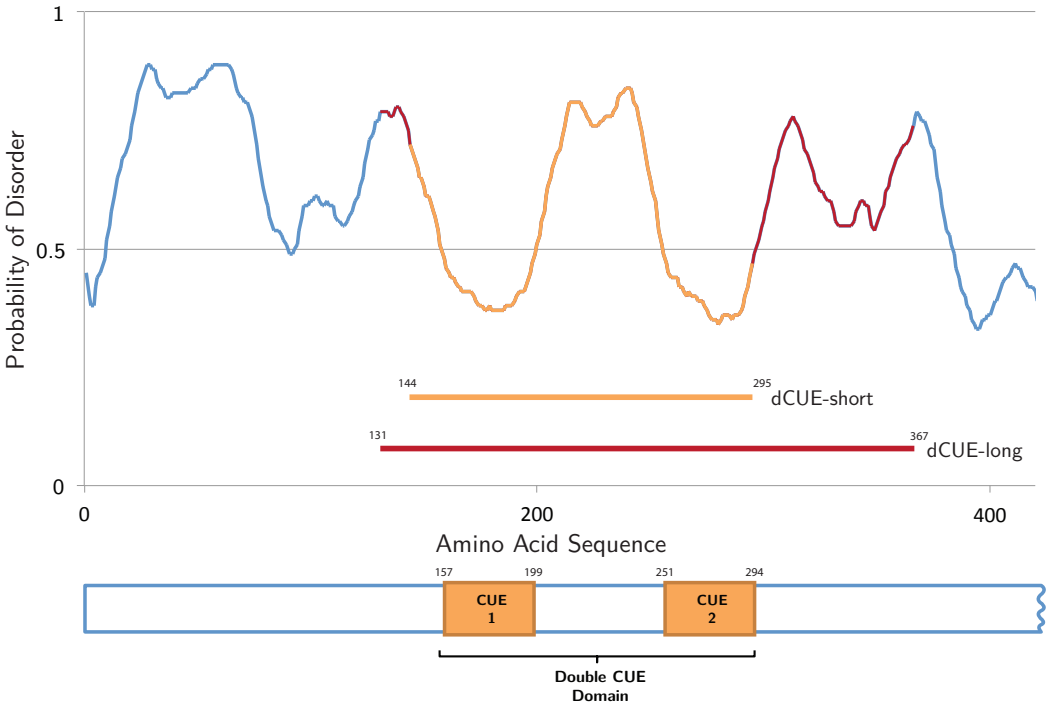


Figure 3.4: Disorder Prediction Plot - Probability of disorder for the N-terminal sequence of SMARCAD1. Data plotted as predicted by Regional Order Neural Network (RONN) (229). Alternative algorithms, DISOPRED2 (230) and PrDOS (231), gave comparable results (data not shown). Two major dips correspond to annotated CUE motifs (section 3.2.1) and reflect the structured nature of the CUE domains. The region between the two CUE domains is predicted as disordered, suggesting a flexible linker region. Regions of the plot correlating with two different expression constructs dCUE-short and dCUE-long are shown in orange and red respectively.

3.2.4 Construct Design

A schematic of domain organisation was produced (figure 3.5) based on the sequence analysis described above (sections 3.2.1, 3.2.2 and 3.2.3). The region covered by the double CUE domain represents approximately a 150 aa stretch in the middle of the non-ATPase part of the protein.

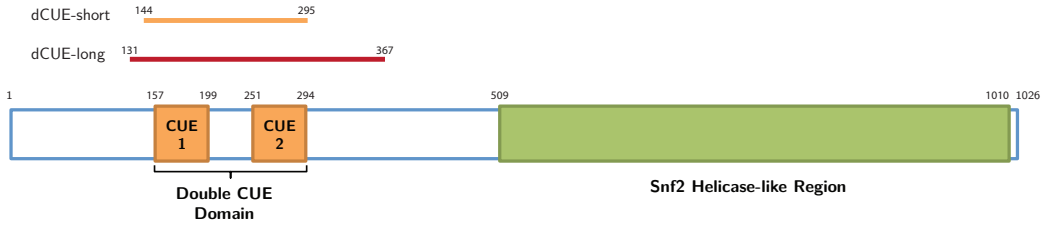


Figure 3.5: Domain Structure of SMARCAD1 - Bioinformatic analysis of SMARCAD1 identified two CUE motifs in the N-terminal half of SMARCAD1 separated by approximately 50 amino acids. The C-terminal half of the protein consists of the common ATP-dependent Snf2 helicase-like region. Two His_6-tagged constructs, dCUE-long and dCUE-short, were cloned covering the indicated region of SMARCAD1.

An expression construct of the N-terminal region of Human SMARCAD1 (amino acid sequence 131-367) in pGEX-4T-1 (GE Healthcare) (i.e. a GST-tagged SMARCAD1 131-367 construct) was provided by Patrick Varga-Weisz (Babraham Institute, Cambridgeshire) (108). This GST-dCUE construct was found to overexpress at high levels in *E. coli* but could not sufficiently be cleaved and separated from the GST tag for crystallisation trials (data not shown). Both the GST tag and dCUE region in this construct are approximately 27 kDa, which complicated complete separation.

Given the high-level of protein expression, GST-dCUE was used as the basis for

designing two constructs to be cloned into a vector providing a much smaller hexa-His tag. dCUE-long (SMARCAD1 131-367) and dCUE-short (SMARCAD1 144-295) were subsequently designed to optimise coverage of the CUE domains and purification (figure 3.5). dCUE-long is the same sequence as that used by Varga-Weisz, including flanking regions around the CUE domains. dCUE-short was optimised to cover the CUE domains with minimal additional disordered region. It was decided to produce constructs that kept both CUE domains together as they occur in SMARCAD1, with the 52 amino acid linker region, so that the domains could be studied together.

3.3 Protein Production

The dCUE-long and dCUE-short constructs were successfully cloned and expressed using a bacterial expression vector. High levels of soluble protein were purified (figures 3.6 and 3.7). Protein size and identity were confirmed by mass spectrometry and western blot respectively (figure 3.8 and 3.9).

3.4 Oligomeric State Determination

CUE motifs have a range of binding mechanisms with examples of monomers and dimers representing the functional unit (167). In order to establish the nature of the SMARCAD1 CUE domain, the oligomeric state was investigated. Results from analytical gel filtration and dynamic light scattering (DLS) were inconclusive, broadly suggesting 2-3

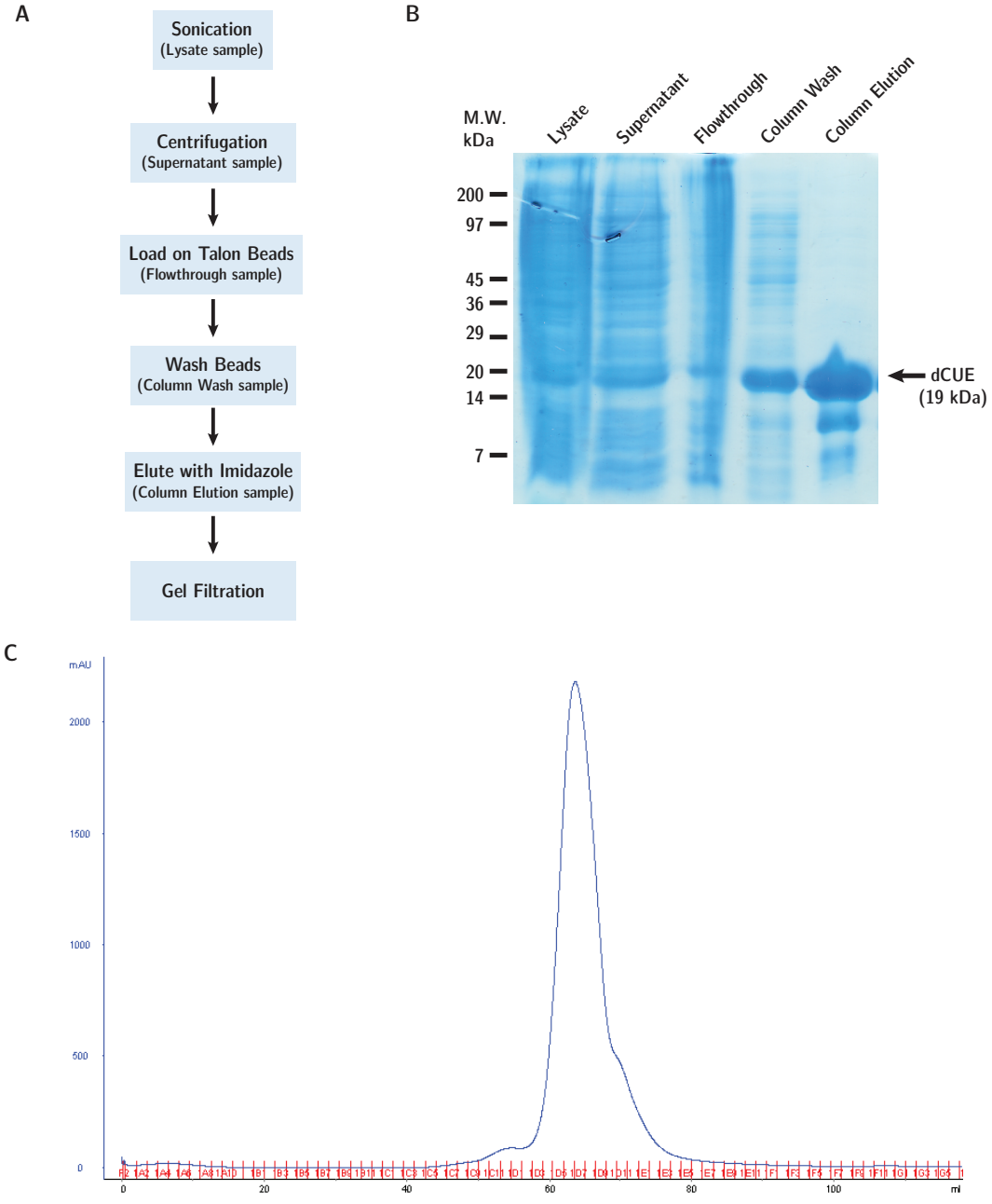


Figure 3.6: Protein Purification of dCUE - **A** Overview of purification protocol. Resuspended *E. coli* cell pellets were lysed by sonication and the supernatant loaded onto Ni-NTA beads, which bind His-tagged proteins. The beads were washed and protein eluted with imidazole. dCUE protein was finally purified using a gel filtration column and fractions containing only dCUE pooled. See section 2.2.1 for full methods. **B** Coomassie gel showing samples from purification of dCUE-short: initial *E. coli* lysate, supernatant (soluble protein), column flow-through (i.e. protein that does not bind to talon beads), column wash and column elution. The eluted fraction has contaminants (possibly degradation products), which were removed by gel filtration. **C** UV_{280} trace (mAU) of dCUE-short eluant sample after Ni-NTA column. The gel filtration step separates molecules by size and the large peak corresponds to highly enriched dCUE-short protein.

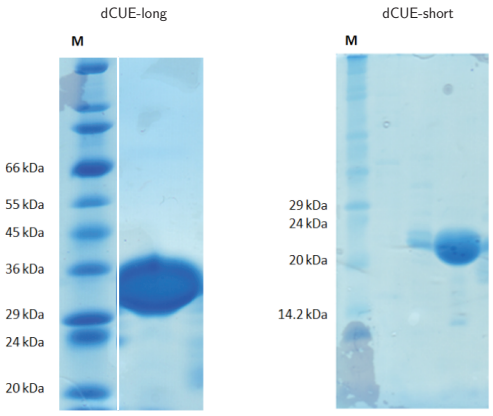


Figure 3.7: Expression and Purification of Constructs - Both dCUE-long and dCUE were highly expressed and purified, with little contaminating proteins. Gels shown after final gel filtration purification step.

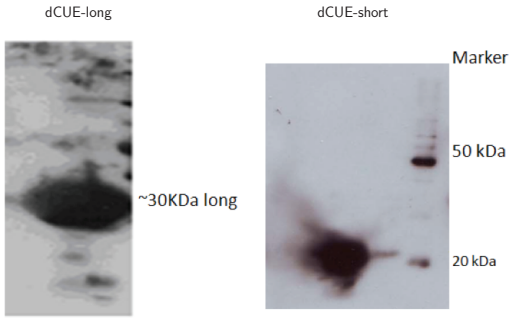


Figure 3.8: Western Blot Confirmation of His_6-expression - Western blots of cell lysate were probed with anti-hexahistadine tags, identifying His_6-tagged protein bands at the exprected size. This confirmed that full length tagged recombinant protein is being expressed.

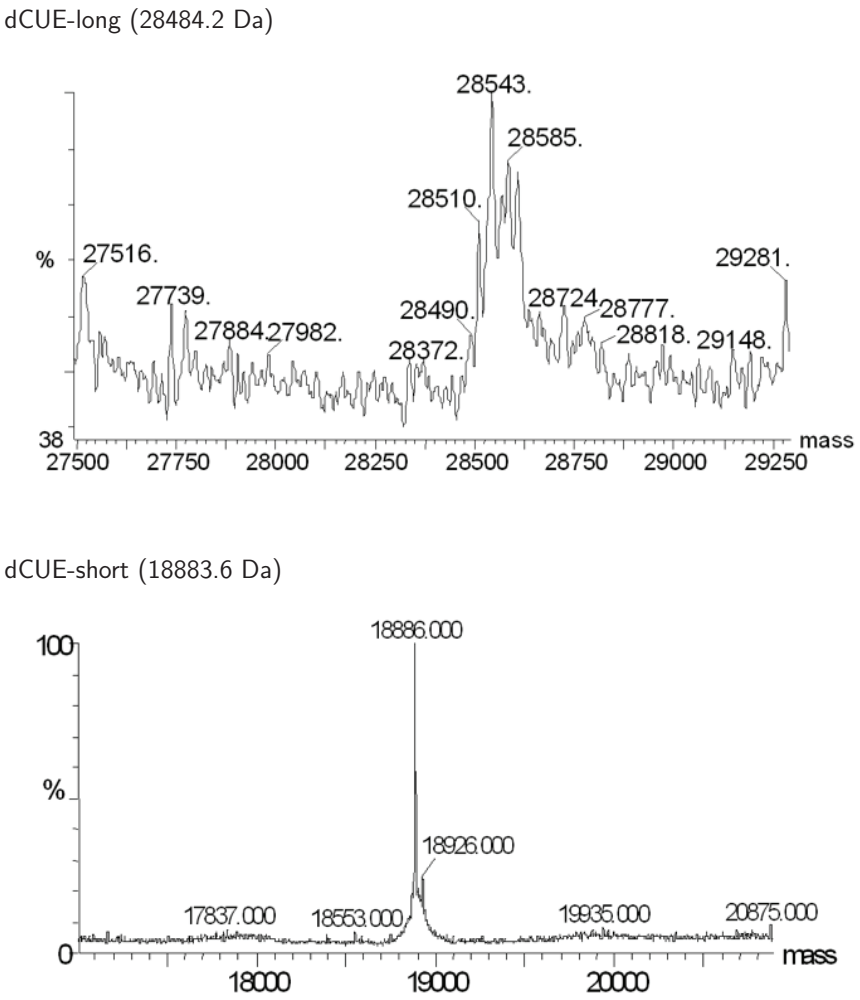


Figure 3.9: Mass spectrometry results - Spectra for purified dCUE-long and dCUE-short proteins. The theoretical average mass is shown in brackets. The spectrum for the dCUE-long sample is very noisy and there is no clear single peak corresponding to specific mass.

times the molecular weight of a single dCUE-short / dCUE-long monomer (data not shown). In contrast, multiple angle light scattering (MALS) determined the size of dCUE-short and dCUE-long to be extremely close to that of monomers (figure 3.10). The discrepancy between DLS, analytical gel filtration and MALS results could be explained by a single monomer with an elongated, flexible structure causing it to appear much like a larger protein in DLS and gel filtration. Unlike DLS, which measures the hydrodynamic size of a protein, MALS is capable of determining the shape independent molecular mass of the protein, taking into account non-spherical structure. Given the accurate match of the MALS mass measurement to the actual mass of the monomers and that it is a superior technique, it is highly likely that the double CUE domain is a monomer in solution.

3.5 Crystallisation

Attempts to crystallise the protein were unsuccessful. Both dCUE-long and dCUE-short were tested using a wide range of crystallisation conditions, including varied protein concentrations from less than 10 mg/ml to more than 150 mg/ml. Surface lysine methylation was tested on both proteins to see if this would enhance precipitation (192). A series of co-crystallisation trials with ubiquitin and dCUE-short also didn't yield any results. One possible explanation for the difficulties forming crystals could

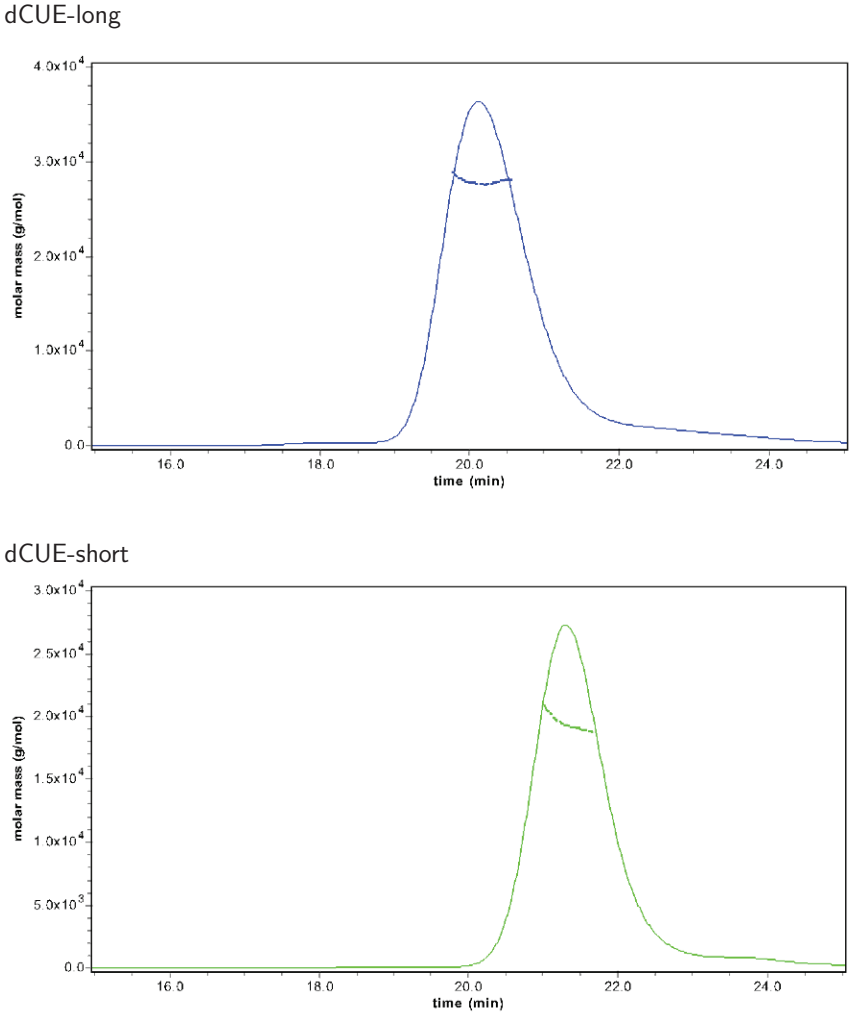


Figure 3.10: Multiple-angle light scattering (MALS) results - Samples were run through a size exclusion column at a concentration of 4 mg/ml. The UV_{280} absorbance trace identifies protein elution (peak) as it enters the MALS chamber. Points under the peaks correspond to molar mass identified by MALS. dCUE-long was found to have a mass of 28.5 kDa kDa, suggesting a monomer (28.6 kDa) and dCUE-short was in solution at 18.9 kDa, which also corresponds to a monomeric species (19.0 kDa).

be the presence of the flexible linker between CUE motifs. This flexibility might have enhanced solubility and prevented crystal packing.

In several trials with very high protein concentration, a solid, gel-like substance formed after several days (figure 3.11). This was not crystalline but was likely to be almost predominantly protein given the very high concentrations involved. This underscored the difficulty of growing crystals from such soluble protein.

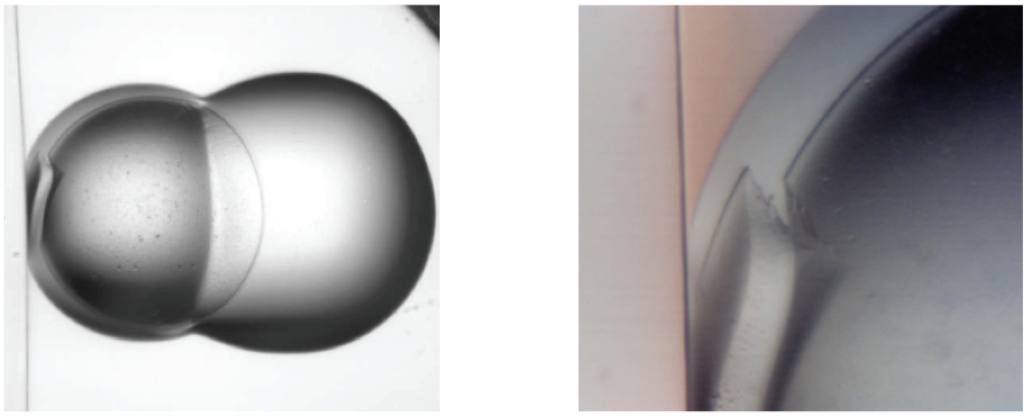


Figure 3.11: Formation of Non-Crystalline Protein 'Gel' - At very high protein concentrations (greater than 120 mg/ml) some crystallisation trials produced an unusual gel-like substance. The image on the left is of a crystallisation trial after several days. The precipitant liquid has been expelled to the right and the left portion of the drop appears solid. The image on the right is of a higher magnification and shows a crack in the substance, highlighting it's solid nature.

3.6 NMR Trials

Nuclear Magnetic Resonance (NMR) experiments can be used to solve protein structures without the need for protein crystals. Given the very high solubility of the double

CUE domain, the inability to grow crystals and the relative small size of dCUE-short (19 kDa), it was considered a good potential target for structural solution by NMR. An initial ^{15}N-labelled sample of dCUE-short was tested using a 2D ^{1}H, ^{15}N-HSQC experiment and provided encouraging results (figure 3.12).

It was decided to solve the structure of the double CUE domain using NMR. Subsequently in this thesis, all references to the 'dCUE' construct are about dCUE-short and all sequence numbers refer to this construct as opposed to SMARCAD1 residue numbers.

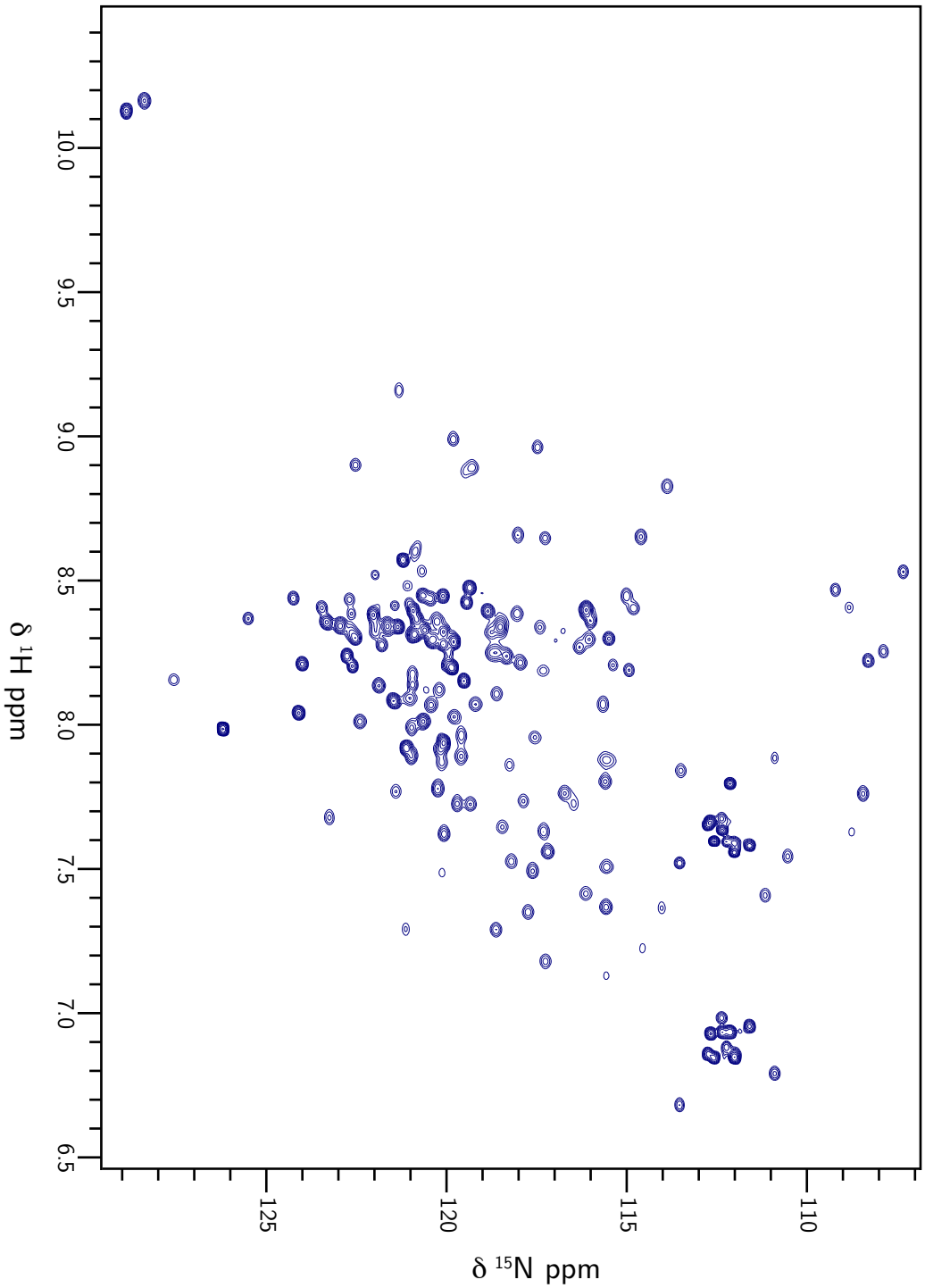


Figure 3.12: ^{1}H, ^{15}N-HSQC from dCUE-short - The initial NMR experiment was an ^{1}H, ^{15}N-HSQC. NMR is a relatively size-sensitive technique, which quickly gets increasingly more complicated to solve when the molecular mass increases. Accordingly, dCUE-short was used as it is smaller but still contains the main points of interest, the two CUE domains. This spectra shows relatively good dispersion in that most peaks are resolved but there is considerable overlap in the middle.

Chapter 4

Assignment of Backbone Resonances in the Nuclear Magnetic Resonance Spectrum of SMARCAD1 Double CUE Domain

4.1 Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy is a widely used technique for determining the structure, interactions and dynamics of protein molecules. NMR is unique among methods used for solving protein structures at atomic resolution as data can be collected using proteins in solution without the need for crystallisation. This makes NMR particularly applicable for studying proteins resistant to crystallisation, including partially folded proteins such as prion proteins (232, 233).

NMR is also an important tool in studying the functional aspects of proteins and

their interactions with ligands. Detecting and measuring binding of small molecules to proteins by NMR is routine in drug discovery. Even relatively weak protein-protein interactions can be investigated using NMR and provide information about binding sites and affinities (234).

Given the failure to crystallise SMARCAD1 double CUE domain (section 3.5) and reasonable chemical shift dispersion in the ^{1}H, ^{15}N heteronuclear single quantum coherence spectrum (^{1}H, ^{15}N-HSQC; section 3.6), further NMR experiments were carried out in order to solve the structure of the SMARCAD1 CUE domains and to investigate functional binding activity. Sequential assignment of backbone resonances is the prerequisite step for interpreting further NMR experiments that give structural, dynamic and functional information. Details of NMR theory are discussed in appendix B.

4.2 Strategy for Assignment of Backbone Resonances

The majority of backbone assignments were successfully made using a series of triple resonance experiments. The broad principles involved collecting a 2D ^{1}H, ^{15}N-HSQC experiment, arbitrarily assigning the peaks a given ID number and then using information from 3D experiments (typically triple resonance experiments) to identify the associated spin system type (i.e. residue type) and linking these to neighbouring residues based on through-bond information. Based on short stretches of linked spin systems

of known types, it was possible to unambiguously assign most of the protein, given the known sequence. Figure 4.1 graphically shows the principles involved (235).

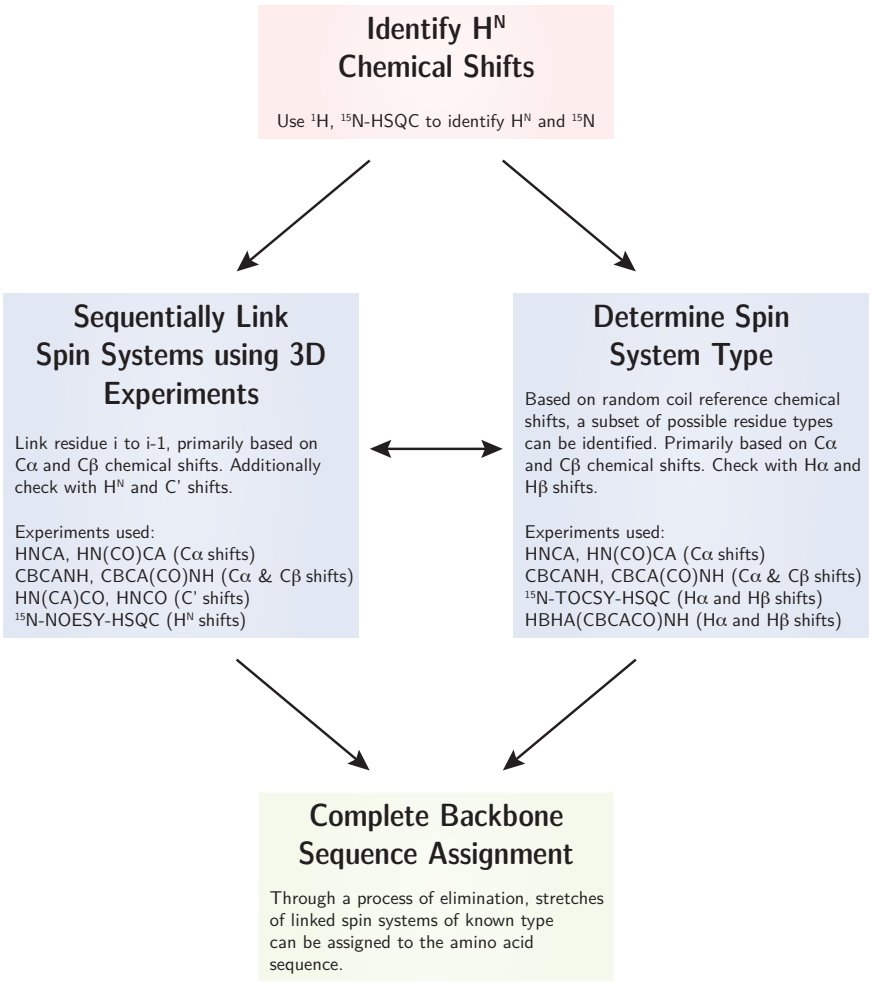


Figure 4.1: Overview of Backbone Assignment Process - A combination of triple resonance experiments and ^{15}N-edited TOCSY / NOESY-HSQC allowed both linking of spin systems and identification of their type. This made it possible to unambiguously assign the majority of residues (235).

4.3 Results

The SMARCAD1 dCUE construct used for NMR assignment has 166 residues, 14 of which were introduced as part of the N-terminal His_6-tag used for protein purification, and approximately 50 belong to a putatively unfolded linker region between the two CUE domains (section 3.2.2, figure 4.2). The ^{1}H, ^{15}N-HSQC spectrum of SMARCAD1 dCUE shows approximately 177 peaks, most of which correspond to the backbone ^{15}N-$^{1}H^{N}$ of each amino acid. The subsequent sections shows typical results for various NMR experiments used to assign the backbone atoms.

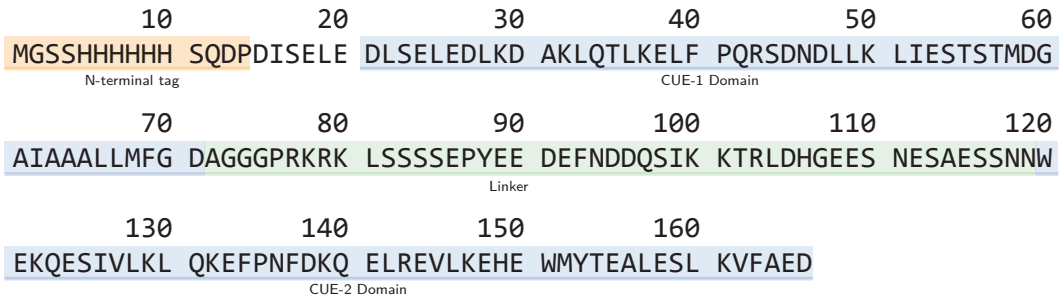


Figure 4.2: SMARCAD1 dCUE-short Construct Sequence - Numbering corresponds to amino acid position of the dCUE construct. Residues 15-166 are SMARCAD1 sequence and corresponds with full-length SMARCAD1 position 144-295. Region shaded in orange is the N-terminal His-tag (not part of full-length SMARCAD1 sequence). The two CUE domains are shown in blue and the linker shaded green.

4.3.1 ^{1}H, ^{15}N-HSQC

A ^{1}H, ^{15}N heteronuclear single quantum coherence (HSQC) experiment utilises $^{1}J_{NH}$ coupling to transfer magnetisation between ^{1}H and ^{15}N nuclei (figure 4.3). The resulting

spectrum has a peak corresponding to each proton attached to labeled, ^{15}N, nitrogen.

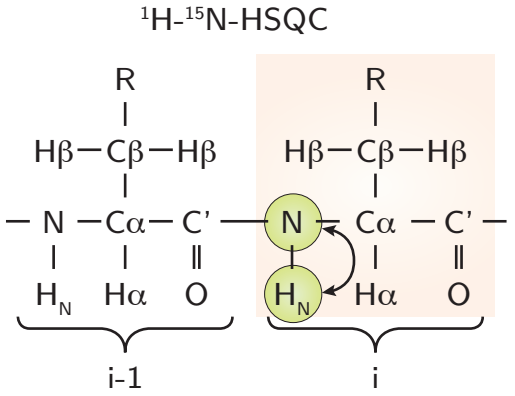


Figure 4.3: ^{1}H, ^{15}N-HSQC Coherence Transfer Diagram - ^{1}H, ^{15}N HSQC experiments are used as the starting point in sequential backbone assignment as they give one peak for each residue, side-chain amide groups and tryptophan H^{ϵ_1} (236).

An ^{15}N-labeled sample of SMARCAD1 dCUE was used to collect an ^{1}H, ^{15}N-HSQC spectrum (237). Dispersion of the spectrum is reasonable, however not all peaks are resolved and there is significant overlap. Approximately 75% of all peaks have a ^{1}H chemical shift within 0.5 ppm of 8.0 ppm (i.e. peaks are between 7.5-8.5 ppm). Figure 4.4 shows the assigned ^{1}H, ^{15}N-HSQC spectrum for SMARCAD1 dCUE. The poorly dispersed region in the centre of the spectrum has been magnified in figure 4.5.

The ^{1}H, ^{15}N-HSQC spectrum was used as the basis for assignment. Each peak was arbitrarily numbered and then 3D experiments built upon this by adding extra information in a third dimension.

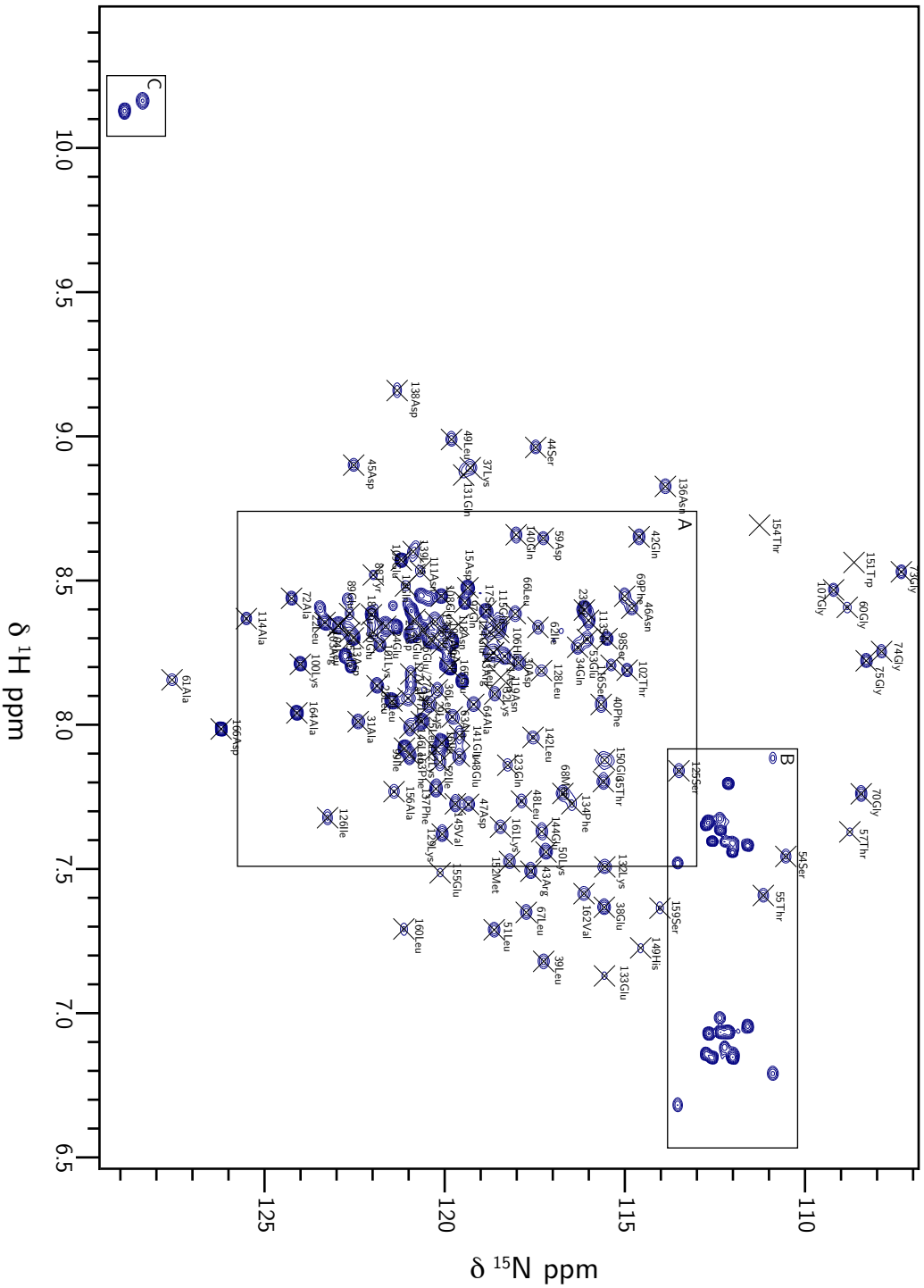


Figure 4.4: Assigned ^{1}H, ^{15}N-HSQC Spectrum for SMARCAD1 Double CUE Domain (dCUE). - Box A Residues in the centre of the spectrum are relatively poorly resolved. A higher magnification of the region shows this in more detail (figure 4.5). **Box B** Asn / Gln NH_2 region (figure 4.6). **Box C** Characteristic tryptophan H^{ϵ_1} from the indole group have a typical shift around 10 ppm.

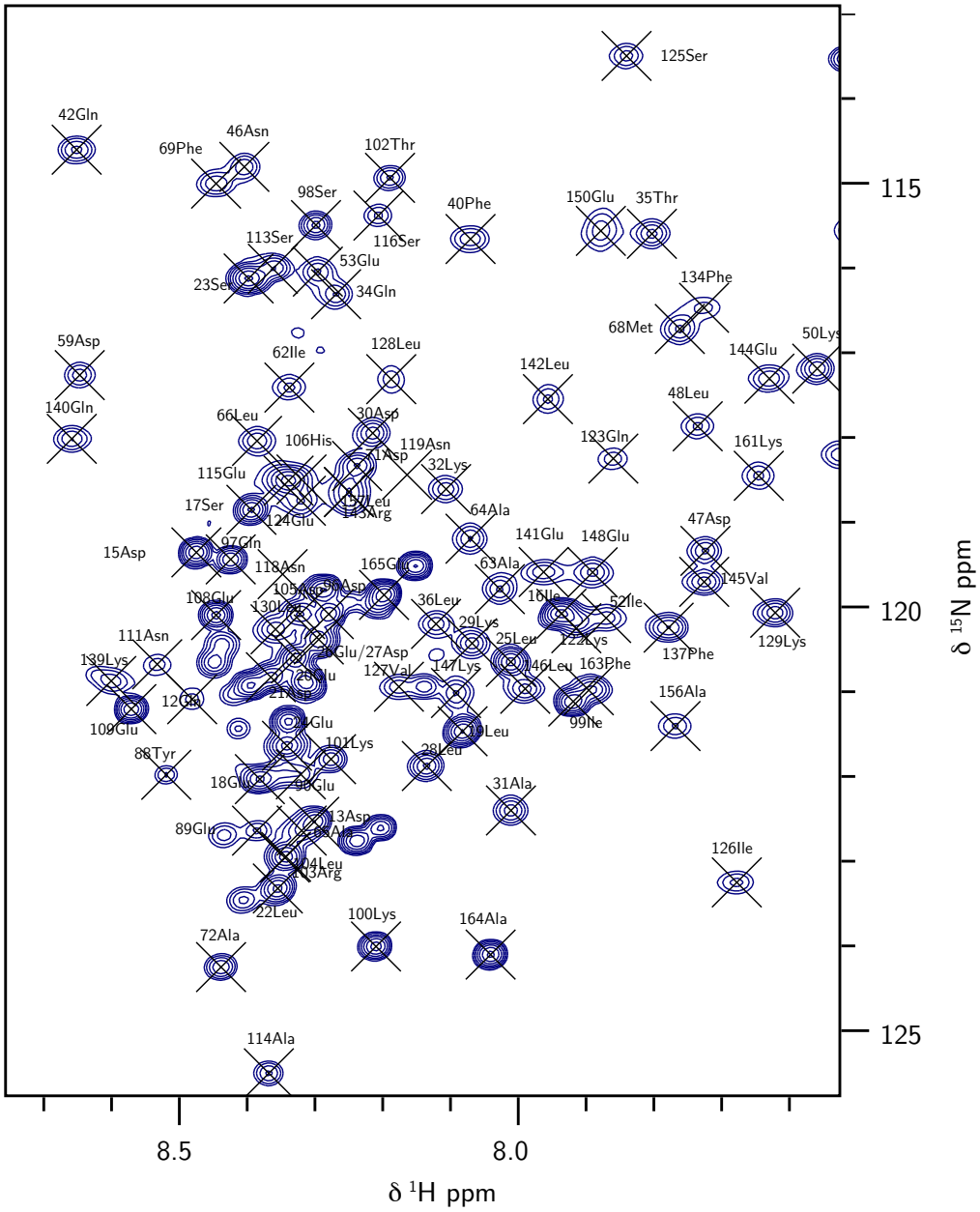


Figure 4.5: ^{1}H, ^{15}N-HSQC - Magnified region of the poorly dispersed central area of the ^{1}H, ^{15}N-HSQC. Many of the peaks overlap and are difficult to assign.

4.3.2 Identification of Side-chain H^N in the ^1H, ^{15}N-HSQC

As the ^1H, ^{15}N-HSQC spectrum gives a peak for all NH groups, this means that side-chain NH, such as those from asparagine and glutamine, also give a peak. An ^1H, ^{15}N-HSQC experiment that selects for NH_2 groups (figure 4.6) identified these peaks as side-chain peaks and thus they could be discounted during backbone assignment. Similarly, it is often possible to identify peaks corresponding to those from tryptophan's indole group as they frequently have a shift around 10 ppm, far from the bulk of the backbone ^{15}N-$^1H^N$ shifts, and give different results in other experiments. For example, no peak is observed in the HNCA spectra but an H^{ϵ_1}-H^{δ_1} peak is seen in the TOCSY spectrum (238).

4.3.3 Triple Resonance, 3D Experiments

As discussed in section 4.2, a strategy using triple-resonance experiments with a ^{13}C, ^{15}N-labeled sample of SMARCAD1 dCUE was used to make sequential backbone assignments (239). Spin systems were linked together based on carbon shifts, particularly Cα and Cβ shifts, and the spin system type determined based on similarity to reference chemical shifts.

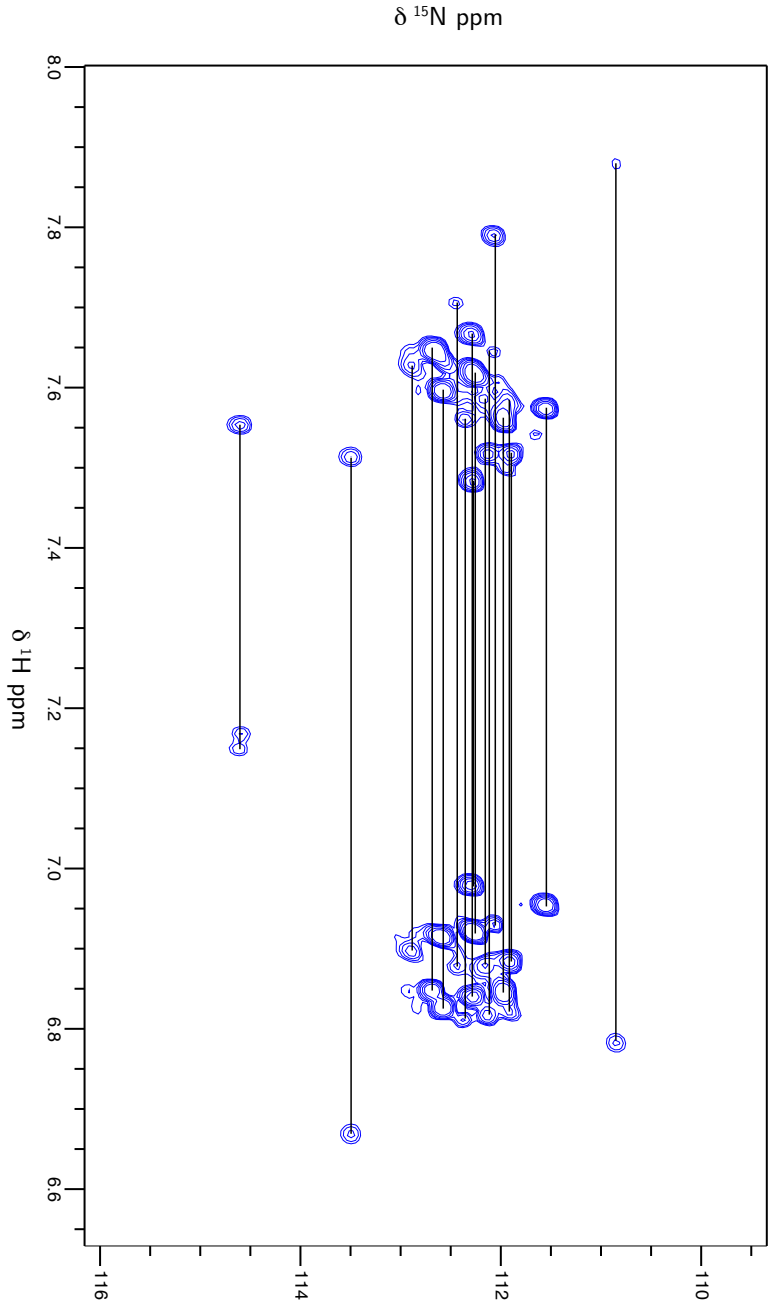


Figure 4.6: HSQC (NH_2 selective) - The HSQC (NH_2 selective) experiment is similar in principle to an ^{1}H, ^{15}N-HSQC but utilises a shorter time delay ($^{1}/_{8}J_{NH}$; appendix B.3.1) to only include resonances from NH_2 groups, such as those from asparagine and glutamine side chains. Therefore this allows for the identification of the majority of non-backbone peaks in the HSQC.

4.3.3.1 Sequential Connections Using Cα Chemical Shifts

HNCA and HN(CO)CA spectra (figure 4.8) were used for sequential connections and collecting Cα chemical shifts. HNCA experiments give two peaks in the carbon dimension, one corresponding to the Cα of the residue with the NH group (i) being directly measured and the other from the previous residue (i-1). HN(CO)CA, by contrast, only gives a peak from the Cα of the previous residue (i-1). Figure 4.7 shows the coherence transfer diagram for these experiments (236, 239, 240).

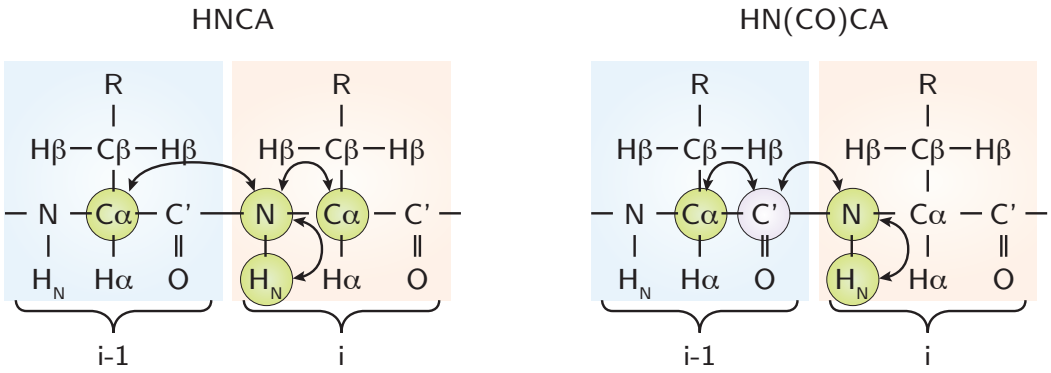


Figure 4.7: HNCA and HN(CO)CA Coherence Transfer Diagram - These experiments give Cα shifts for residues i and i-1; and just i-1 respectively (236, 241).

By overlaying these two spectra, it is possible to identify the i-1 and i Cα shifts from the HNCA experiments and link spin systems together (figure 4.8).

4.3.3.2 Spin System Identification and Connection using Cβ Chemical Shifts

CBCANH and CBCA(CO)NH spectra (figures 4.10 and 4.11) were used for additional sequential connection, collecting Cβ chemical shifts and for determining spin system

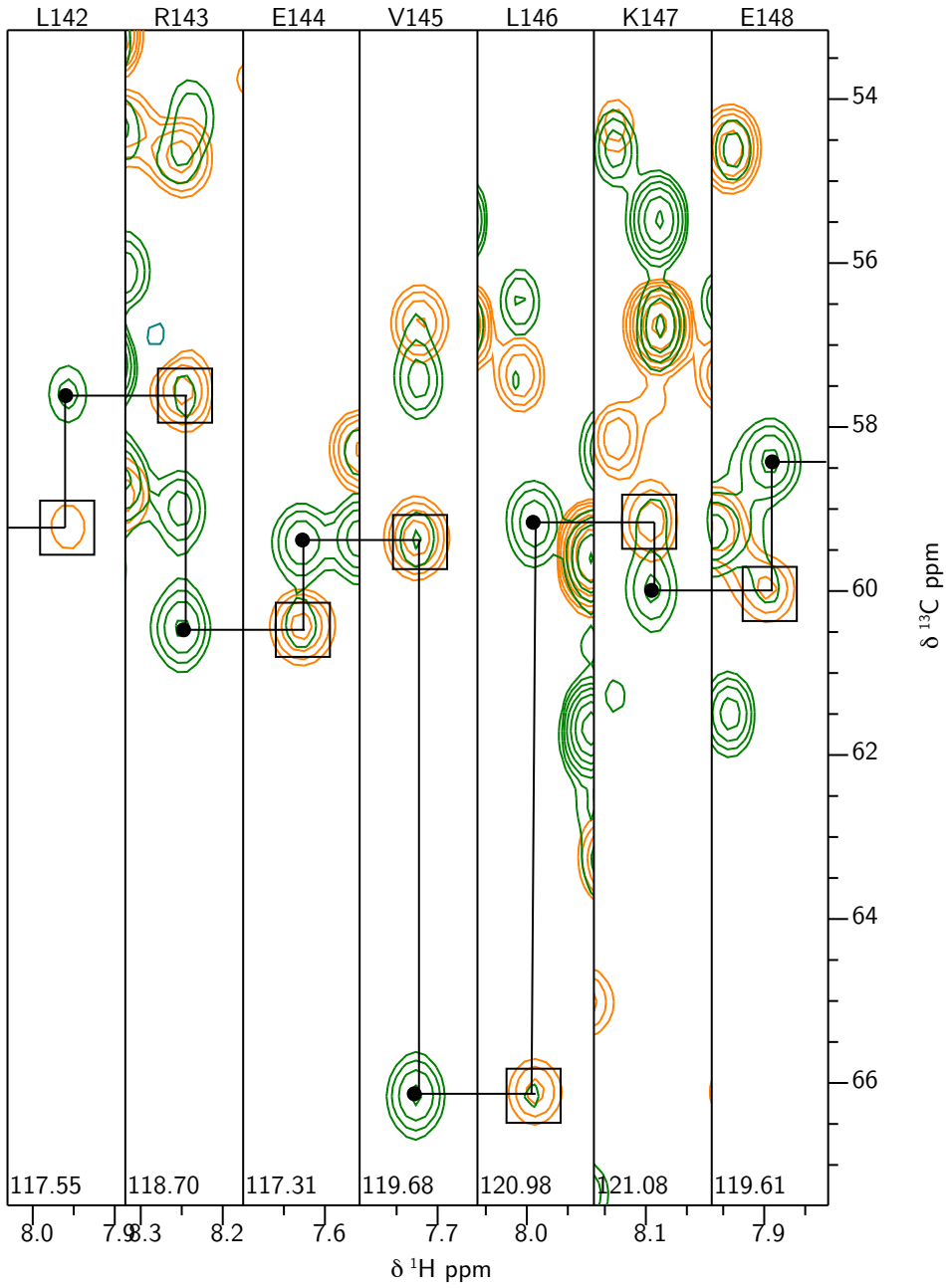


Figure 4.8: Cα Assignments and Connections - Overlay of HNCA (green) and HN(CO)CA (orange) experiments for a stretch of residues. Circles (•) indicate peaks from i and an open square (□) is a peak from residue i-1. Using both experiments together, it is possible to assign Cα chemical shifts and connect spin systems. HNCA gives peaks at both Cα of residue i and i-1. The HN(CO)CA experiment only shows a peak corresponding to the chemical shift of Cα for i-1 (i.e. the previous residue). Data is typical for well defined regions corresponding to the two CUE domains and was used to sequentially connect spin systems. For some, less clear examples, additional experiments were used to confirm connections. Numbers at the bottom of each strip indicate the ^{15}N chemical shift (ppm).

(i.e. residue) type. This combination made it possible to make unambiguous assignments.

CBCANH experiments give two peaks in the Cβ region of the carbon dimension, one corresponding to the Cβ of the residue with the NH group (i) being directly measured and the other from the previous residue (i-1). CBCA(CO)NH, by contrast, only gives a peak from the Cβ of the previous residue (i-1). These experiments also give Cα chemical shifts, which overlapped with HNCA and HN(CO)CA experiments. In some cases, such as serine and threonine, which have Cβ shifts close to the typical Cα range, all spectra were compared to make accurate assignments. There is considerable overlap in much of the data, which made assigning shifts difficult and sometimes impossible. Figure 4.9 shows the coherence transfer diagram for these experiments (236, 239, 240).

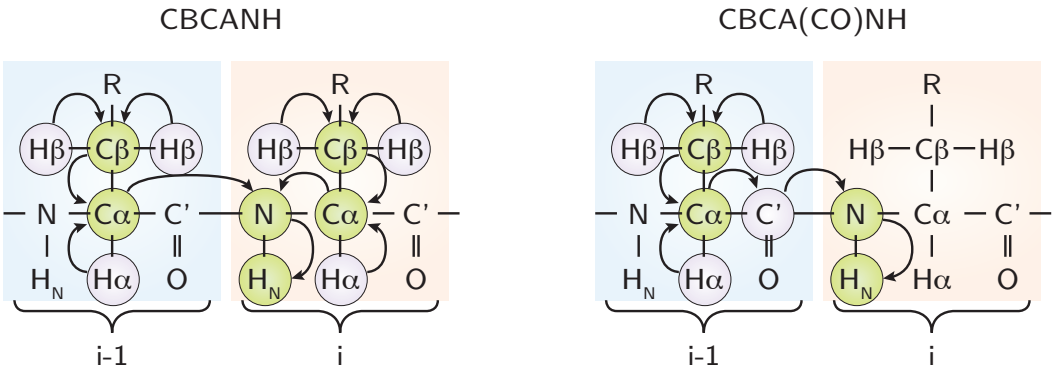


Figure 4.9: CBCANH and CBCA(CO)NH Coherence Transfer Diagram - These experiments give Cα and Cβ shifts for residues i and i-1; and just i-1 respectively (236, 241).

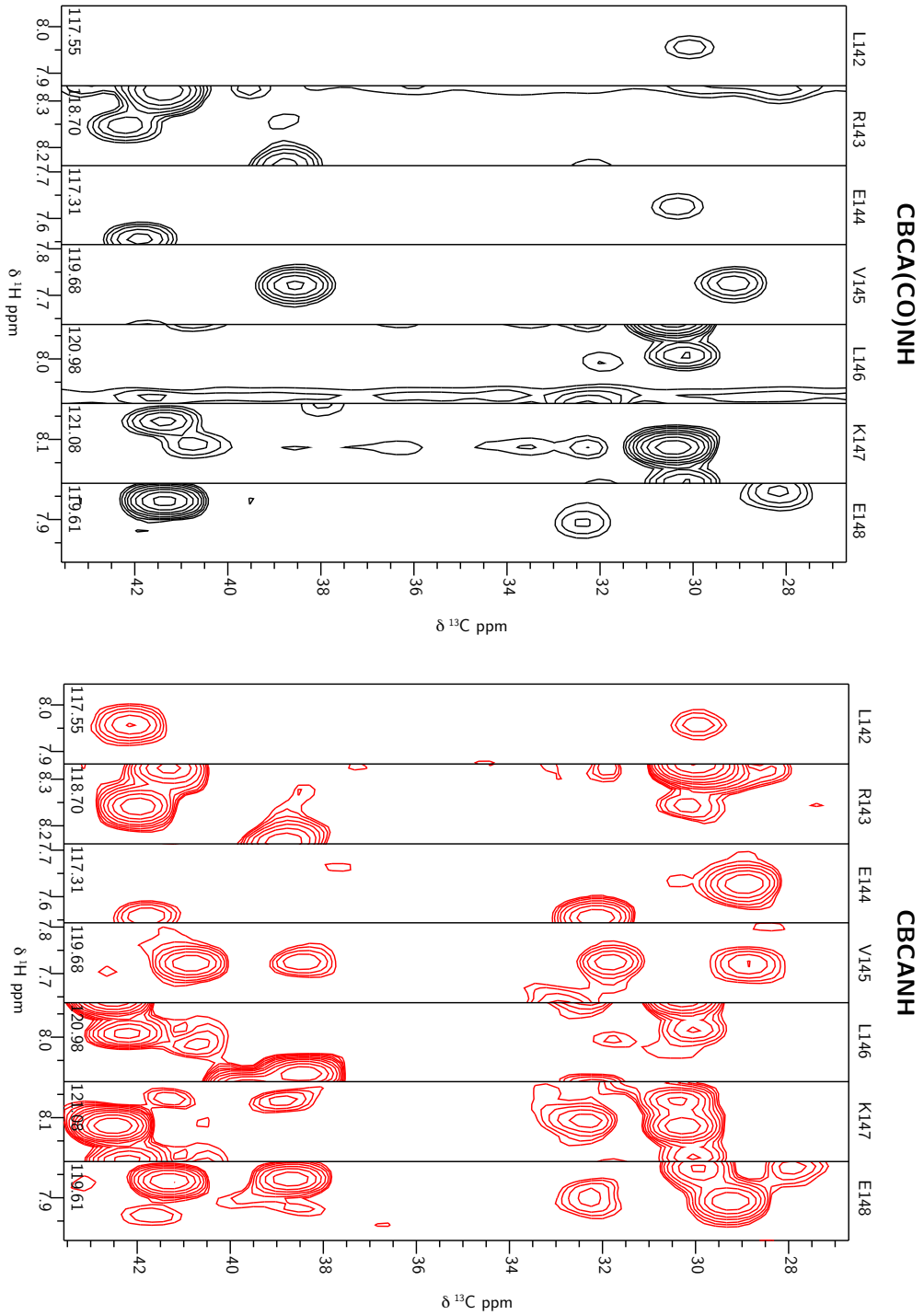


Figure 4.10: Cβ Shift Assignments - Region of **CBCA(CO)NH** spectra showing peaks corresponding to Cβ chemical shifts from residue i-1. **CBCANH** spectra shows peaks corresponding to Cβ chemical shifts from both residues i and i-1. An overlay of these two spectra is shown in figure 4.11. Numbers at the bottom of each strip indicate the ^{15}N chemical shift (ppm).

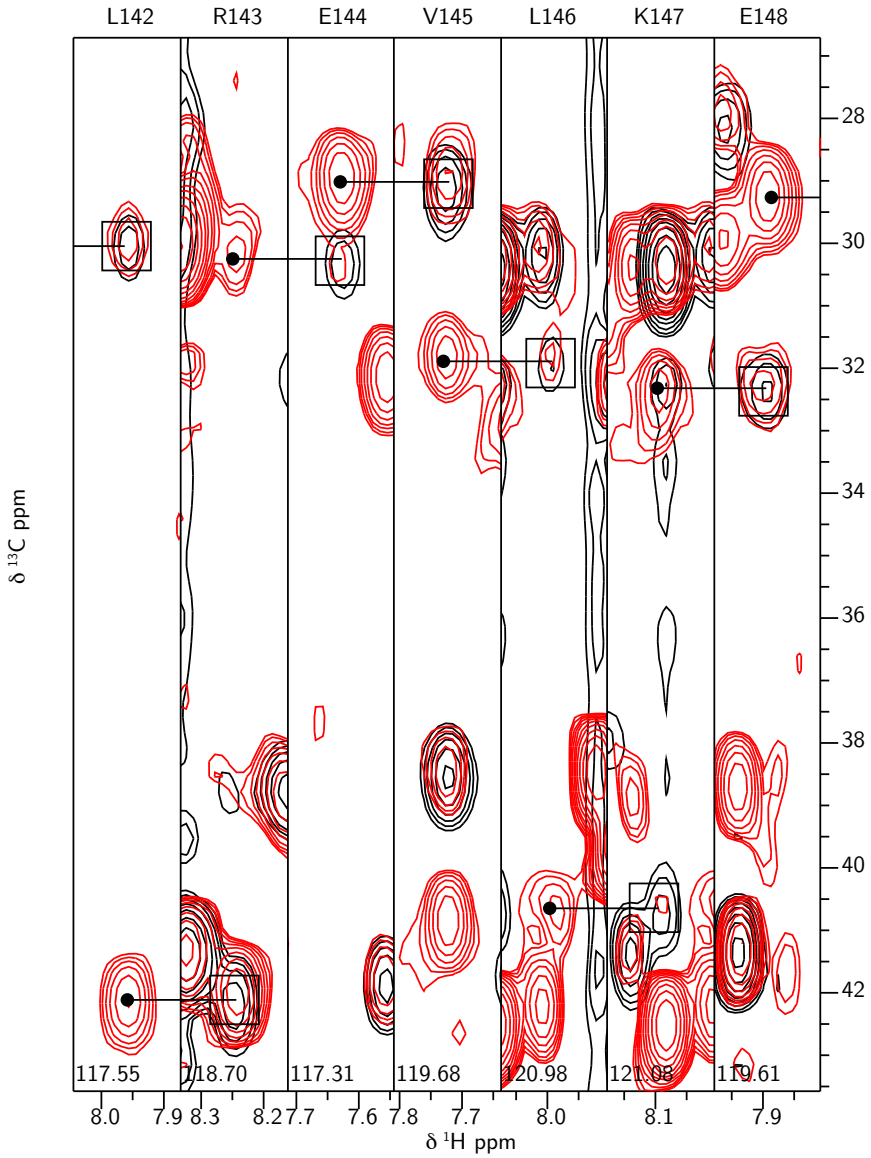


Figure 4.11: Cβ Connections - Overlay of CBCANH (red) and CBCA(CO)NH (black) experiments for a stretch of residues. Circles (•) indicate peaks from i and an open square (□) is a peak from residue i-1. Using both experiments together, it is possible to assign Cβ chemical shifts and connect spin systems. CBCANH gives peaks at both Cβ of residue i and i-1. The CBCA(CO)NH experiment only shows a peak corresponding to the chemical shift of Cβ for i-1 (i.e. the previous residue). In addition to facilitating the linking of spin systems, these experiments were crucial for identifying spin system types. Cβ chemical shifts can vary by a relatively large amount depending on their environment and, in particular, there are characteristic shifts for certain residues. This allowed the possible range of residue types to be narrowed down and eventually assignments made. Data is typical for well defined regions and was used to sequentially connect spin systems. For some, less clear examples, additional experiments were used to ensure connections and determine residue type. Numbers at the bottom of each strip indicate the ^{15}N chemical shift (ppm)

4.3.3.3 CO Assignments

The chemical shifts of the carbonyl carbons, C', in the backbone were assigned using HN(CA)CO (242) and HNCO (240) experiments. The spectra from these experiments result in C' shifts from either both residues i and i-1 or just residue i-1 and therefore a characteristic linking pattern was seen (figure 4.13). Figure 4.12 shows the coherence transfer diagrams from these experiments (236).

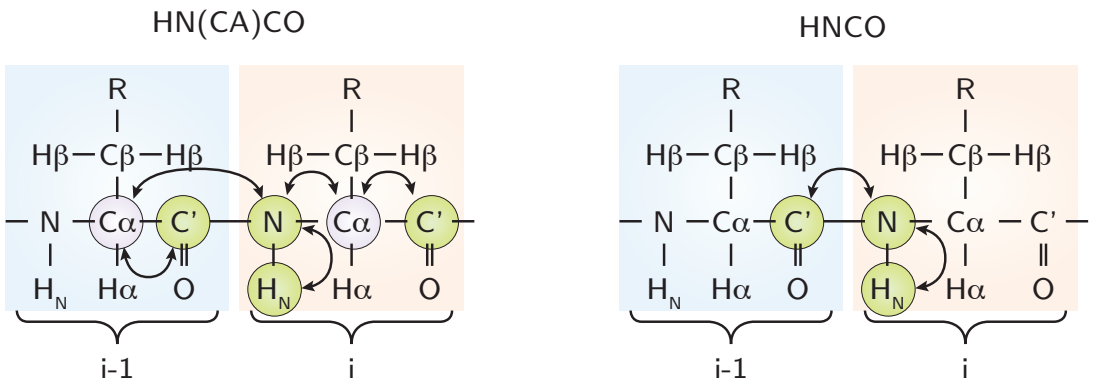


Figure 4.12: HN(CA)CO and HNCO Coherence Transfer Diagram - These experiments give C' shifts for residues i and i-1; and just i-1 respectively (236, 243).

Unlike the experiments used to measure Cα and Cβ shifts, the dispersion in the carbon dimension is fairly low. This meant that these experiments were not particularly useful for determining residue type or unambiguously linking spin systems together but instead served as validation of already linked strips based on other experiments. In other words, the linkages needed to be consistent with those from other experiments but were often not distinct enough to link spin systems on the basis of their C' shift

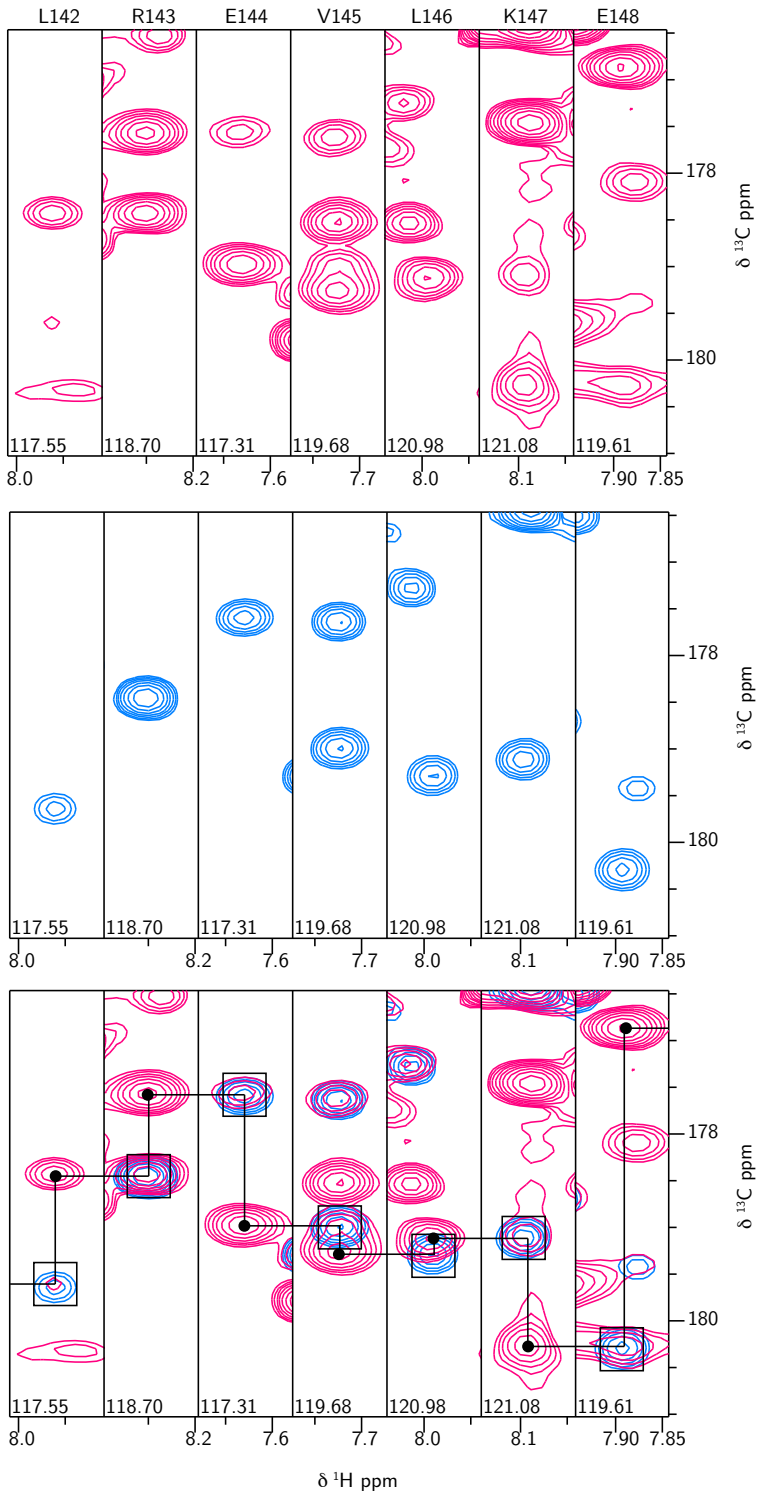


Figure 4.13: Carbonyl Assignments - Top HN(CA)CO experiment gives C' shifts for residues i and i-1. **Middle** HNCO spectrum shows peaks from only residue i-1. **Bottom** Overlay of both experiments with connections between residues shown. Circles (•) indicate peaks from i, whereas an open square (□) is a peak from residue i-1.

alone.

4.3.4 NH-NH

The ^{15}N-NOESY-HSQC experiment is a 3D experiment showing peaks from pairs of nearby protons (figure 4.14). This can be used in sequential backbone assignment as neighbouring NH groups will typically give peaks to one another due to their proximity, especially in helical secondary structure. As proline residues do not have an NH group on the backbone, breaks in sequential NH-NH connectivities occur for proline. These NH-NH NOES were used as an additional way to validate linkages between spin systems, particularly when it wasn't clear from other experiments (figure 4.15) (236).

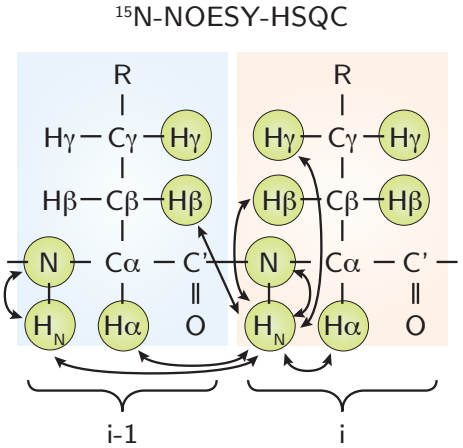


Figure 4.14: ^{15}N-NOESY-HSQC Coherence Transfer Diagram - The ^{15}N-NOESY-HSQC experiment gives through space connections to close protons: NOEs from all ^{1}H to $^{1}H^{N}$. Magnetisation is exchanged between all ^{1}H using the NOE before being transferred to adjacent ^{15}N nuclei and back to $^{1}H^{N}$ for detection (236).

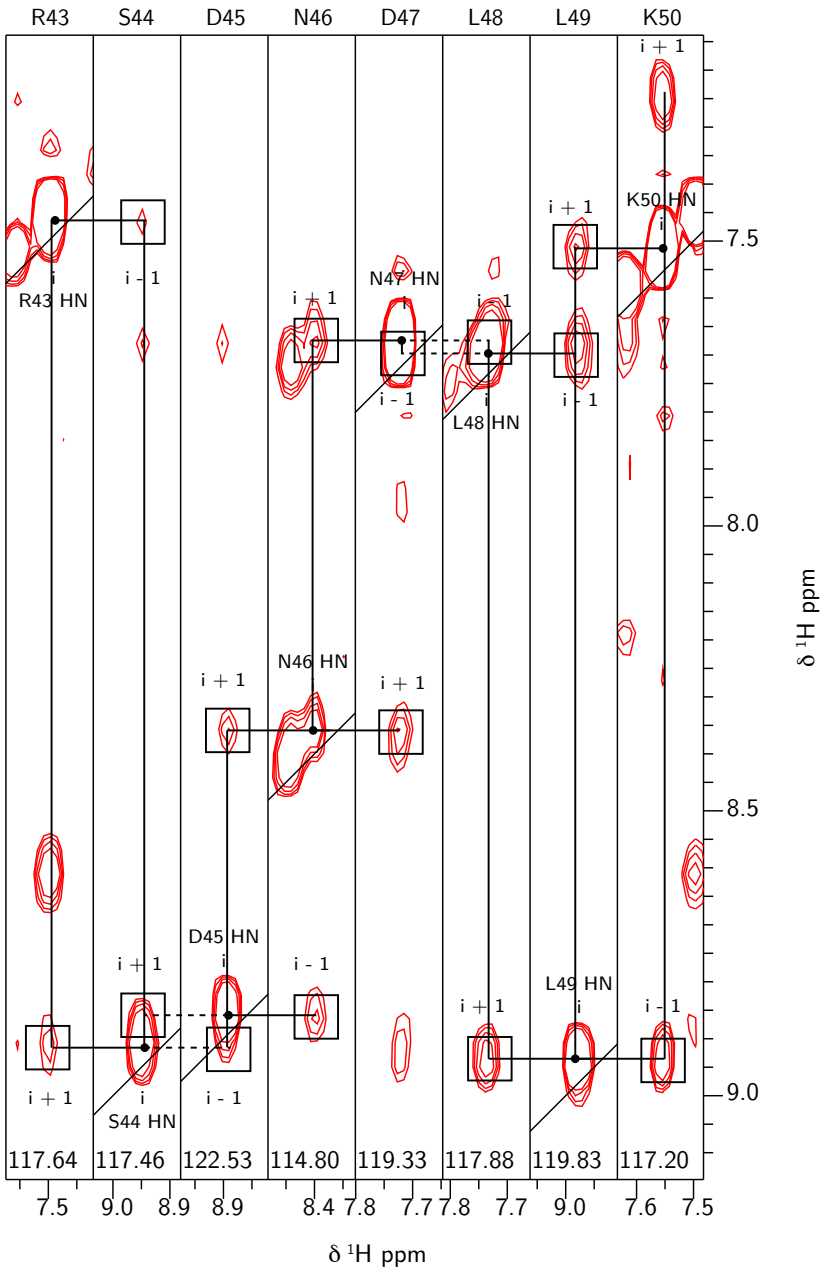


Figure 4.15: NH-NH Connections from ^{15}N-NOESY-HSQC Experiment - $^{1}H^{N}$ region of an ^{15}N-NOESY-HSQC. Peaks from nearby protons appear in strips from linked residues (i-1 and i+1) and give relatively strong peaks, as expected. Circles (•) indicate H^{N} peaks from residue i and open squares (□) are peaks from residues i-1 or i+1. Due to similar $^{1}H^{N}$ chemical shifts, connections between residues 44 and 45, and residues 47 and 48 are not resolved from the diagonal (dashed lines). This is an example of peak overlap creating ambiguity and demonstrates why various different experiments were used to make unambiguous assignments. Numbers at the bottom of each strip indicate the ^{15}N chemical shift.

4.3.5 Final Backbone Assignment

Of the 152 amino acids corresponding to SMARCAD1 dCUE sequence, 129 have unambiguously assigned ^{15}N-^{1}H^{N} chemical shifts (figure 4.16). 19 residues remain partially unassigned, 4 of which are prolines, which do not give peaks in the HSQC because they lack an amide proton. Most of the unassigned residues lie within the unfolded linker region between the two CUE domains, which makes it difficult to unambiguously assign given limited chemical shift dispersion (244). Table 4.1 details the extent of the backbone assignments for the total construct and for each domain individually.

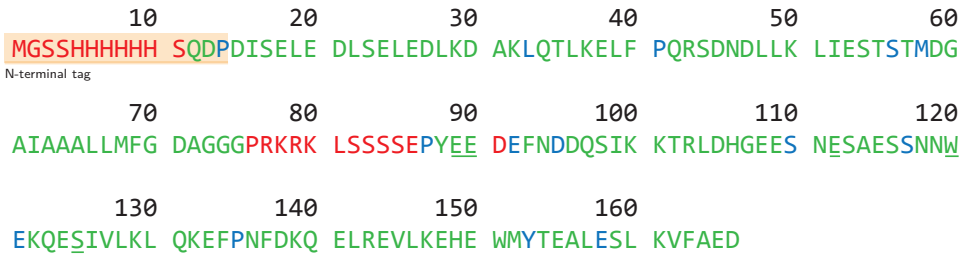


Figure 4.16: Backbone Assignments of SMARCAD1 dCUE - Colour-coded sequence showing completeness of backbone assignments. Residues coloured red have no backbone assignments. Green residues: unambiguously assigned ^{15}N-^{1}H^{N} resonances. Those underlined have some missing Cα / C' assignments otherwise complete backbone assignment. Blue residues are partially assigned but the ^{15}N-^{1}H^{N} has not been assigned (i.e. Cα / C' are assigned). Region shaded in orange is the N-terminal His-tag.

	Percentage Assigned			
Resonance	Full-length (1-166)	CUE-1 (21-71)	CUE-2 (120-166)	Linker (72-119)
^{15}N	77.7%	92.2%	91.5%	64.6%
^{1}H^{N}	80.0%	94.0%	91.3%	67.4%
^{13}Cα	86.1%	100%	97.9%	75.0%
^{1}Hα	83.2%	100%	97.9%	69.2%
^{13}C'	80.1%	100%	93.6%	60.4%

Table 4.1: Extent of Backbone Assignments

Chapter 5

Structure Determination of Double CUE Domains of SMARCAD1

5.1 Introduction

As discussed in section 3.4, SMARCAD1 dCUE is monomeric in solution. This refers to the 19 kDa construct containing two CUE domains. As a tandem domain, one key question surrounding the structure of the double CUE domain was whether the two CUEs interacted with one another to form a single structural unit or whether each CUE was relatively distinct. CUE domains are typically a bundle of three α-helices. The CUE domain from Vps9p (vacuolar protein sorting) protein forms a domain-swapped dimer to bind ubiquitin (169) but other CUE domains are monomeric (section 1.4.3). No structure of a tandem CUE domain has previously been published, although a single CUE domain from the double CUE of yeast protein Cue2 has been shown to be a distinct

three α-helix bundle like other CUE domains (167). In the case of Cue2 the distance between the two CUE domains is just 4 amino acids, as opposed to approximately 50 in the case of SMARCAD1.

According to the Protein Data Bank (PDB), approximately 11.5% of the *ca.* 85,000 protein structures deposited have been determined using NMR (*http://www.pdb.org; September 2012*). Compared to X-ray crystallography, structure determination using NMR is both time consuming and restricted to relatively small proteins. A significant advantage of NMR, however, is that proteins can be studied in solution without the need for crystallisation (245). As discussed in section 3.5, the dCUE construct did not crystallise, despite exhaustive trials. Given that the SMARCAD1 dCUE construct was found to be highly soluble and gave reasonable NMR chemical shift dispersion (section 3.6), the protein structure was determined using NMR.

Figure 5.1 gives an overviews of the process used for determining the structure of the SMARCAD1 CUE domains. A series of NMR experiments were used with either single (^{15}N) or double (^{15}N and ^{13}C) labelled protein. Chemical shift values were assigned to both resonances corresponding to the backbone (as discussed in chapter 4) and those from the side-chains. NOE distance restraints were measured based on relevant peak intensities, dihedral angles predicted from measured chemical shift patterns and residual dipolar couplings (RDC) determined with a specific set of experiments (appendix

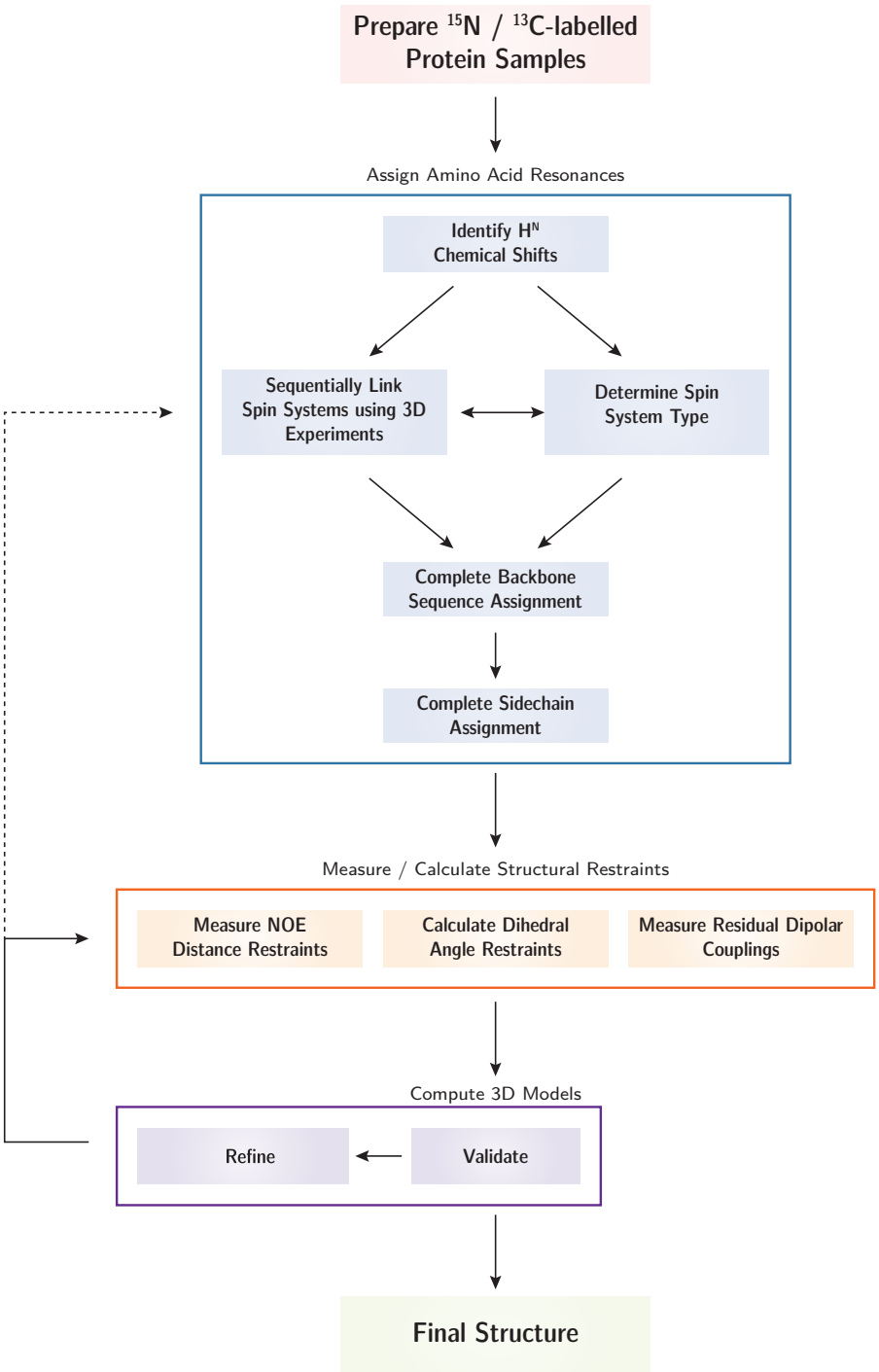


Figure 5.1: Process of Protein Structure Determination using NMR - Structure determination using NMR is an iterative process involving three main stages; assignment of resonances, determination of structural restraints and 3D model computation. This process was repeated until valid, reproducible structures were achieved based on the collected data.

B.3.2.3). Additional information about secondary structure was also incorporated and 3D models computed. This process was iterative and repeated until valid structures were produced. The results of this process are presented below.

5.2 Side-chain Assignments

5.2.1 Hα, Hβ Assignments

Hα and Hβ chemical shifts were assigned using the HBHA(CO)NH experiment (241). This spectrum gives Hα and Hβ shifts corresponding to the previous residue (figure 5.2). Figure 5.4 shows assigned resonances with Hα and Hβ shifts in characteristic ranges (246). Hβ shifts are particularly indicative of residue type and were used to validate and assist some of the backbone assignment (chapter 4). Some regions of the spectra were less well resolved, containing overlapping peaks. Use of an ^{15}N-edited TOCSY (figure 5.3), which shows Hα and Hβ (and other ^{1}H) peaks in the same strip as the ^{15}N-^{1}H^{N} shifts (i.e. residue i) made it possible to assign some of these more ambiguous resonances as the same shifts would be seen in neighbouring strips (figure 5.5).

5.2.2 Additional Side-chain assignments

Remaining side-chain assignments were made using the HCCH-TOCSY experiment. This correlates all protons within a spin system through ^{13}C nuclei (figure 5.6), pro-

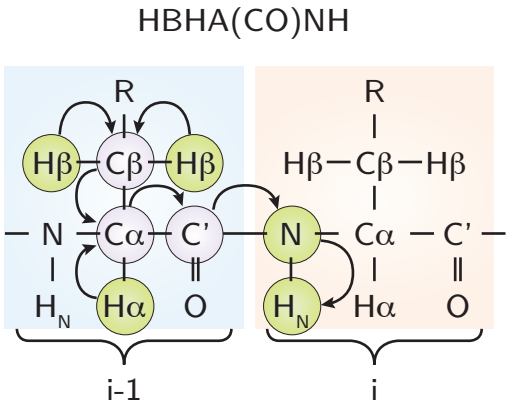


Figure 5.2: HBHA(CO)NH Coherence Transfer Diagram - Magnetisation is transferred from ^{1}Hα and ^{1}Hβ to the carbon atoms (but not evolved on the carbon atoms) and then onto the $^{1}H^{N}$ via the $^{15}N^{N}$ (241, 247).

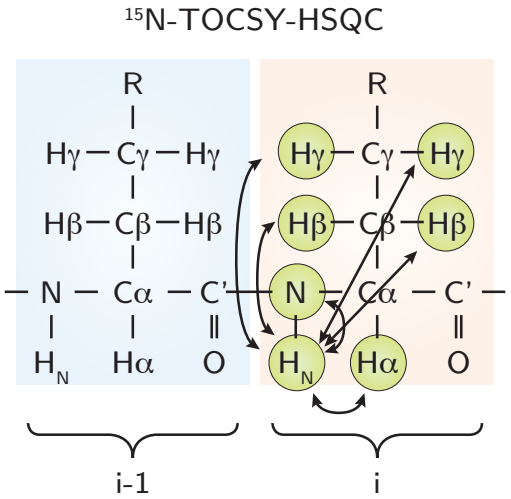


Figure 5.3: ^{15}N-TOCSY-HSQC Coherence Transfer Diagram - Magnetisation is transferred between ^{1}H and then onto the ^{15}N and back to ^{1}H, where it is detected (247, 248).

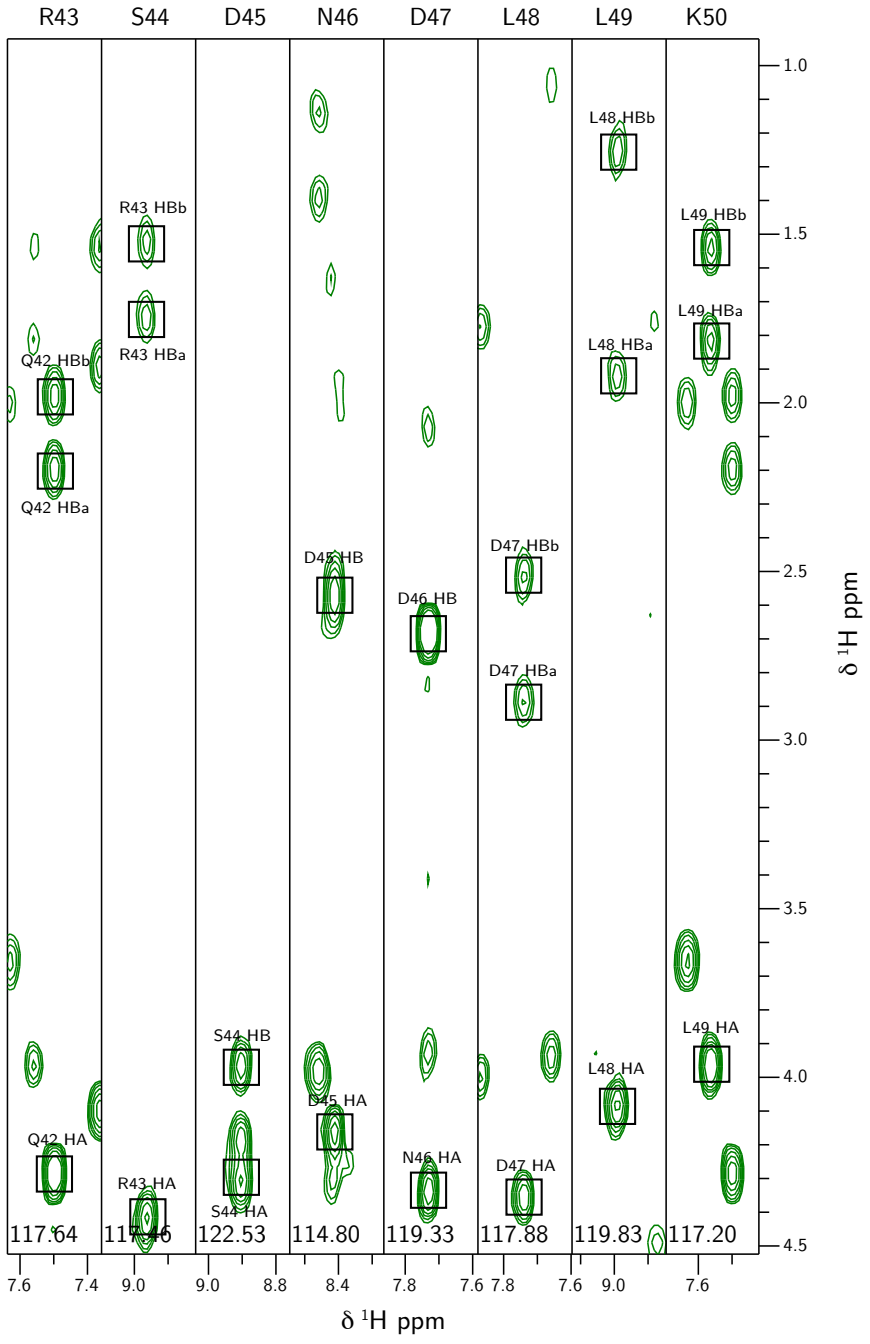


Figure 5.4: Assignment of Hα and Hβ Shifts - The HBHA(CO)NH experiment shows peaks corresponding to the chemical shifts of Hα and Hβ for residue i-1 (i.e the previous residue). Open squares (□) highlight assigned peaks corresponding to these peaks. Hα and Hβ have characteristic shifts depending on residue type, which also helped in backbone assignments (chapter 4). Numbers at the bottom of each strip indicate the ^{15}N chemical shift (ppm).

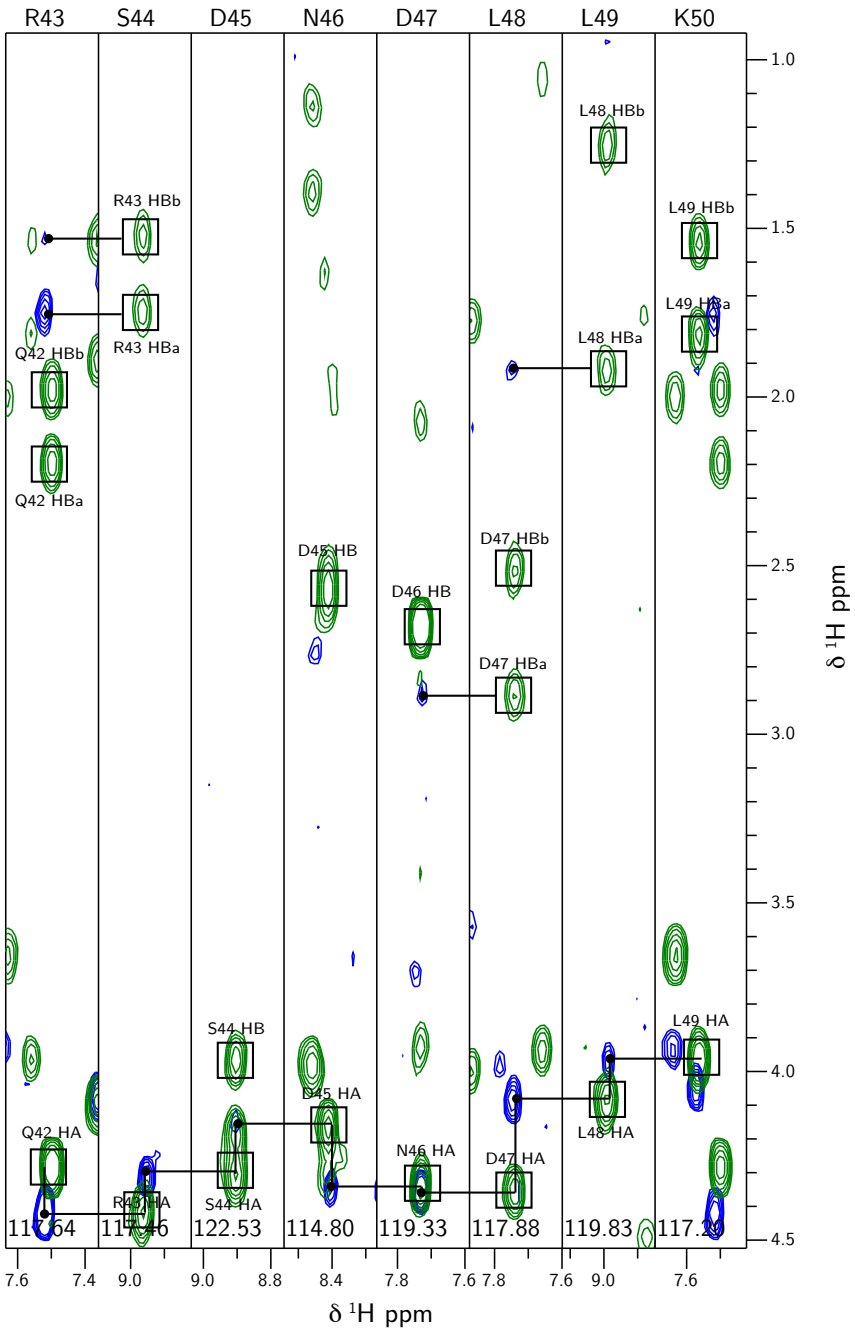


Figure 5.5: HBHA(CO)NH and ^{15}N-Edited TOCSY - Overlay of HBHA(CO)NH (green) and ^{15}N-edited TOCSY (blue) spectra for a stretch of connected residues. Circles (•) indicate peaks from residue i and an open square (□) is a peak from residue i-1. Using both experiments together for already connected strips helped assign peaks in crowded regions of the spectra as Hα and Hβ peaks should appear at the same shift in both spectra. This overlay was also used to confirm some connections in backbone assignment (chapter 4). Numbers at the bottom of each strip indicate the $^{15}N^{H}$ chemical shift (ppm).

ducing a spectrum with chemical shift information for most ^{13}C and ^{1}H nuclei. Figure 5.7 shows HCCH-TOCSY strips for Isoleucine 126, which gives well resolved peaks in the spectrum.

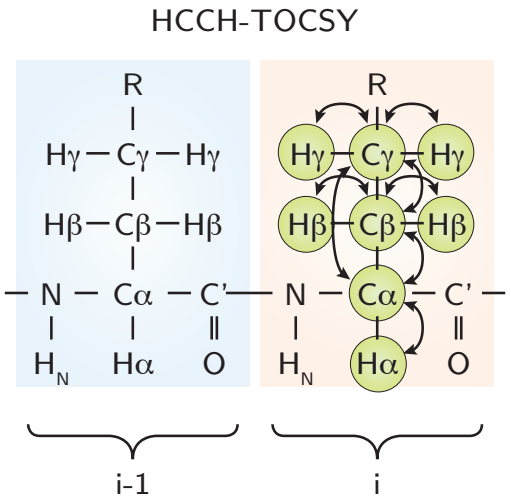


Figure 5.6: HCCH-TOCSY Coherence Transfer Diagram - Magnetisation is transferred from side-chain ^{1}H to ^{13}C, isotropically mixed and transfered back to ^{1}H, where it is detected (247, 249, 250).

5.2.3 Summary of Assignment Statistics

Table 5.1 gives a summary of the side-chain assignment statistics.

5.3 Secondary Structure

As discussed in section 3.2.2, sequence analysis of SMARCAD1 dCUE predicted the protein to be alpha helical, with two groups of three alpha helices, as would be expected with CUE domains (section 1.4.3). Several key experimental results confirmed this to be the case.

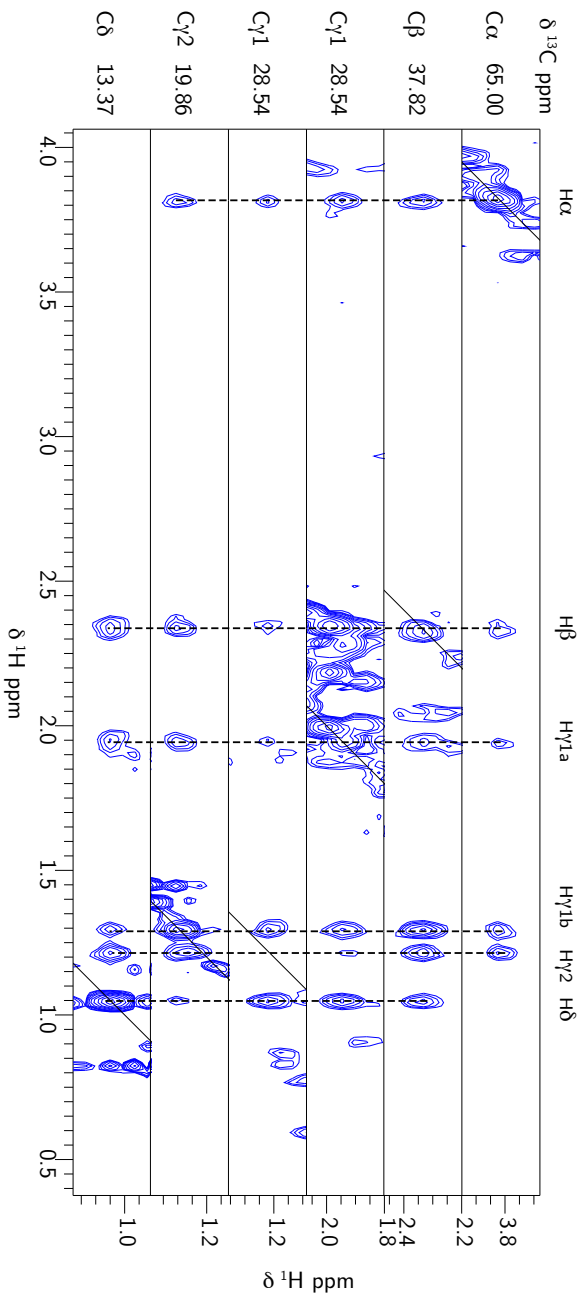


Figure 5.7: HCCH-TOCSY Strips for Isoleucine 126 - Strips showing resonances from isoleucine 126. Coherence is transferred along the spin system, which allows strips for the same spin system to be connected and identify peaks belonging to the same spin system (dashed lines). Peaks at the diagonal give the ^{1}H chemical shift for the proton connected to the ^{13}C of the carbon plane and ^{1}H shifts from other nuclei in the spin system appear at their corresponding shift along the strip.

	Percentage Assigned			
Resonance	Full-length (1-166)	CUE-1 (21-71)	CUE-2 (120-166)	Linker (72-119)
$^{1}H\beta$	80.4 %	100 %	95.4 %	57.1 %
$^{13}C\beta$	86.2 %	100 %	97.9 %	72.7 %
$^{1}H\ \gamma$	53.7 %	64.6 %	50.0 %	40.5 %
$^{13}C\gamma$	42.5 %	56.1 %	42.6 %	29.4 %
$^{1}H\delta$	43.1 %	65.0 %	52.6 %	11.8 %
$^{13}C\delta$	38.3 %	60.5 %	37.2 %	15.4 %
$^{1}H\epsilon$	19.7 %	35.3 %	24.1 %	6.3 %
$^{13}C\epsilon$	18.4 %	50.0 %	42.1 %	14.3 %
Sidechain H	64.0 %	82.0 %	68.9 %	46.3 %
Sidechain non-H	46.9 %	66.2 %	47.5 %	34.6 %

Table 5.1: Extent of Side-Chain Assignments

5.3.1 Characteristic Alpha Helical NOE Pattern

Within an alpha helix, a characteristic pattern of H^{α}-H^{N} NOE connections (through-space) can be seen (figure 5.8). This is a result of the intrinsic structure of an alpha helix. For a given residue (i), the H^{α} nuclei is relatively closer to the H^{N} nuclei from residues i+1 and i+3 than i+2. The distance between H^{α}_{i} and H^{β}_{i+3} in a helical turn is approximately 2.5 - 4.5 Å which also gave characteristic NOE connections in a ^{13}C-edited NOESY spectrum. NOE peak intensity is dependent on magnetisation transfer, which drops off with increased distance ($\propto {}^{1}/_{r^{6}}$, where r is the internuclei distance) (251).

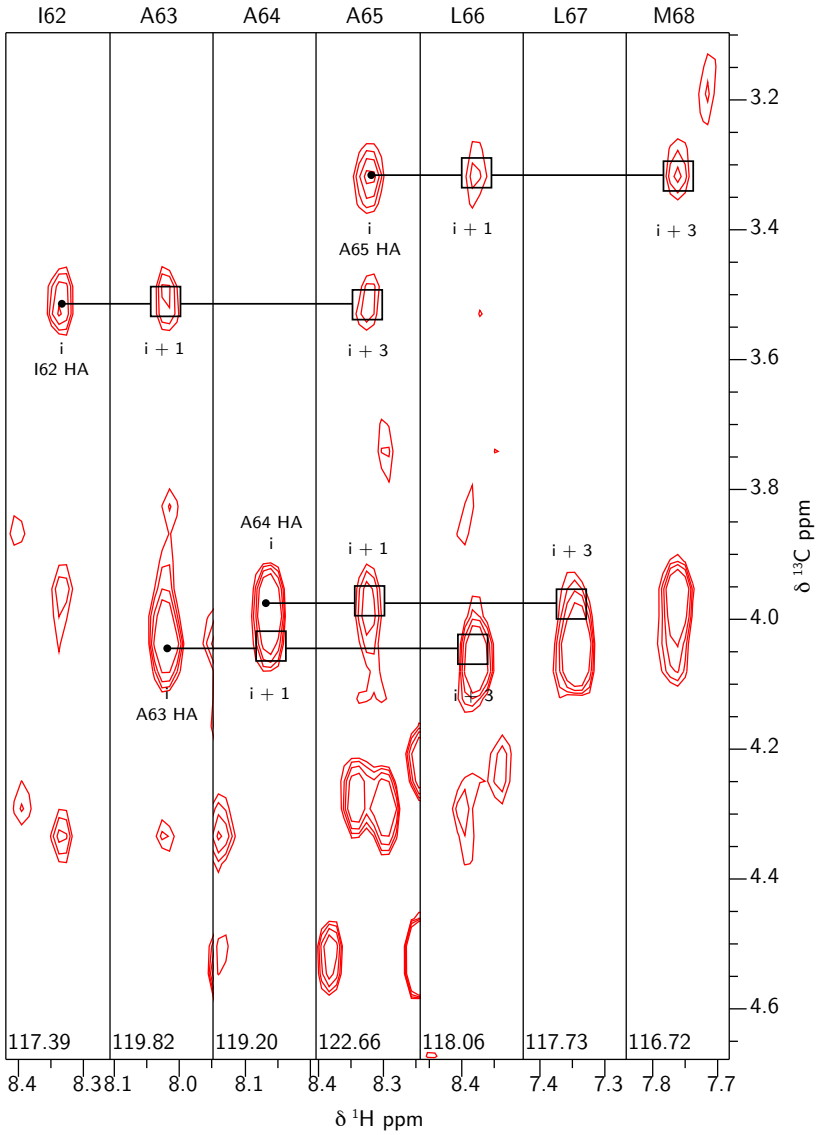


Figure 5.8: NOE Connections Within an Alpha-helix - ^{15}N-NOESY-HSQC spectrum showing Hα chemical shifts within an α helix. Circles (•) indicate H^{α} peaks from residue i and open squares (□) show how these peaks also appear on strips corresponding to H^{N} residues i+1 and i+3. This is a characteristic NOE pattern seen in alpha helices as Hα are close to the H^{N} from residues i+1 and i+3 (*ca* 3.5 Å) but not i+2 and can transfer magnetisation through space. For some residues, the peaks are not always resolved and it is somewhat ambiguous (252).

5.3.2 Dihedral Angles

Backbone torsion angles ϕ and ψ are key in defining protein structure and, in particular, regions of secondary structure (253, 254). Local protein structure greatly influences NMR chemical shifts. Analysis of chemical shift differences by residue types compared to a reference database can accurately predict regions of secondary structure (204). Two different software algorithms, TALOS+ (section 5.3.2.1) and DANGLE (section 5.3.2.2), were used to analyse chemical shifts and predict regions of secondary structure (204, 255). Dihedral angles predicted using TALOS+ were subsequently used in the structural calculations (section 5.7 and appendix A.2).

5.3.2.1 TALOS+

TALOS+ utilises an artificial neural network and empirical relationship between chemical shift values and protein structure to make dihedral angle predictions based on supplied chemical shift values. Chemical shifts for all assigned ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{1}H^{\alpha}$ and $^{1}H^{N}$ nuclei were used. Two clusters of three helices were predicted, which corresponds to the two CUE domains. TALOS+ also predicted some short helical structure in the middle of the linker region (figure 5.9) (204).

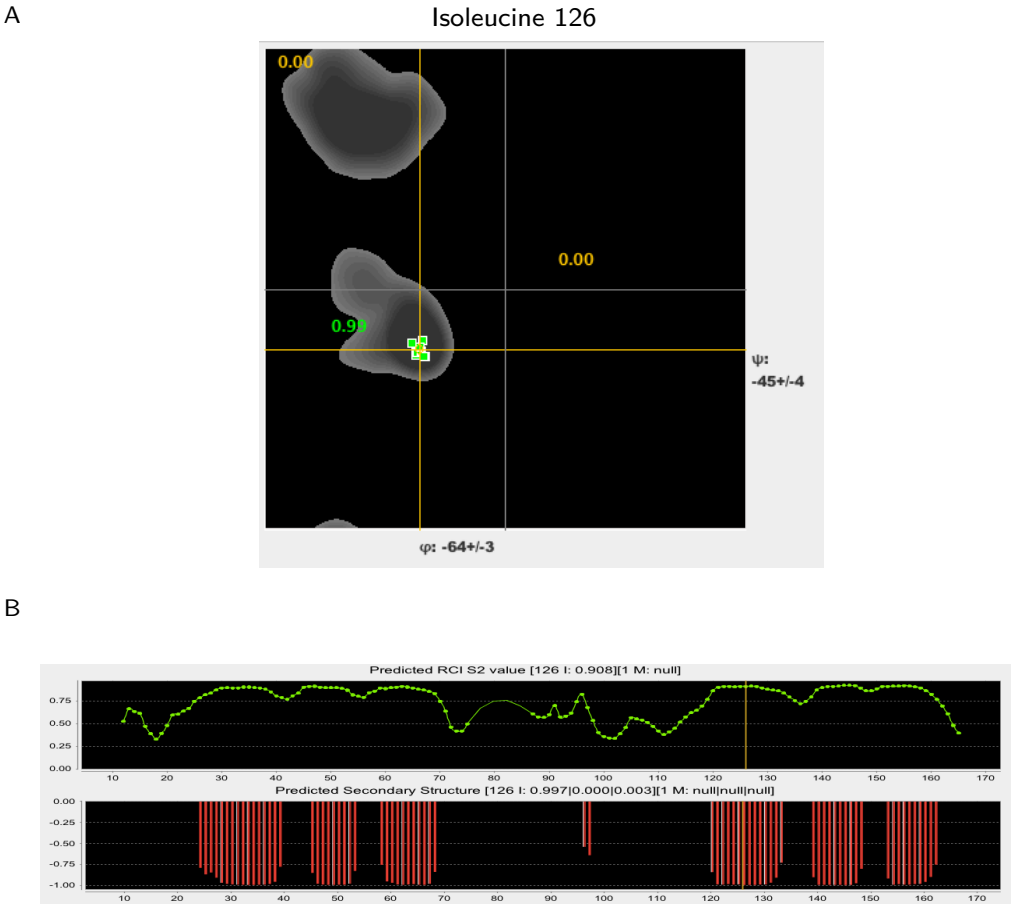


Figure 5.9: TALOS+ Results Indicate Two Domains of Three Helices Separated by an Unstrucutred Linker - **A** Distribution of ten closest database matches for Isoleucine 126 on a Ramachandran plot. Peaks are clustered together, with a high confidence ϕ / ψ prediction characteristic of an alpha-helix. **B** Predicted secondary structure as a function of residue number. TALOS+ predicts two clusters of three helices with high confidence (red bars). Green plot shows RCI-S^2, a measure of rigidity based on random coil index (RCI), derived from chemical shift values (205).

5.3.2.2 DANGLE

Dihedral angles from global likelihood estimations (DANGLE) is an alternative algorithm to TALOS+, which utilises Baysesian inference (a probability technique) to assess possible ϕ / ψ values. This analysed likelihood of 10° bins of possible angles against SMARCAD1 dCUE chemical shift data (figure 5.10). The DANGLE algorithm uses a database generated using five-residue fragments, longer than TALOS+ which uses three residue fragments. Results are grouped as islands of predicted angles on a Ramachandran-like GLE diagram (255). DANGLE results also confirmed a helical structure.

5.3.3 Helix-capping and β-Turn Prediction

In addition to dihedral angle and secondary structure prediction, helix capping motifs and the probability of certain β-turns were predicted based on chemical shift data. The MICS (motif identification from chemical shifts) software package (256), which is related to TALOS+ (section 5.3.2.1), was used. Like TALOS+, MICS also utilises an artificial neural network-derived empirical relationship between chemical shift values and structural features (256). For SMARCAD1 dCUE, MICS confirmed the limits of the helices and ruled out most types of β-turns between helices (figure 5.11).

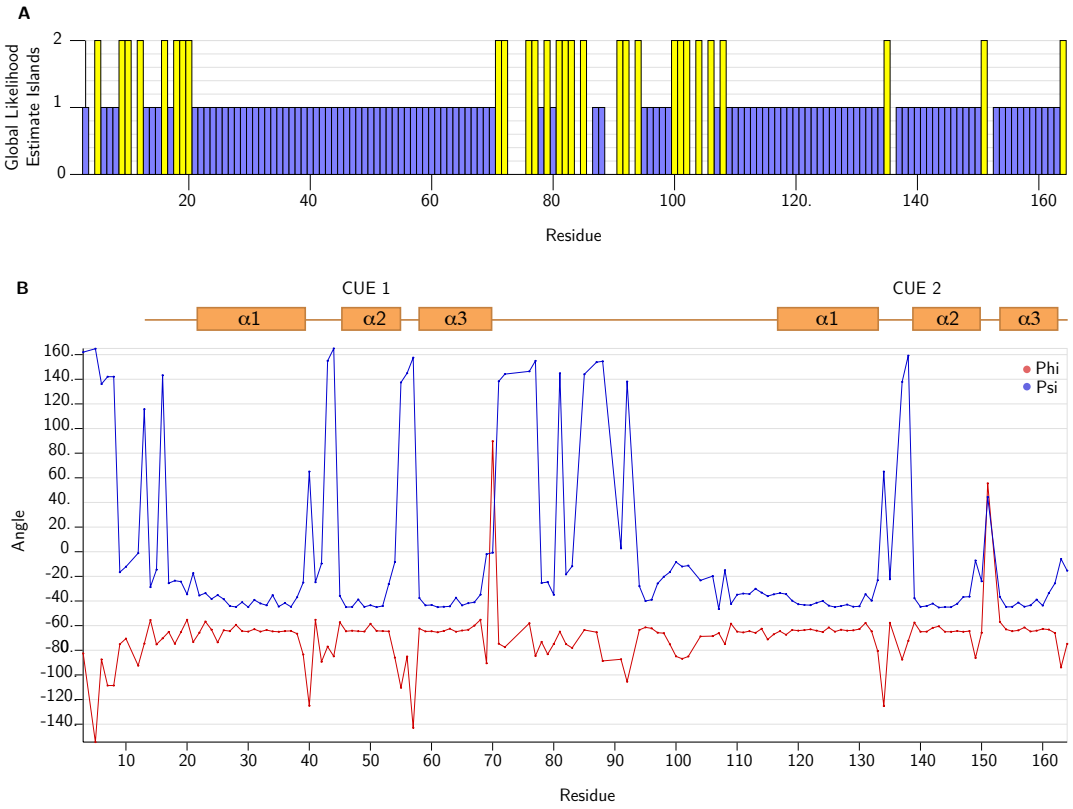


Figure 5.10: DANGLE Results Suggest Helical Structure of two CUE Domains. - **A** Multiple ϕ / ψ angle predictions for any given residue can be grouped based on their clustering on a Ramachandran plot. Discrete clusters are *global likelihood estimate (GLE) islands*, which arises from the variety in ϕ / ψ angle predictions for a particular residue. Predictions are made by comparing secondary chemical shift values (i.e. the difference in chemical shift compared with random coil values) with a database of torsion angle and chemical shift data. Each discreet island represents a grouping of similar predicted angles for any given residue. Residues that give predicted angle values that cluster into more than one *GLE island* may be indicative of backbone flexibility due to an averaging effect (multiple conformations producing a averaged secondary chemical shift that doesn't tightly correlate with a particular dihedral angle) or simply degeneracy (i.e. certain secondary chemical shift motifs could indicate several ϕ / ψ angles). Some residues weren't fully assigned, which added to the ambiguity and correlate with the higher number of GLE islands (particularly a problem in part of the linker region, residues 70 - 95). A smaller numbers of islands indicates that the angles predicted all fall within a narrower range and increase confidence in the prediction (255). **B** Predicted ϕ / ψ angles using DANGLE are similar to those predicted by TALOS+ and fit for helical regions. Backbone conformational angles for a right-handed α-helix are approximately ϕ = -63° and ψ = -43° (253), which compare well with predicted helical regions. Peaks on the chart corresponding to non-helical angles (approximately residues 43, 55, 135 and 150) represent turns between helices. Prediction of ϕ / ψ angles between the two CUE domains is less meaningful as the region is more flexible and unstructured (section 5.4.1).

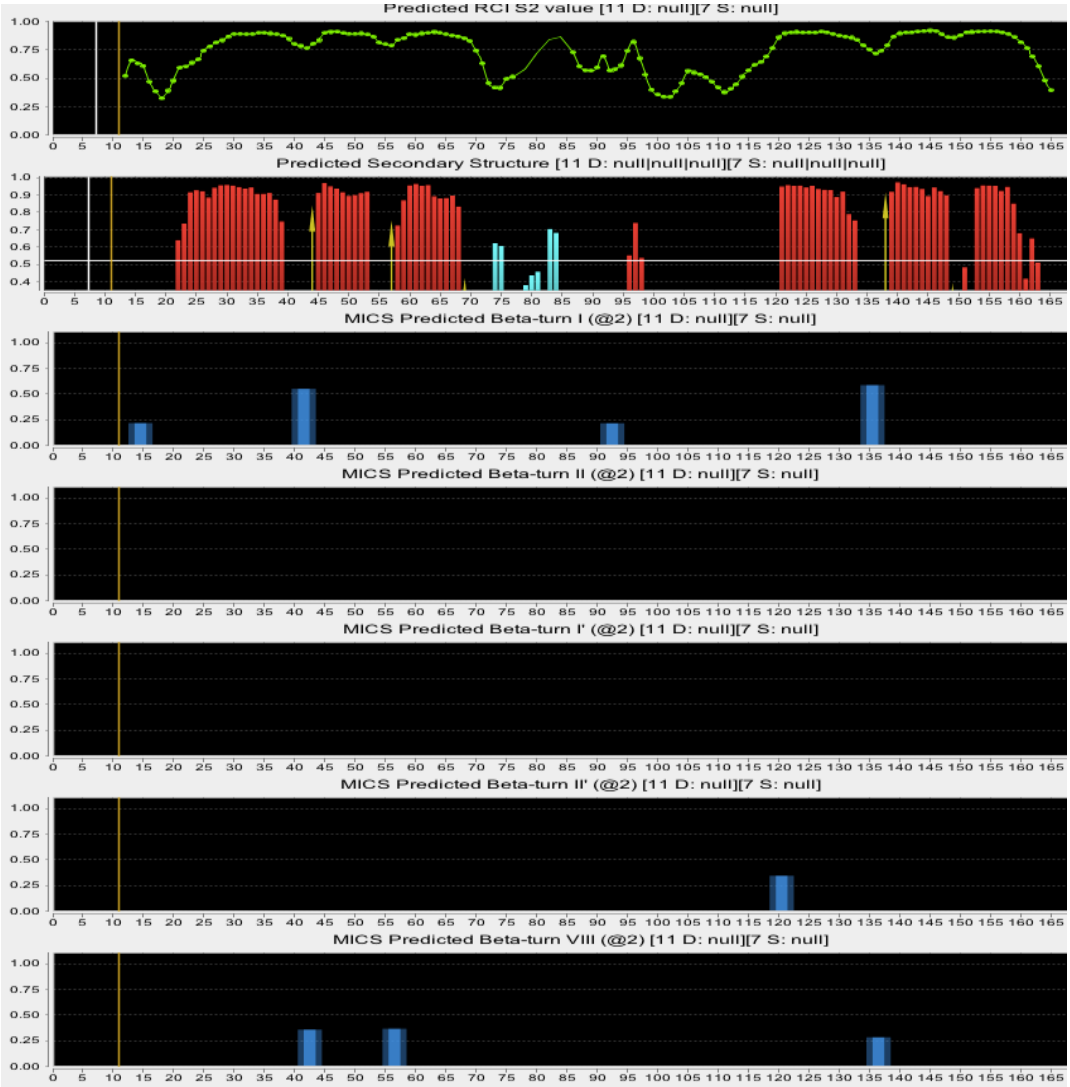


Figure 5.11: Helix Capping and β-Turn Prediction Results - The MICS software predicted several helix capping motifs (yellow arrows) at the end of predicted helical residues (red bars). These correlated well with the expected ends of the helices from other analysis. MICS also uses chemical shift data to identify several β-turn types based on predicted dihedral angles. The dCUE data gave predictions for either a β-turn type I or VIII between the first two helices of each CUE domain. The confidence value (blue bar) is higher for the prediction of a type I turn than it is for type VIII, although this is only just above 0.5. β-turn type I and VIII are similar in that the ϕ/ψ angles of residue i+1 within the four residue turn are very similar (ϕ/ψ approximately -68°/-30° respectively). (256).

5.3.4 Helix Packing

Helical wheels were plotted for each of the dCUE helices (figure 5.12). As CUE domains are helical bundles, the helices have a buried surface. This gave an indication of how the helices may pack together, which was useful for speeding up assignment of long and medium-range NOE connections.

5.4 Backbone Dynamics

Proteins are not rigid structures and many have genuinely flexible regions. During the structural calculation stage, it is important to be aware of mobile regions as their position will be less well defined than more structured parts of the molecule (i.e. the root mean square deviation (rmsd) will be higher). NMR techniques can be particularly useful in identifying and quantifying residue mobility (245).

5.4.1 Heteronuclear NOE

Steady state $\{^1H\}$-^{15}N heteronuclear NOE experiments were used with SMARCAD1 dCUE to study fast timescale residue mobility and, in particular, to investigate the flexibility of the linker region between the two CUE domains. A heteronuclear NOE ratio was measured by collecting two spectra; one with ^1H saturation and the other without it. The spectrum without ^1H saturation resembles an ^1H, ^{15}N-HSQC whereas the spectrum with ^1H saturation has lower peak heights due to the $\{^1H\}$-^{15}N NOE. The

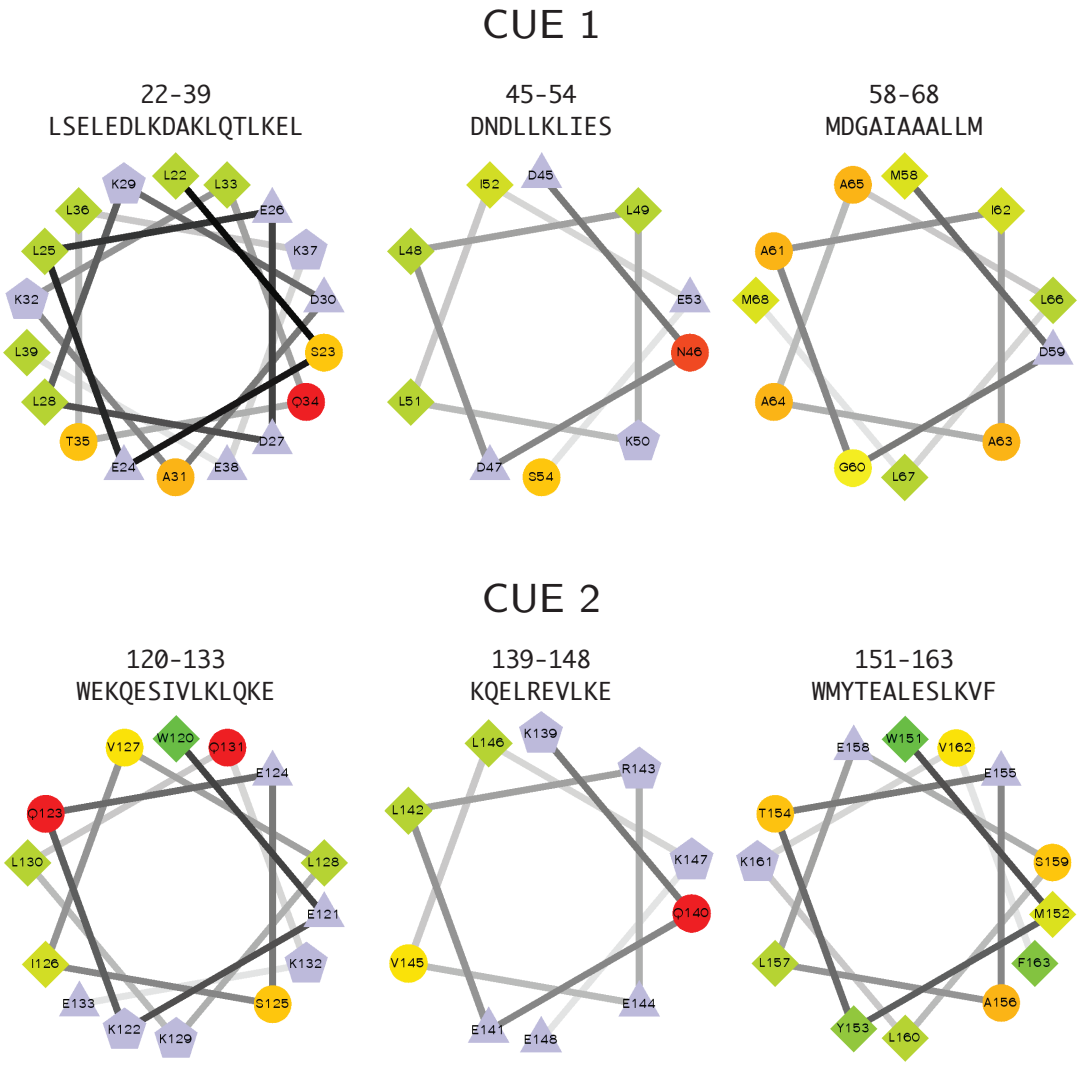


Figure 5.12: Helical Wheels for CUE Domains - Helical wheel representations (257) of residues in the helices of each CUE domain. Sequence and residue number for each helix are shown above the helix wheel. Hydrophobic residues are shown as a green / yellow diamond (◇); green diamonds are more hydrophobic than yellow residues. Circles (○) are hydrophilic residues. Triangles (△) and pentagons (⬠) indicates residues that potentially can be negatively or positively charged respectively. Helix 1 in CUE1 is a striking example of an amphiphilic helix, with clear spatial clustering of hydrophobic leucine residues on one side of the helix. As hydrophobic residues packing together is energetically favourable (to hide their hydrophobic surface) this helped with identification of inter-helical NOEs (i.e. by knowing which side of a helix packs against another). Helical wheels were plotted with the helical wheel tool by Don Armstrong (258).

relative decrease in peak intensity for each peak 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 (appendix B.3.2.2) (245, 259, 260, 261).

A plot of NOE ratio and residue position gives an indication of flexibility along the backbone (figure 5.13). The residues corresponding to each of the CUE domains have a high heteronuclear NOE ratio, which is as expected for fairly rigid globular structure. In contrast, the linker region between the two CUE domains has a very low heteronuclear NOE ratio, signifying considerable flexibility. Given that the linker region has little notable secondary structure (section 5.3.2) and is highly flexible, it seems probable that the linker region between the two CUEs may act as a spacer to separate the two CUE domains, which are essentially independent of one another.

5.5 Residual Dipolar Couplings

Residual dipolar couplings (RDCs) arise from weak alignment of a protein in an anisotropic medium and are useful for structure calculations (appendix B.3.2.3). RDCs can be used to orientate each NH bond relative to a global alignment tensor and thus produce long-range structural information. ^{1}H-^{15}N RDCs were measured from $^{1}J_{\mathrm{NH}}$ splitting in the ^{15}N dimension of an HSQC spectrum collected with the IPAP (in-phase and anti-phase) method (figure 5.14) (236, 262). An anisotropic sample was made

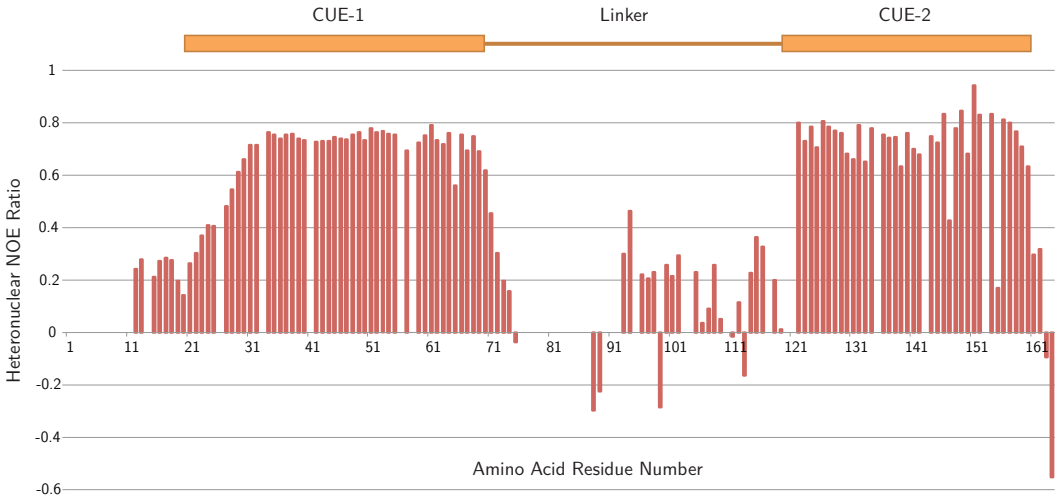


Figure 5.13: Heteronuclear NOE - Plot of heteronuclear NOE ratio and amino acid residue number. The heteronuclear NOE ratio is the ratio of peak intensity between two spectra used in the experiment, one with proton saturation and one without. Higher numbers correspond to higher rigidity such as that found in globular proteins, whereas lower numbers signify highly flexible regions. In the case of SMARCAD1 dCUE, the two globular CUE domains have high heteronuclear NOE ratios whereas residues corresponding to the linker are highly flexible (245). The N-terminal end of CUE-1 (aa21-29) have relatively low heteronuclear ratios. The ratio values gradually decrease for residues further from the centre of the CUE-1 domain whereas the other domain boundaries have transitions from high-low ratio values which occur over just 2-3 residues. This may suggest a dynamic N-terminal structure and is discussed more in section 5.7.1.4.

using 2% w/v $C_{12}E_6$/hexanol in the same buffer as the isotropic sample. Doublet peak separation for each sample was measured and the difference between isotropic and partially ordered samples used to calculate the RDC magnitude and sign. RDC values for all measured residues are shown in appendix A.3 and figure 5.15, with changes in RDC values correlated to structured regions.

Given the considerable length of the linker between the two CUE domains, its lack of secondary structure and high degree of flexibility (section 5.4), it was concluded that the two CUE domains likely behave independently of one another. This suggests that each CUE domain is mobile and can tumble in solution separately. As such, RDC data was separated for each CUE domain and analysed independently. The axial and rhombic components of the alignment tensor, D_a^{NH} and R, (appendix B.3.2.3) are used along with the residual dipolar coupling values for refinement structure calculations. One common way for estimating these values is by using the histogram method (appendix B.3.2.3). Figure 5.16 shows the distribution of RDC values, binned into 2 Hz increments for both domains. From these plots, it is not possible to make reliable estimates for D_a^{NH} and R as the distribution of RDCs is fairly limited and the classic distribution shape is not seen. One possible explanation is due to the helical nature of the protein, which means that NH vectors do not sample a large enough range of orientations in space.

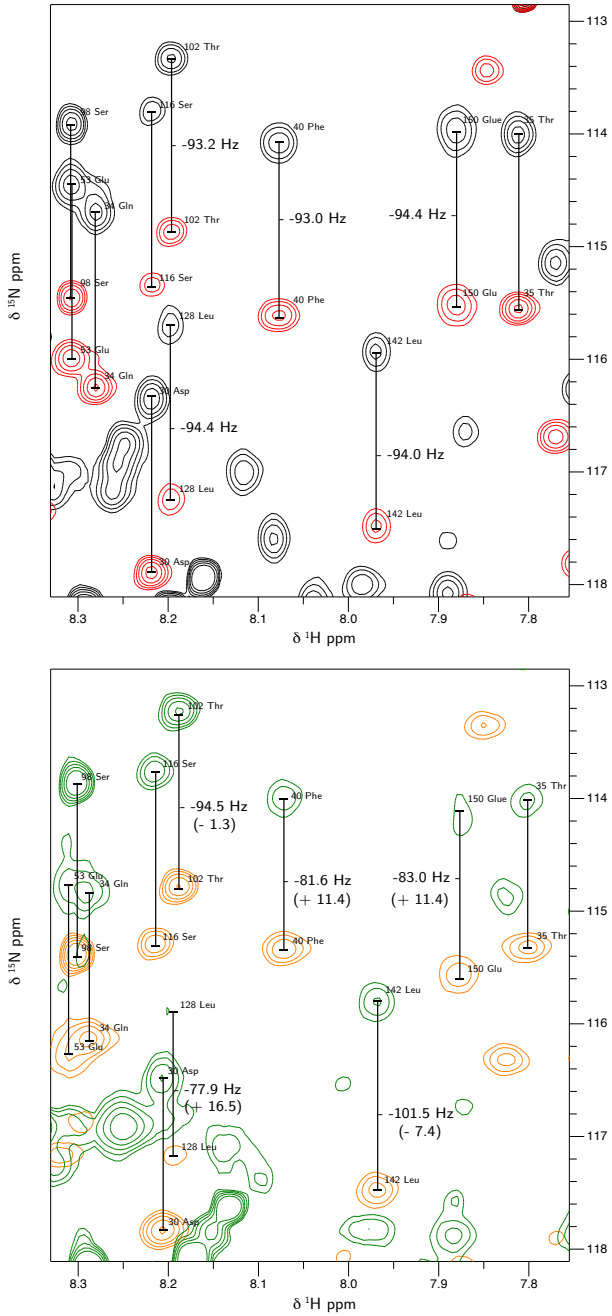


Figure 5.14: Dipolar Couplings from HSQC-IPAP Spectra - RDC values were determined by measuring peak separation in isotropic (**top**) and partially aligned (**bottom**) samples. Splitting in the isotropic solution is uniform (no contribution from RDCs) and corresponds to the $^{1}J_{\mathrm{NH}}$ scalar coupling (magnitude approximately -93 Hz). Upfield and downfield doublet components are shown in red and black respectively. The partially aligned protein sample was made with 2% w/v $C_{12}E_6$/hexanol, which gave different splitting values and thus observable RDCs. Differences in peak separation between the isotropic solution and aligned medium were measured (shown in brackets). Upfield and downfield doublet components for the anisotropic sample are shown in green and orange respectively.

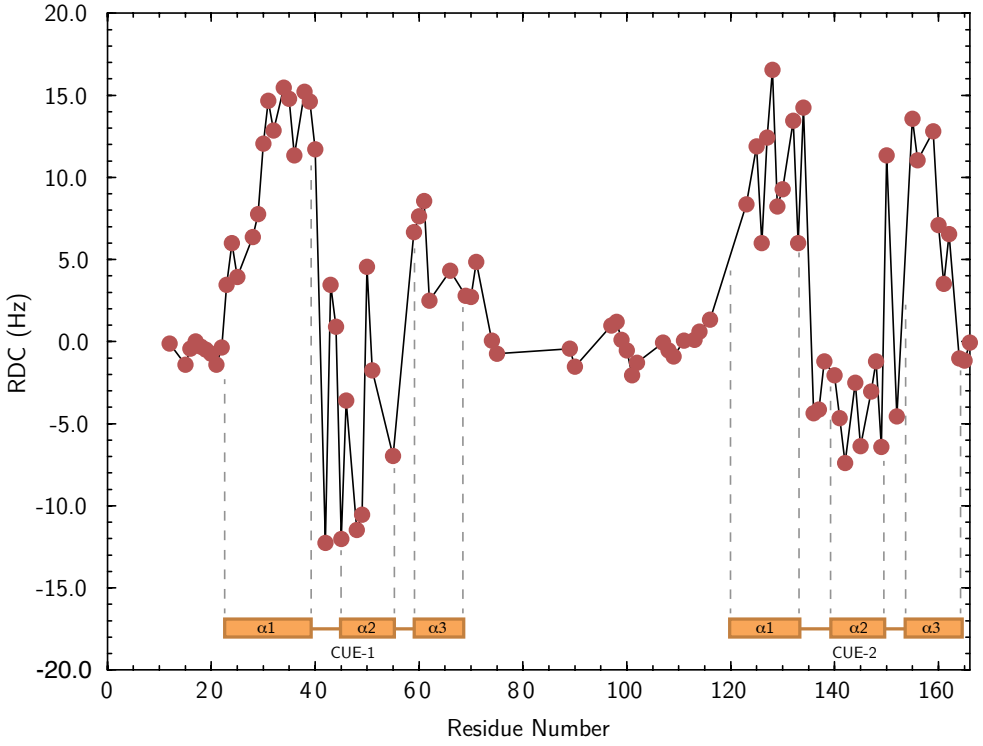


Figure 5.15: Residual Dipolar Couplings - Plot of NH RDC values with dCUE sequence. The CUE domains are marked with individual helices shown. Similar RDC values are observed for all residues within an individual helix as all NH bond vectors are aligned parallel to the helix axis and thus to one another. RDC values corresponding to the second helix, $\alpha 2$, of both CUE domains are substantially different to those for $\alpha 1$ or $\alpha 3$ and negative in sign. This suggests that $\alpha 2$ is orientated along a non-parallel, significantly different plane than that of either $\alpha 1$ or $\alpha 3$ (263).

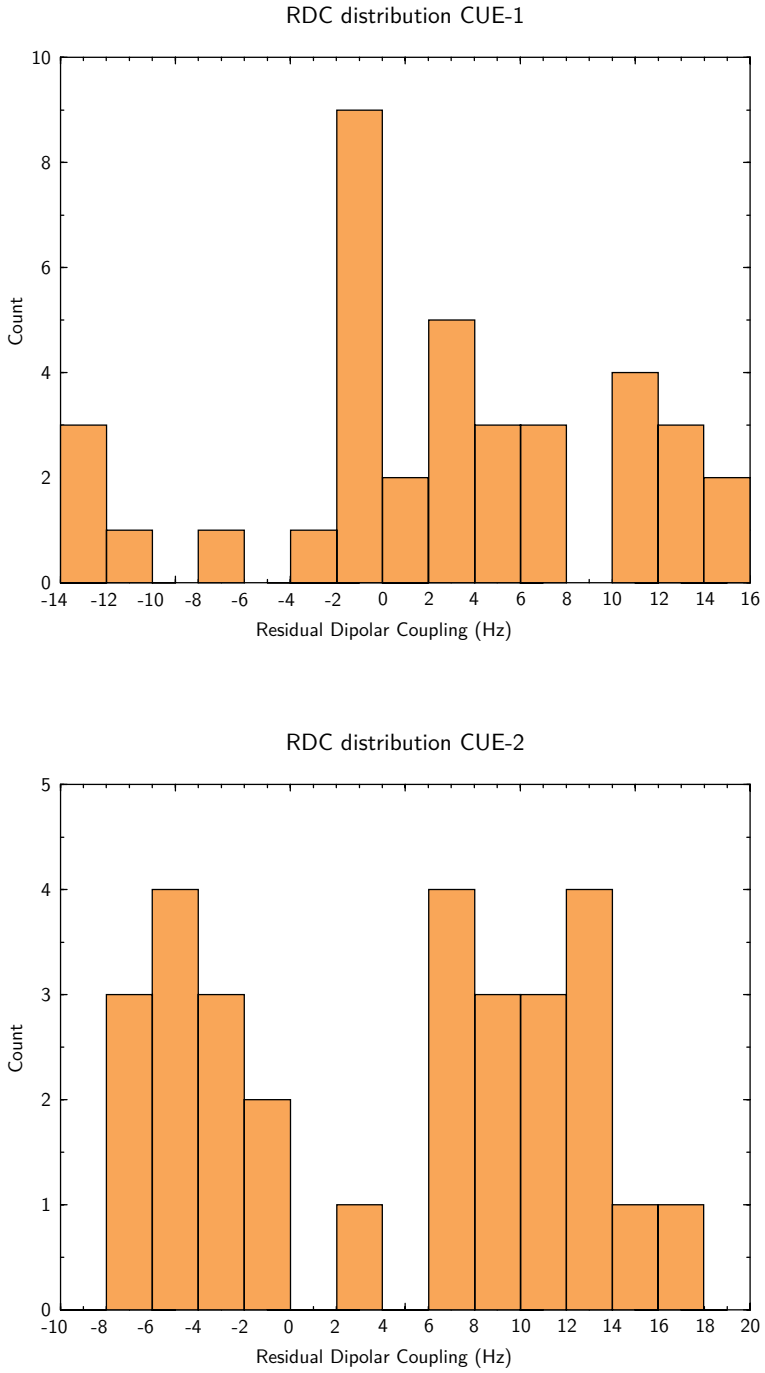


Figure 5.16: RDC Distribution for each CUE Domain - High confidence estimates for D_a^{NH} and R for either CUE domain were difficult to make using the histogram method. This may be a result of the relatively small number of RDCs per domain not producing a complete distribution and the helical nature of the domains restricting the range of orientations adopted by NH vectors (264).

An alternative, homology-based, approach was used to estimate the D_a^{NH} and R values used in structure calculations (section 5.7). The sequence of both SMARCAD1 CUE domains were aligned to that of human CUE Domain Containing Protein 1 (CUEDC1) (section 3.2.1). The published ensemble of 20 NMR structures for CUEDC1 (PDB: 2DHY) was used to calculate RDC values which closest matched the experimentally measured ones for the first two helices of the SMARCAD1 CUE domains. D_a^{NH} and R values for the 10 models with the best fit to experimental data calculated from the structures and averaged for use in structural calculations. (265). D_a^{NH} and R values for CUE-1 were -9.0 and 0.5 respectively. CUE-2 D_a^{NH} and R values were calculated at 6.9 and 0.3 respectively.

5.6 NOE Distance Restraints

Long, medium and short range distance restraints were measured using nuclear Overhauser effect (NOE) through-space peak connections. Peak intensity is inversely proportional to distance ($1/r^6$) and was used to generate restraints between different nuclei. The distance corresponding to particular peak intensities was referenced using several fixed distances on side-chains (266). A ^{13}C-NOESY-HSQC was used to find sidechain NOE connections (figures 5.17 and 5.18).

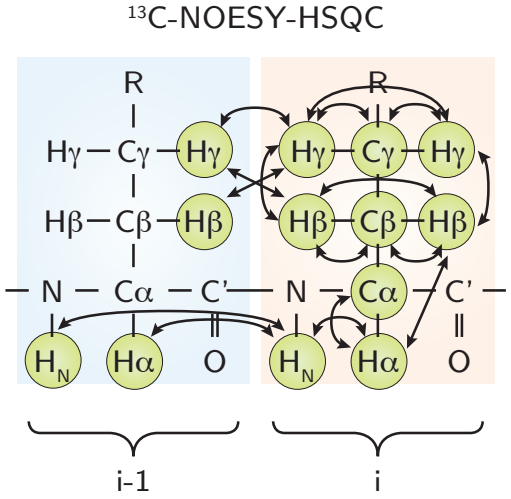


Figure 5.17: ^{13}C-NOESY-HSQC Coherence Transfer Diagram - Magnetisation is exchanged between all protons using the NOE and then transferred to neighbouring ^{13}C nuclei and back to ^{1}H for detection.

5.7 Structure Calculations

No long-range NOE distance restraints were assigned between the two CUE domains and very little secondary structure was identified in the linker region based on chemical shift and NOE analysis. Given that the heteronuclear NOE (section 5.4.1) experiment confirmed that the linker was highly dynamic it was concluded that, in solution, the two CUE domains of the dCUE construct were able to move independently and thus no single, fixed structure for the double CUE domain exists. The data was split for the ordered residues in each CUE domain and structure calculations carried out separately. Subsequently, calculations were run using the full double CUE domain sequence (CUE-1, CUE-2 and linker) to confirm and visualise a range of conformations.

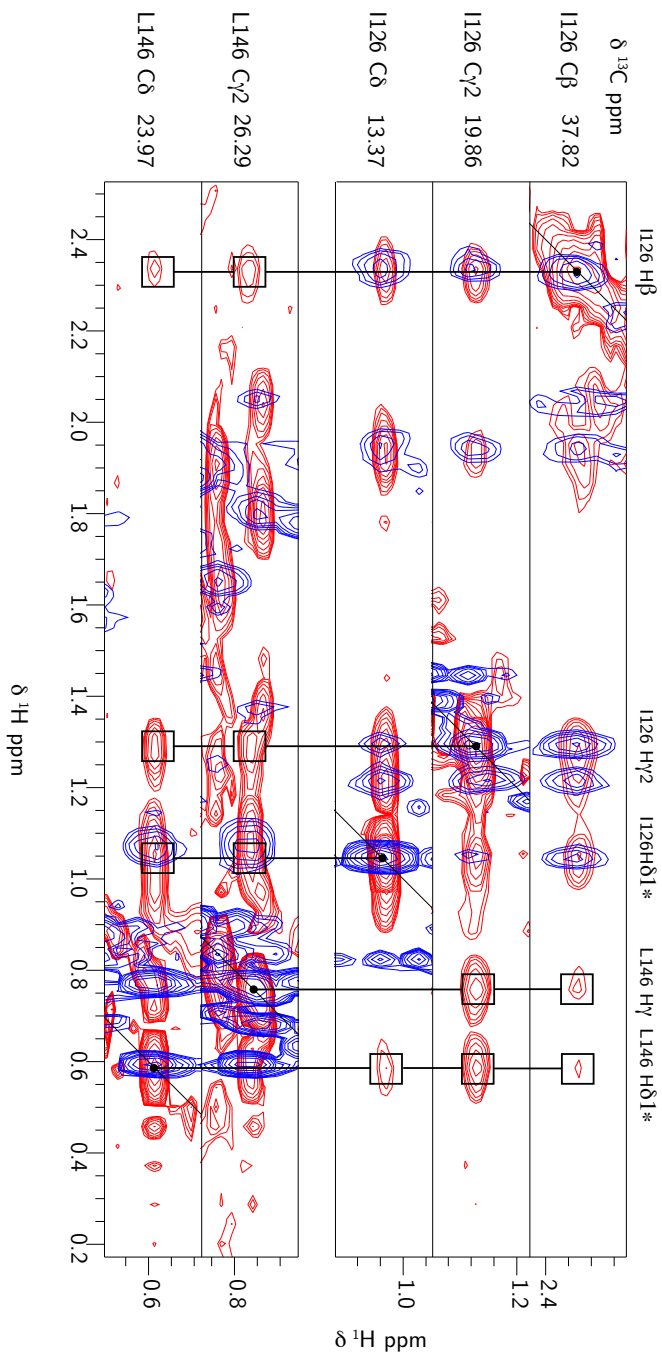


Figure 5.18: Long Range Distance Restraints - A ^{13}C-NOESY-HSQC experiment (red) was used with an HCCH-TOCSY experiment (blue) to identify connections between nuclei, and in particular, those between sidechain nuclei from different residues (long and medium range). HCCH-TOCSY spectra gives peaks to nuclei within a same residue and the ^{13}C-NOESY-HSQC gives peaks to all nearby nuclei, regardless of a through-bond connection. Circles (•) indicate the cross peak of the strips and open squares (□) mark NOE peaks on different residues. NOE connections between I126 and L146 are shown.

5.7.1 CUE-1

From a total of 2000 refined structures, 60 unrelated structures with low potential energy were analysed. These structures were calculated independently from a randomly disordered polypeptide chain corresponding to the CUE-1 sequence using experimentally derived restraints. Figure 5.19 shows an ensemble superposition of the calculated low energy structures. Based on the same restraints, low energy structures fell into one of two structurally similar but distinct families. Both families are a three-helix bundle, with α-helices 1 and 3 approximately parallel and the second α-helix orientated diagonally connecting the two. This conformation is similar to previously reported CUE proteins (section 1.4.3).

5.7.1.1 Two Families of Structures

The two energy-minimised families of CUE-1 can be classified by structural features. α-helices 1 and 3 are either parallel or kinked relative to one another along the long axis of the helix. This is a result of differing ϕ / ψ dihedral angles occurring within the loops between helices which translates into particular helix orientations. Two residues, Q42 and T55, have significant differences between families (figure 5.19).

Figure 5.20 shows separate ensembles for each of the two families of structures. For each family, backbone residues show little deviation from the mean. In addition to

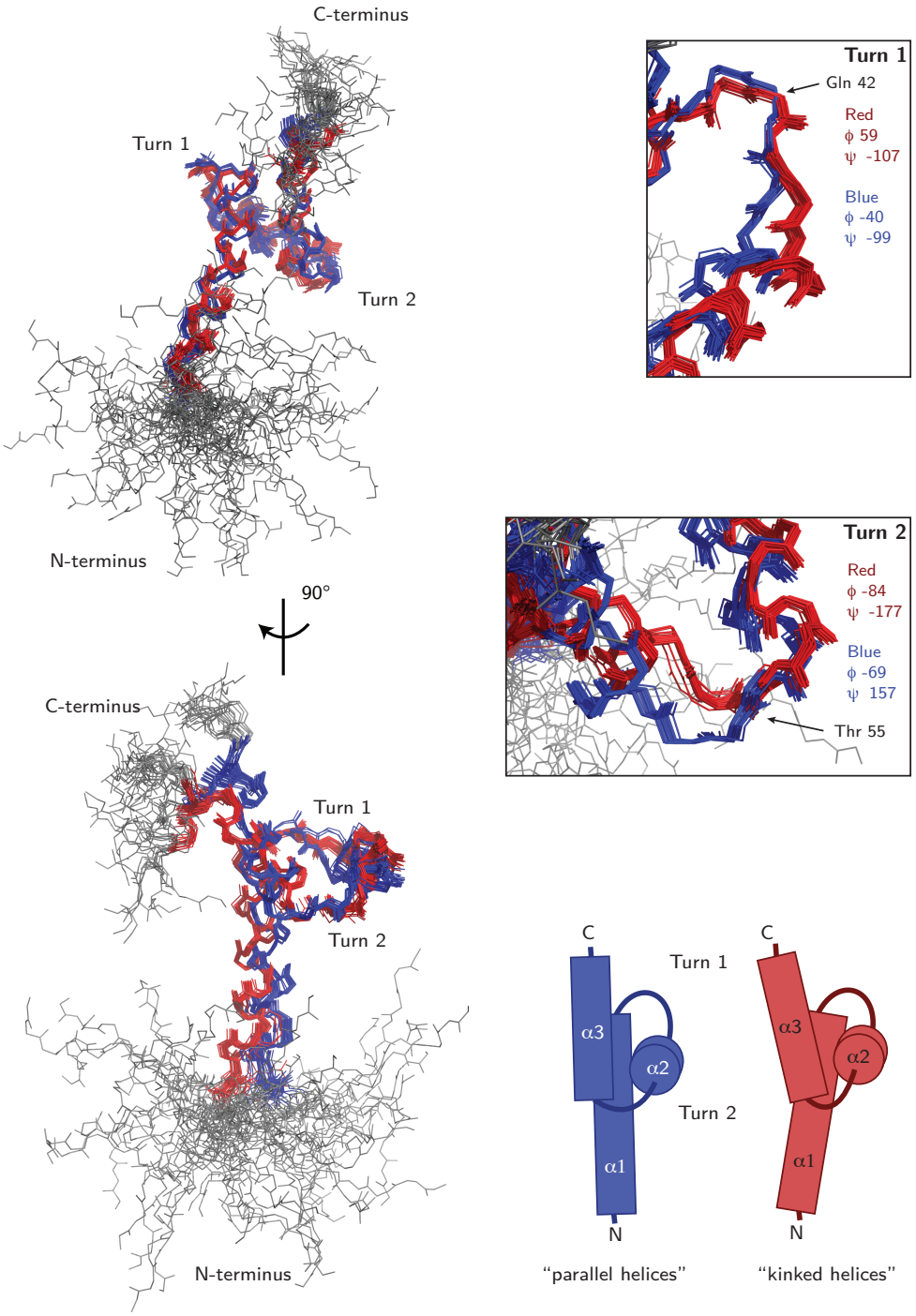


Figure 5.19: Ensembles of Refined Energy Minimised Structures of CUE1 - Backbone representation (Cα, C', O and N) of superpositions of 60 low-energy CUE-1 structures. Two main structural families were identified from the ensemble of low-energy structures (red and blue; non-helical residues at C- and N-termini are shaded grey). Inset boxes show the different loop structures between the two families of structures. In the blue family, helices 1 and 3 are parallel to one another whereas structures shaded red are kinked relative to one another (schematic).

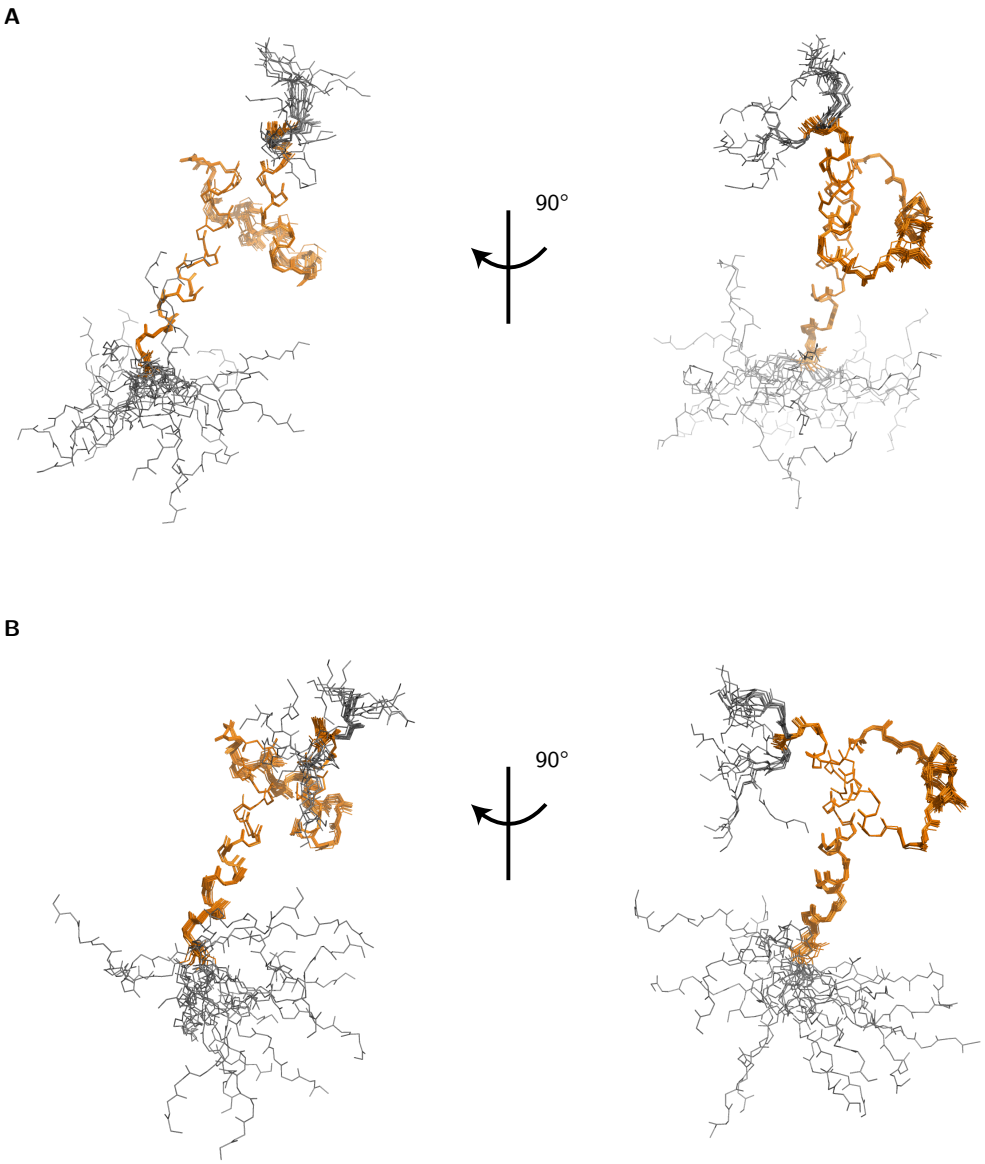


Figure 5.20: CUE-1 Low-energy Structural Families - Backbone atom representation of the two CUE-1 families of structures."Parallel" (**A**) and "kinked" (**B**) helix conformations of CUE-1 were calculated and validated (table 5.2). 20 structures are shown as an ensemble with ordered residues (21-70) in orange and more flexible, non-helical, residues in grey.

Distance and dihedral constraints	**CUE-1"parallel"**	**CUE-1 "kinked"**
Distance constraints	385	385
Total NOE	335	335
Intra-residue	194	194
Inter-residue	141	141
Sequential [| i - j | = 1]	60	60
Medium-range [1 < | i -j | ≤ 4]	57	57
long range [| i -j | > 4]	24	24
Hydrogen bonds	50	50
Total dihedral angle restraints	98	98
ϕ	49	49
ψ	49	49
Structure Statistcs		
Potential energies (Xplor)[1] (mean ± s.d.)		
Total (kcal mol^{-1})	111.32 ± 3.54	112.59 ± 2.52
Bonds (kcal mol^{-1})	2.06 ± 0.15	2.57 ± 0.14
Angles (kcal mol^{-1})	40.42 ± 0.48	38.44 ± 1.42
Improper (kcal mol^{-1})	9.68 ± 0.26	8.52 ± 0.31
van der Waal (kcal mol^{-1})	37.38 ± 1.29	36.89 ± 2.00
NOE (kcal mol^{-1})	14.34 ± 1.85	19.60 ± 0.96
Dihedral Angles (kcal mol^{-1})	1.65 ± 0.43	1.49 ± 0.35
RDC (Sani) (kcal mol^{-1})	5.79 ± 0.32	5.07 ± 0.37
NOE Violations (> 0.2)	1	3
Deviations from ideal covalent geometry		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.5	0.5
Close Contacts[2]	50	77
Ramachandran plot statistics[3]		
Residues in most favored regions	94.4 %	93.8%
Residues in allowed regions	5.6 %	6.2%
Residues in disallowed regions	0%	0%
Root-mean-square deviation[4] (RMSD) (Å)		
All heavy atoms	0.8	0.8
Backbone atoms (N^H, C^α, C')	0.3	0.2

Table 5.2: NMR Structure refinement statistics for both families of CUE-1 structures. [1]Based on 25 lowest total energy structures. [2]Number of close contacts within 1.6 Å for H atoms and 2.2 Å for heavy atoms. [3]Ramachandran statistics generated using the protein structure validation software suite (PSVS) / Molprobity (208, 210). [4]RMSD calculated for ordered residues only aa22-aa69.

structural characteristics that distinguish the two families, the various energy potential terms involved with structure calculations have different values associated with them (table 5.2). On average, the two families have a comparable *total* potential energy but the "parallel" family of structures all share a significantly lower NOE energy (14.3 kcal mol^{-1} compared to the "kinked" conformations' average of 19.6 kcal mol^{-1}). The kinked family of structures gave 3 NOE restraint violations > 0.2, whereas the structures which adopt the parallel conformation only require one significant NOE restraint violation. In the case of structures belonging to the "parallel" family, most other potential energy terms have, on average, a slightly higher value than for the "kinked" family, which explains why the total potential energy for the two structure conformations are approximately equal.

Structural validation analysis of ensembles for each family of structures found that both have only small deviations from idealised covalent geometry, low root-mean-square deviations for ordered residues and all residues in allowed regions of the Ramachandran plot with the vast majority (93-95 %) in most favoured regions. The "parallel" family of structures has notably fewer close contacts than for "kinked" family members.

5.7.1.2 Energy-minimised Average Structure

A single, energy-minimised average structure was generated for both families based on the lowest-energy 20-structure ensembles (figure 5.21). The ordered residues are shown

by low root-mean square deviation scores (figure 5.22).

5.7.1.3 Comparison of CUE-1 Families with other CUE Domains

It is difficult to categorically conclude which of two CUE-1 conformations represents an accurate structure. Both satisfy allowed regions of the Ramachandran plot favourably, have low RMSD values, low deviations from ideal geometry and comparable potential energies. The "parallel" family of structures has a smaller number of large NOE violations and close contacts. Comparison of both CUE-1 structures with existing CUE structures (figures 5.23, 5.24 and 5.25) identifies the "parallel" family of structures as having greatest structural homology to other CUE structures. The RMSD across all atoms for the "parallel" CUE-1 / CUEDC1 and "parallel" CUE-1 / Cue2-1 superpositions are 2.9 Å and 3.9 Å respectively. A similar comparison with the kinked CUE-1 gives RMSD values of 4.4 Å and 4.6 Å. The greater CUE structural homology of the "parallel" CUE-1 conformation suggests that this structure is more accurate. The CUE-1 energy-minimised average structure of the CUE-1 "parallel" ensemble was used as the CUE-1 structure for subsequent analysis.

5.7.1.4 N-terminal End of α-helix 1 of CUE-1

Analysis of heteronuclear NOE data (figure 5.13), residual dipolar couplings (figure 5.15) and Cα secondary chemical shift data (figure 5.26) suggests that the protruding

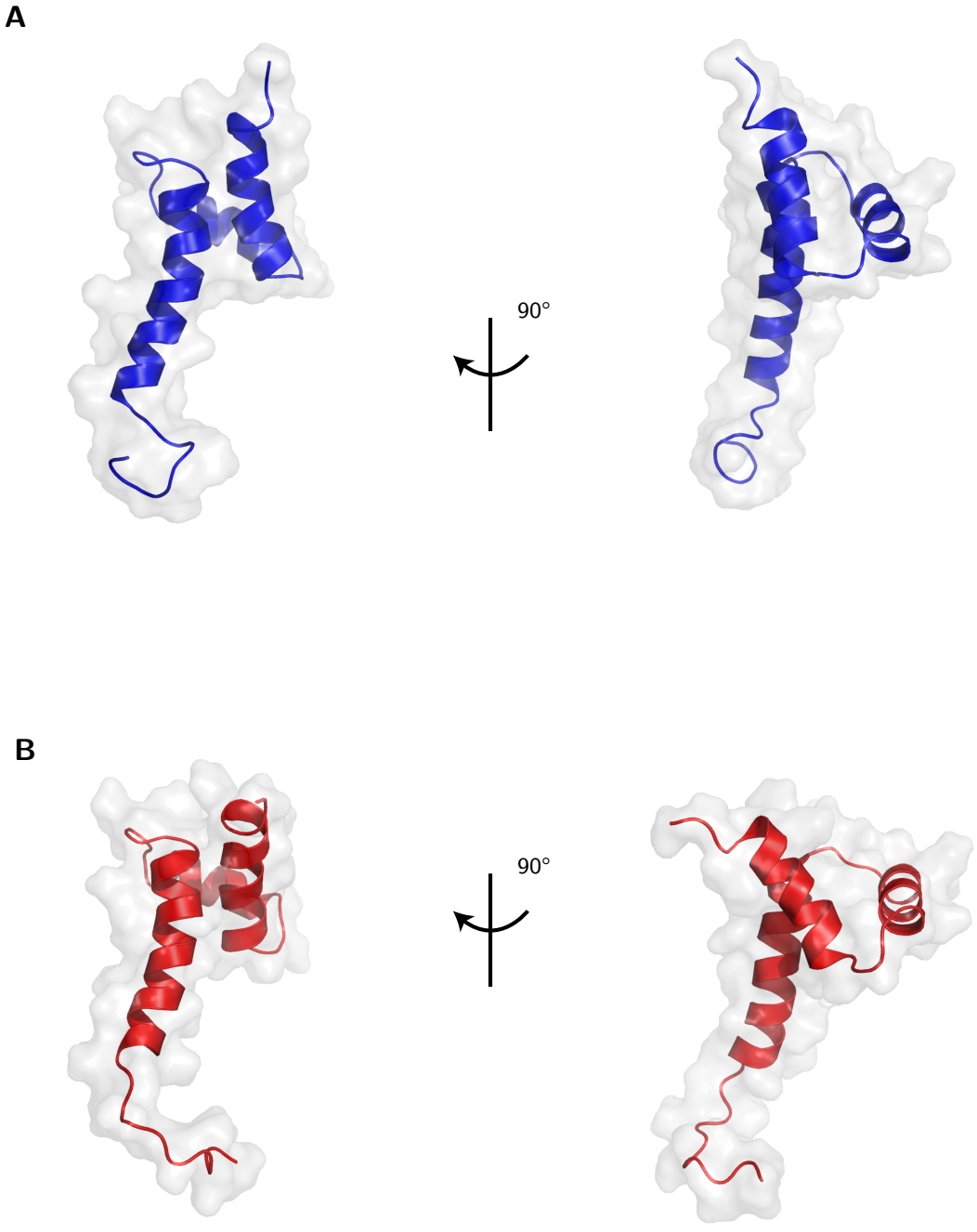


Figure 5.21: Energy Minimised Average Structures for Both CUE-1 Conformations - **A** Parallel CUE-1 **B** Kinked CUE-1

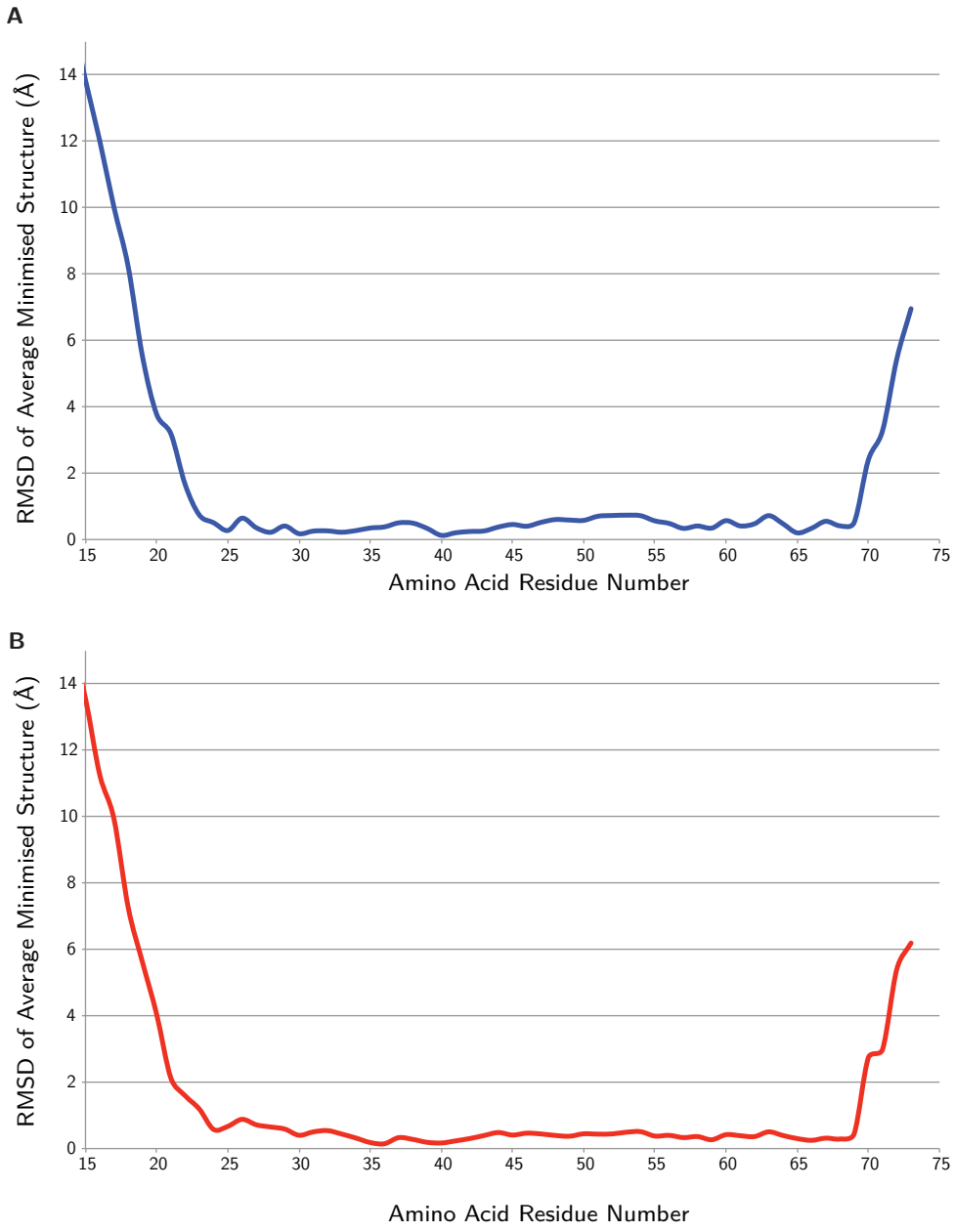


Figure 5.22: Root-mean-square deviation of Cα atoms in CUE-1 Ensembles - Average deviation of lowest-energy ensemble (20 structures) from energy-minimised average structure **A** Parallel CUE-1 **B** Kinked CUE-1.

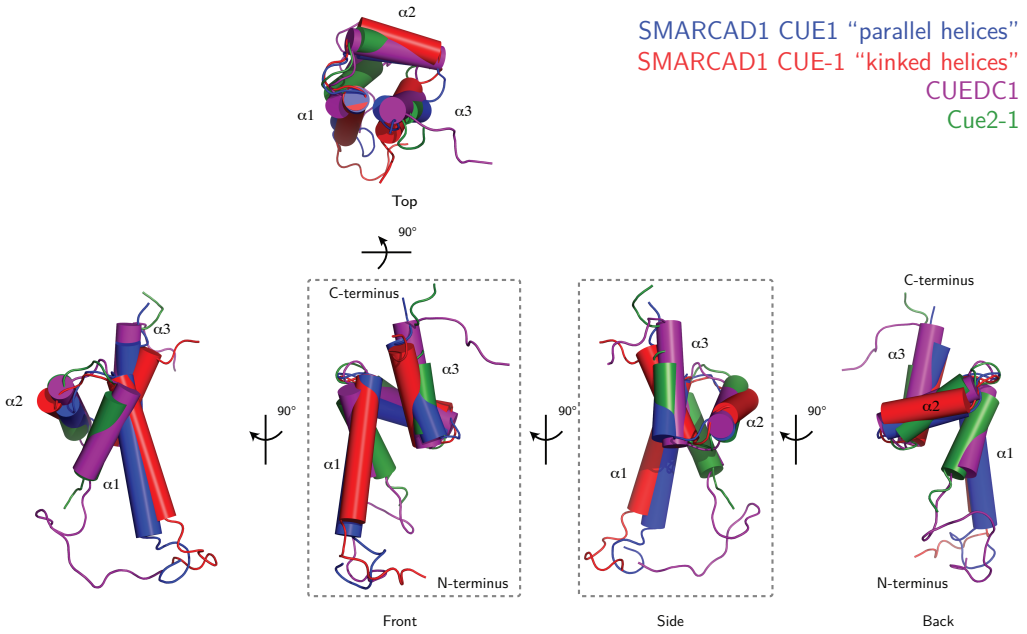


Figure 5.23: Overlay of CUE-1 Structures with Other CUE Domains - The energy minimised structures for both CUE-1 families (red and blue) were superimposed with the deposited structures for the CUE domain from human CUEDC1 (CUE Domain Containing Protein 1; purple) (267) and the N-terminal CUE domain from yeast Cue2 protein (green) (167). α-helices drawn as cylinders. Both CUEDC1 and Cue2-1 CUEs are similar in overall topology to SMARCAD1 CUE-1 although α-helix 1 in CUE-1 is extended N-terminally compared to both CUEDC1 and Cue2-1 and is at a different angle relative to the core of the helix bundle. In the case of both α-helices 2 and 3, the structure from the "parallel helices" CUE-1 family (blue) superimposes more closely with the deposited structures. Figures 5.24 and 5.25 show cartoon representations of the 'front' and 'side' orientations (boxed).

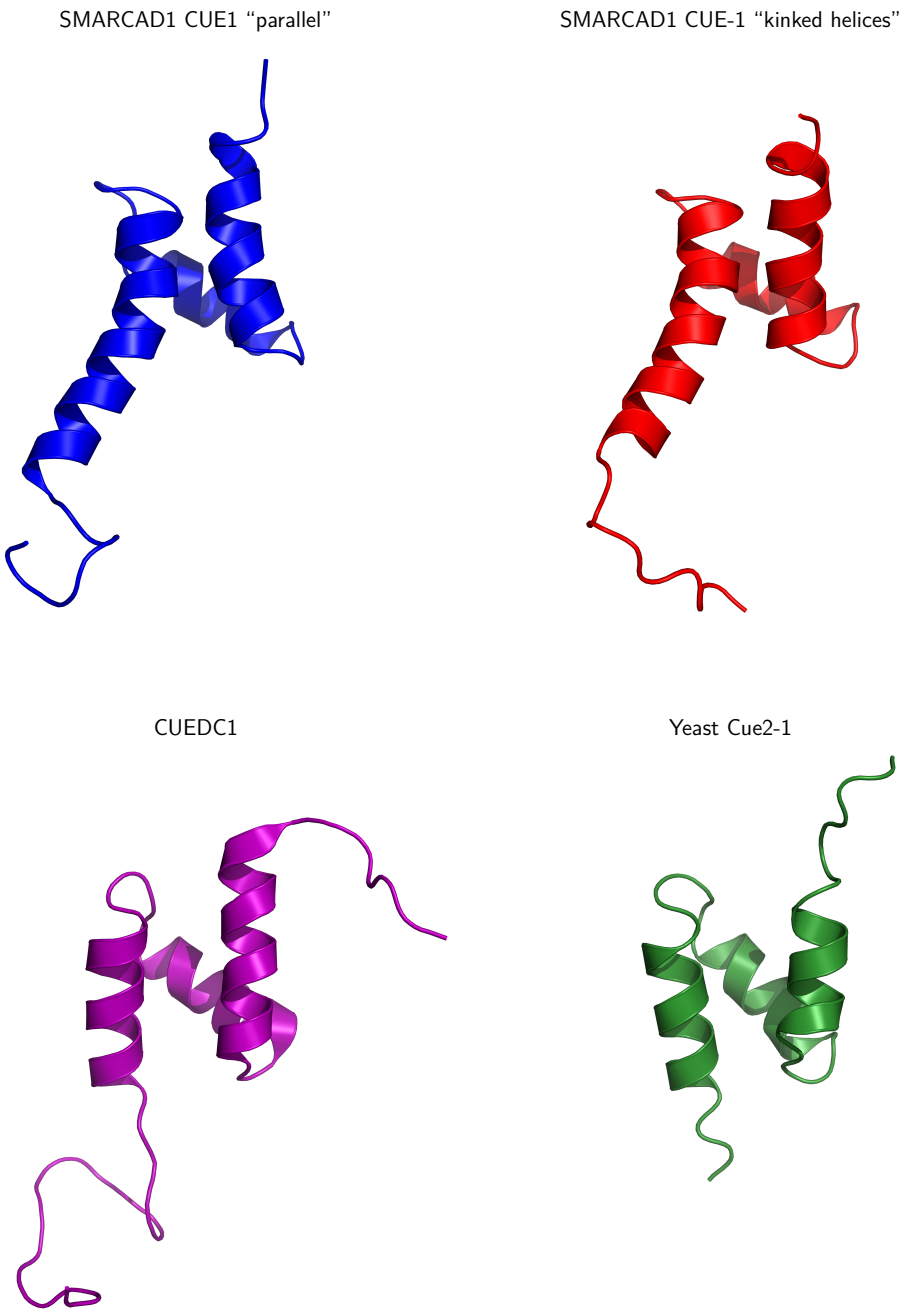


Figure 5.24: Comparison of CUE-1 Structures with Other CUE Domains (‘Front’ View) - Magenta CUEDC1 (PDB: 2DHY). Green Cue2-1 (PDB: 1OTR)

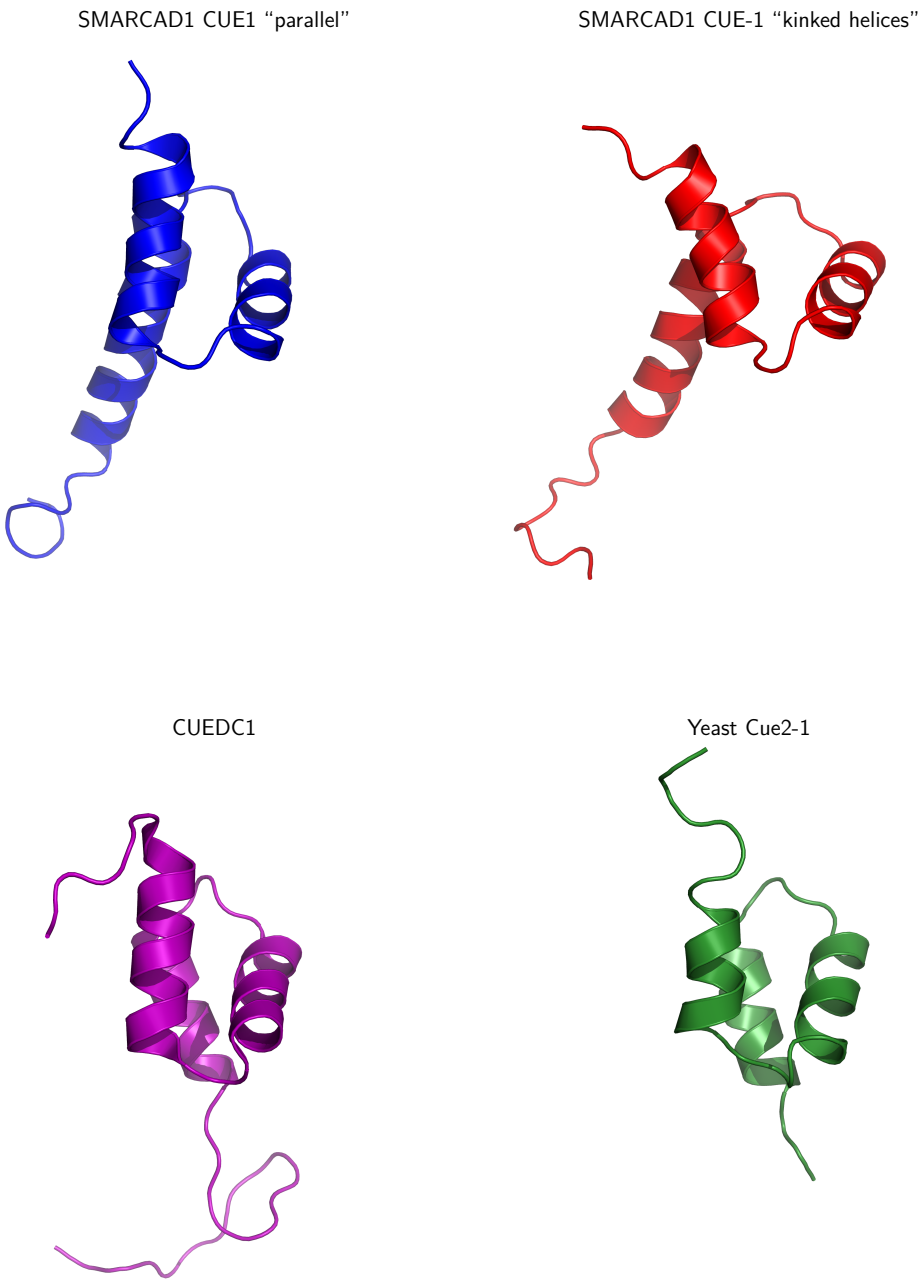


Figure 5.25: Comparison of CUE-1 Structures with Other CUE Domains (‘Side’ View) - Magenta CUEDC1 (PDB: 2DHY). Green Cue2-1 (PDB: 1OTR)

N-terminal end of α-helix 1 in CUE-1 may be somewhat less fully helical than the main core of the helix bundle. In the case of the heteronuclear NOE experiment, there is a gradual change in heteronuclear NOE ratio from highly dynamic to rigid across a number of residues (aa21-29). This is not seen at the other boundaries between helices and linker / disordered residue and implies that some of the residues at the N-terminal end of the first helix have a greater degree of flexibility but are not fully dynamic. The RDC values for the same residues (aa21-28; RDC: *ca* +5) also cluster at a different value than the rest of the helix (aa30-40; RDC: *ca* +15)

These observations suggest a degree of dynamic behaviour in the N-terminal end of the first helix in CUE-1. It may be that these observations are the result of averaging between two different states, where residues aa21-28 either form an extended helix or random coil (figure 5.27). Given that this particular part of helix 1 is not part of the main helical bundle or even canonical CUE motif, it may be that it has an additional function, for example having a long helix could expand a binding surface. This helix is heavily amphipathic, with one side rich in hydrophobic leucine residues, which this N-terminal section continues. One interesting possibility is that, upon certain interactions, this helix becomes more rigid (induced structure) and serves as an extended binding site. Conceivably, as a ubiquitin binding domain, the longer helix may be involved with binding particular ubiquitin species, such as polyubiquitin / diubiquitin chains or

specific ubiquitinated proteins.

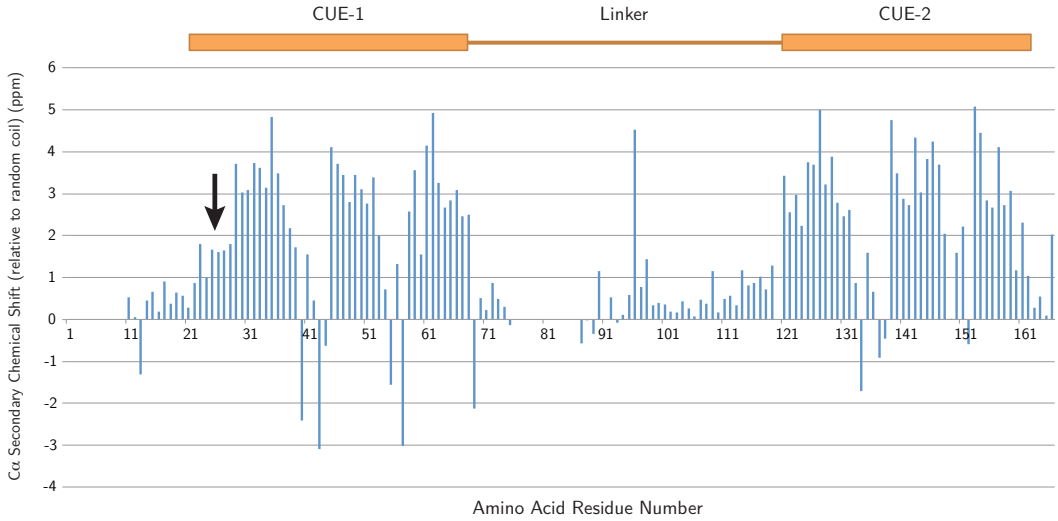


Figure 5.26: Chemical Shift Values for Cα Relative to Random Coil Values - Helical residues typically have a downfield shifted Cα chemical shift relative to the random coil index (268). It is striking to note that the majority of regions identified as helical within the double CUE domain have significant (+2-5 ppm) secondary chemical shifts except the N-terminal 8 residues (aa21-28; arrow) of the first helix in CUE-1. This suggests that the helical nature of the very N-terminal section of the first helix of CUE-1 may be slightly different than the other part. One possibility for the step-change in secondary chemical shift may be an effect of averaging if that section of the helix is only helical part of the time. Given that this part of the helix protrudes from the main CUE bundle, it is tempting to speculate as to whether this part of the helix experiences induced structural changes.

5.7.2 CUE-2

The structure of the second SMARCAD1 CUE domain, CUE-2, was calculated as for CUE-1. An ensemble of 20 lowest-energy structures was used to calculate the energy minimised average structure (figures 5.28 and 5.29). Unlike for CUE-1, all calculated low-energy structures were structurally very similar with a low backbone RMSD of 0.3 Å. No residues are in disallowed regions of the Ramachandran plot and 94.5 % are in the most favoured regions. Table 5.4 details full refinement statistics. CUE-2 is a three

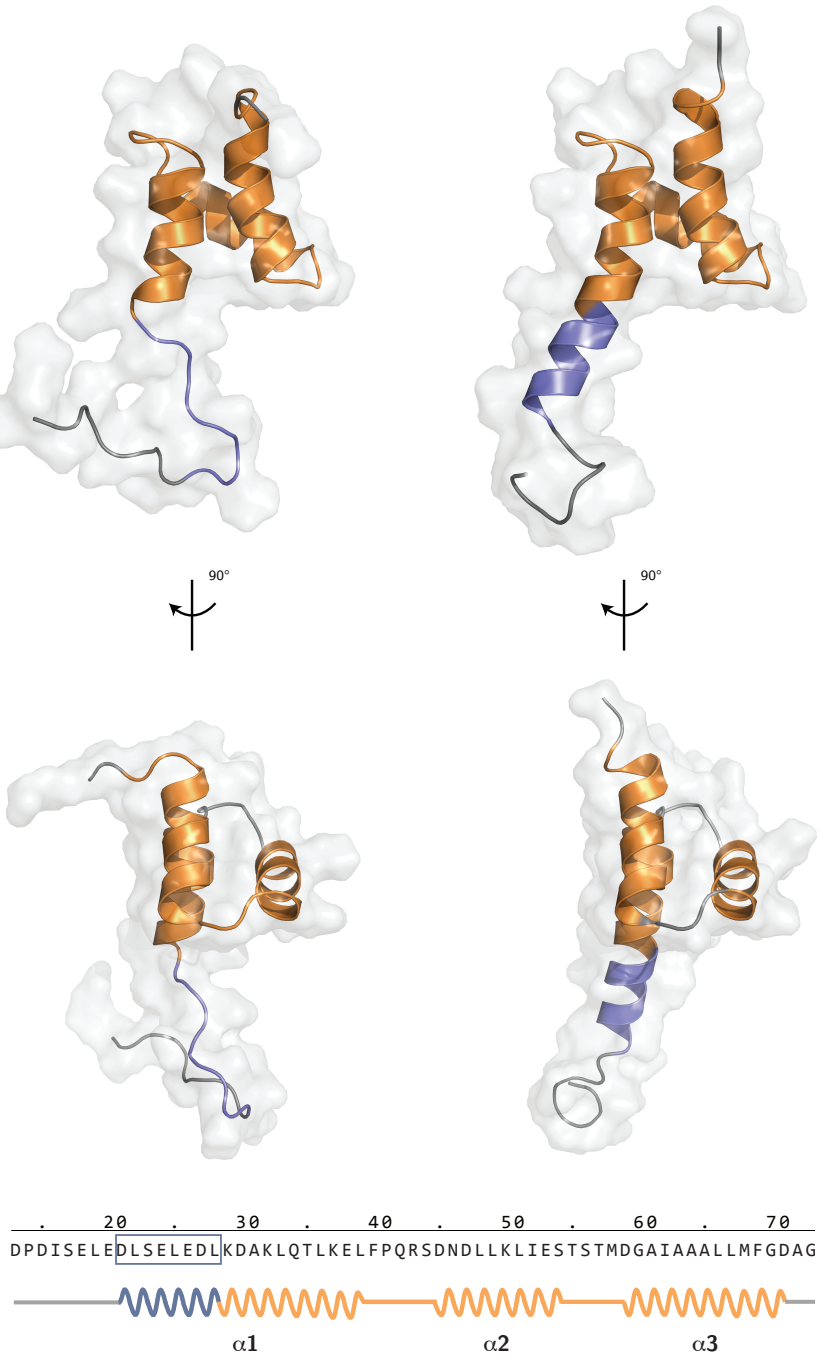


Figure 5.27: The N-terminal end of CUE-1 helix 1 - Structural calculations were carried out with all restraints (RDC, NOE, H-bond) removed for residues prior to D30 (left; disordered residues aa 21-28 shaded purple). Solved structure of CUE-1 (right; section 5.7.1) with more highly dynamic residues aa21-28 shaded in blue. It is possible that CUE-1 rapidly interconverts between these two states.

helix bundle with remarkable structural homology to two other CUE domains. RMSD across all atoms between SMARCAD1 CUE-2 domain and CUEDC1 and Cue2-1 CUE domains is 1.8 Å and 2.3 Å respectively. Figure 5.30 shows a structural comparison between SMARCAD1 CUE domains (CUE-1 and CUE-2) and CUEDC1 and Cue2-1 CUE domains.

5.7.3 Complete Double CUE-Domain

In order to visualise the full double CUE domain from SMARCAD1, structure calculations were run with a complete set of restrains for both domains. Given there were very few restraints in the linker region, the calculated structures have a high degree of variability for the linker. This reflects the genuine flexibility and dynamic nature of the protein. Figure 5.31 shows a selection of low-energy dCUE structures, which display a variety of CUE orientations with respect to one another and no fixed linker position. The distance between the CUE domains also varies. RMSD values for the structures are very high, 12.7 Å for backbone residues (table 5.4), which, as expected, indicates that the calculated structures are not very similar or superimposable. Figures 5.32 and 5.33 show dCUE superpositions, aligned by CUE-1 and CUE-2 respectively. The non-superposed CUE domain samples a range of positions across the various structures as the two CUE domains are independent of one another. Unsurprisingly, the distance

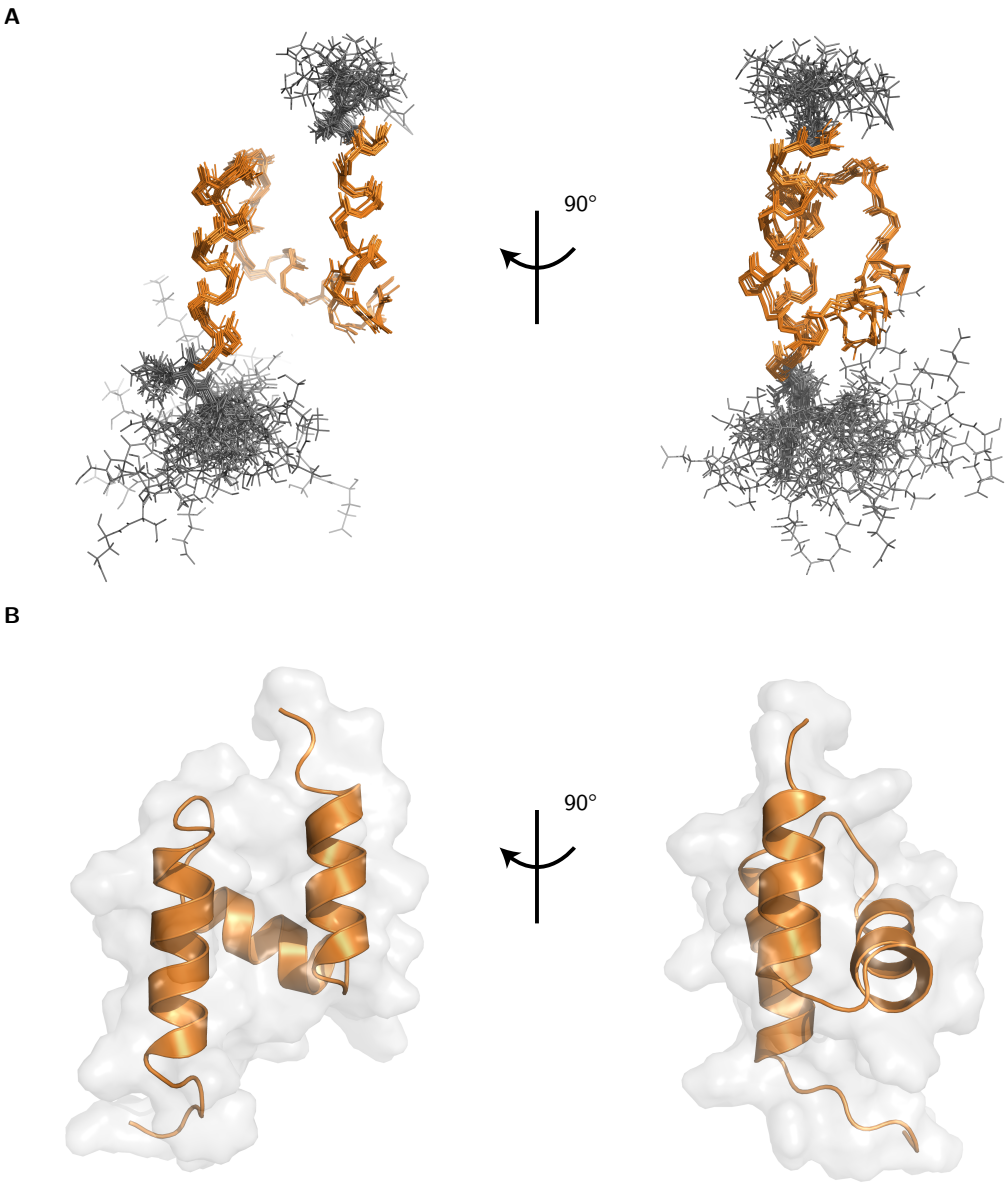


Figure 5.28: CUE-2 Ensemble of Structures and Energy Minimised Average Structure - **A** Lowest-energy 20 structures **B** Energy minimised average structure. Ordered residues (120-163) in orange.

Distance and dihedral constraints	**CUE-2**
Distance constraints	429
Total NOE	387
Intra-residue	268
Inter-residue	119
Sequential [| i - j | = 1]	59
Medium-range [1 < | i -j | ≤ 4]	23
long range [| i -j | > 4]	38
Hydrogen bonds	42
Total dihedral angle restraints	81
ϕ	40
ψ	41
Structure Statistcs	
Potential energies (Xplor)[1] (mean ± s.d.)	
Total (kcal mol^{-1})	52.45 ± 0.79
Bonds (kcal mol^{-1})	0.54 ± 0.03
Angles (kcal mol^{-1})	21.28 ± 0.48
Improper (kcal mol^{-1})	2.44 ± 0.13
van der Waal (kcal mol^{-1})	14.08 ± 0.49
NOE (kcal mol^{-1})	1.33 ± 0.34
Dihedral Angles (kcal mol^{-1})	1.08 ± 0.46
RDC (Sani) (kcal mol^{-1})	11.7 ± 0.86
NOE Violations (> 0.2)	0
Deviations from ideal covalent geometry	
Bond lengths (Å)	0.001
Bond angles (°)	0.3
Close Contacts[2]	10
Ramachandran plot statistics[3]	
Residues in most favored regions	94.5 %
Residues in allowed regions	5.5 %
Residues in disallowed regions	0%
Root-mean-square deviation[4] (RMSD) (Å)	
All heavy atoms	1.0
Backbone atoms (N^H, C^α, C')	0.3

Table 5.3: NMR Structure refinement statistics for SMARCAD1 CUE-2 domain. [1]Based on 20 lowest total energy structures. [2]Number of close contacts within 1.6 Å for H atoms and 2.2 Å for heavy atoms. [3]Ramachandran statistics generated using the protein structure validation software suite (PSVS) / Molprobity (208, 210). [4]RMSD calculated for ordered residues only aa121-aa134, aa137-aa163 .

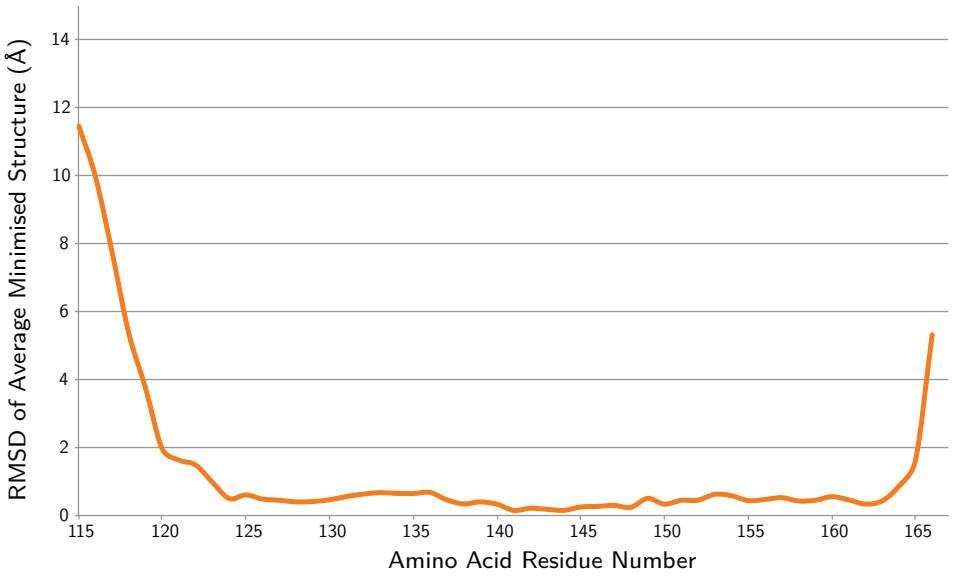


Figure 5.29: Room-mean-square deviation of Cα atoms in CUE-2 Ensemble - Average deviation of lowest-energy ensemble (20 structures) from energy-minimised average structure

between the two CUE domains also various somewhat with a range of *ca* 35 - 128 Å in 20 calculated structures (table 5.5).

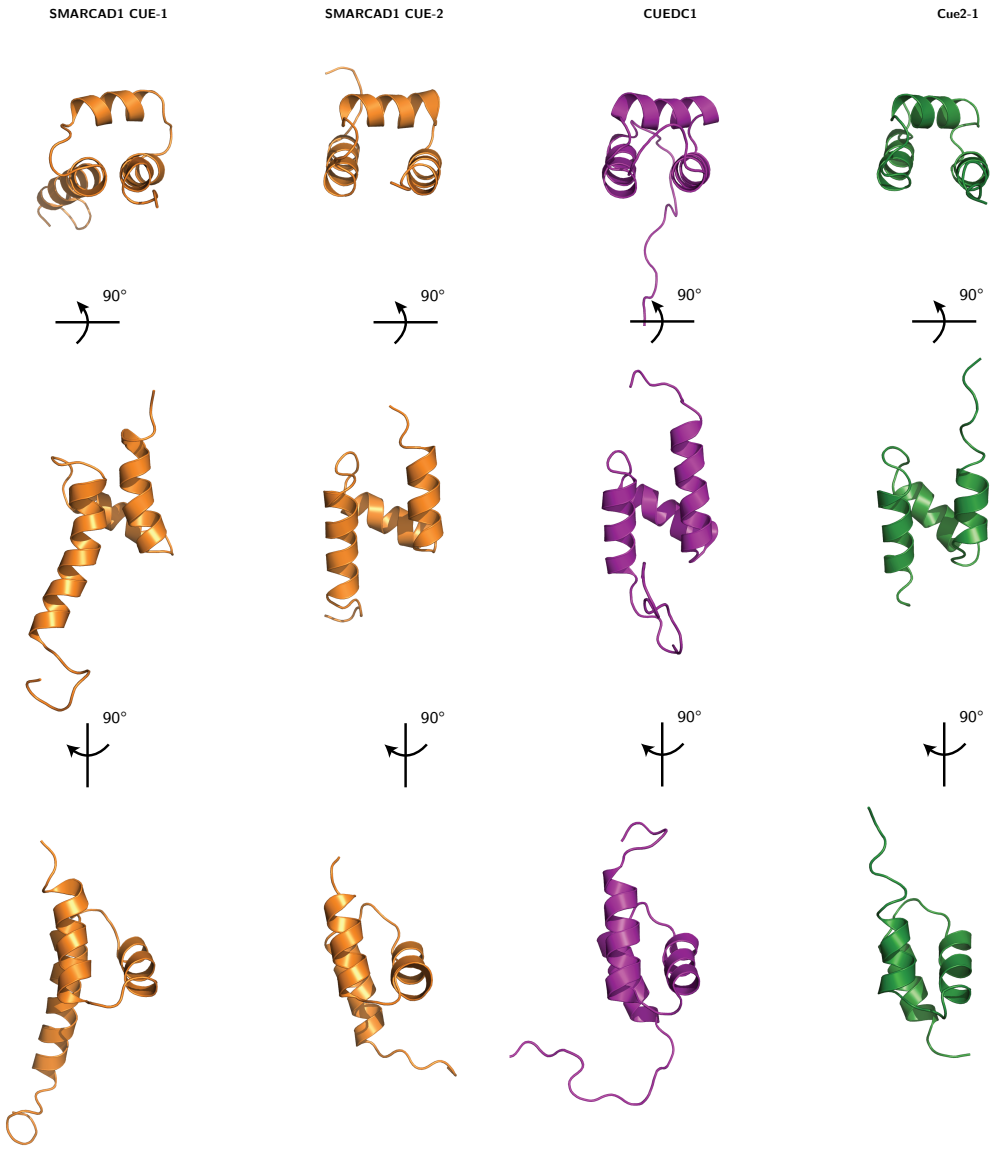


Figure 5.30: Comparison of SMARCAD1 CUE Domains with Other CUE Domains - Solution structures for SMARCAD1 CUE-1 and CUE-2 show significant conformational similarity with other CUE domains.

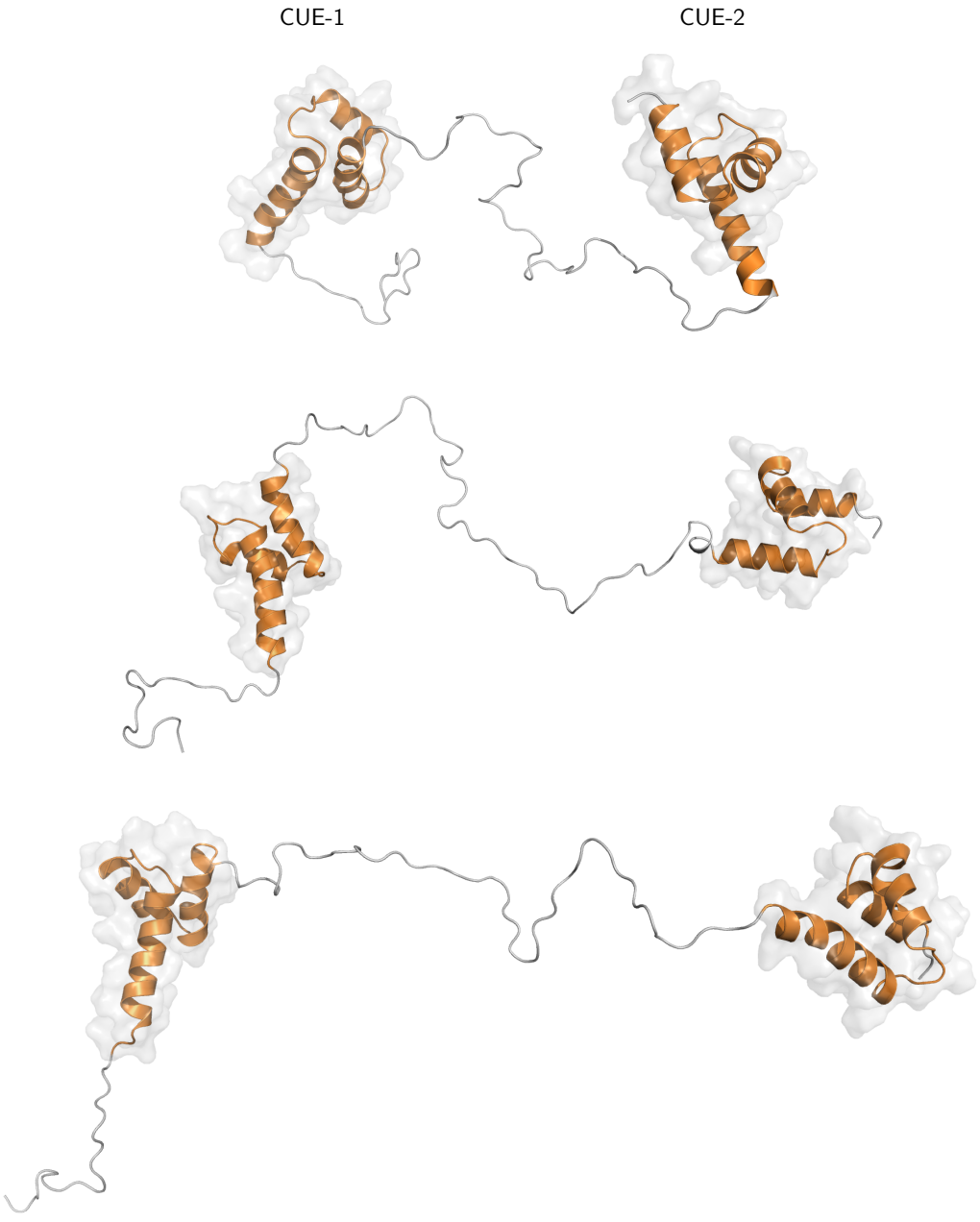


Figure 5.31: Full-length dCUE - Examples of three low energy structures calculated using a full set of restraints and complete. Variety of length of linker and orientation.

Distance and dihedral constraints	**Full dCUE**
Distance constraints	901
Total NOE	722
Intra-residue	462
Inter-residue	237
Sequential [| i - j | = 1]	118
Medium-range [1 < | i -j | ≤ 4]	80
long range [| i -j | > 4]	62
Inter-domain connections (CUE-1-CUE-2)	0
Hydrogen bonds	92
Total dihedral angle restraints	179
ϕ	89
ψ	90
Structure Statistcs	
Potential energies (Xplor)[1] (mean ± s.d.)	
Total (kcal mol^{-1})	185.64 ± 4.34
Bonds (kcal mol^{-1})	3.10 ± 0.30
Angles (kcal mol^{-1})	72.77 ± 2.35
Improper (kcal mol^{-1})	11.55 ± 0.47
van der Waal (kcal mol^{-1})	56.50 ± 2.19
NOE (kcal mol^{-1})	18.31 ± 2.52
Dihedral Angles (kcal mol^{-1})	3.67 ± 0.89
RDC (Sani) (kcal mol^{-1})	19.73 ± 1.0
Deviations from ideal covalent geometry	
Bond lengths (Å)	0.002
Bond angles (°)	0.3
Close Contacts[2]	78
Ramachandran plot statistics[3]	
Residues in most favored regions	94.3 %
Residues in allowed regions	5.7 %
Residues in disallowed regions	0%
Root-mean-square deviation[4] (RMSD) (Å)	
All heavy atoms	12.1
Backbone atoms (N^H, C^α, C')	12.7

Table 5.4: NMR Structure refinement statistics for SMARCAD1 dCUE domain (CUE-1 and CUE-2). [1]Based on 20 lowest total energy structures. [2]Number of close contacts within 1.6 Å for H atoms and 2.2 Å for heavy atoms. [3]Ramachandran statistics generated using the protein structure validation software suite (PSVS) / Molprobity (208, 210). [4]RMSD calculated for ordered residues only (aa22-aa69, aa116-aa119, aa121-aa134 and aa137-aa163).

Structure	Distance (Å)
1	76.6
2	35.7
3	122.8
4	50.9
5	59.4
6	95.8
7	66.1
8	68.5
9	70.3
10	90.7
11	72.9
12	128.1
13	64.0
14	67.5
15	48.0
16	80.1
17	65.1
18	69.4
19	64.4
20	67.5
Mean ± s.d.	73.2 ± 22.3

Table 5.5: Distance between CUE domains in complete dCUE calculations. Measured from the position of 48 Cγ and 146 Cγ in all structures in the ensemble.

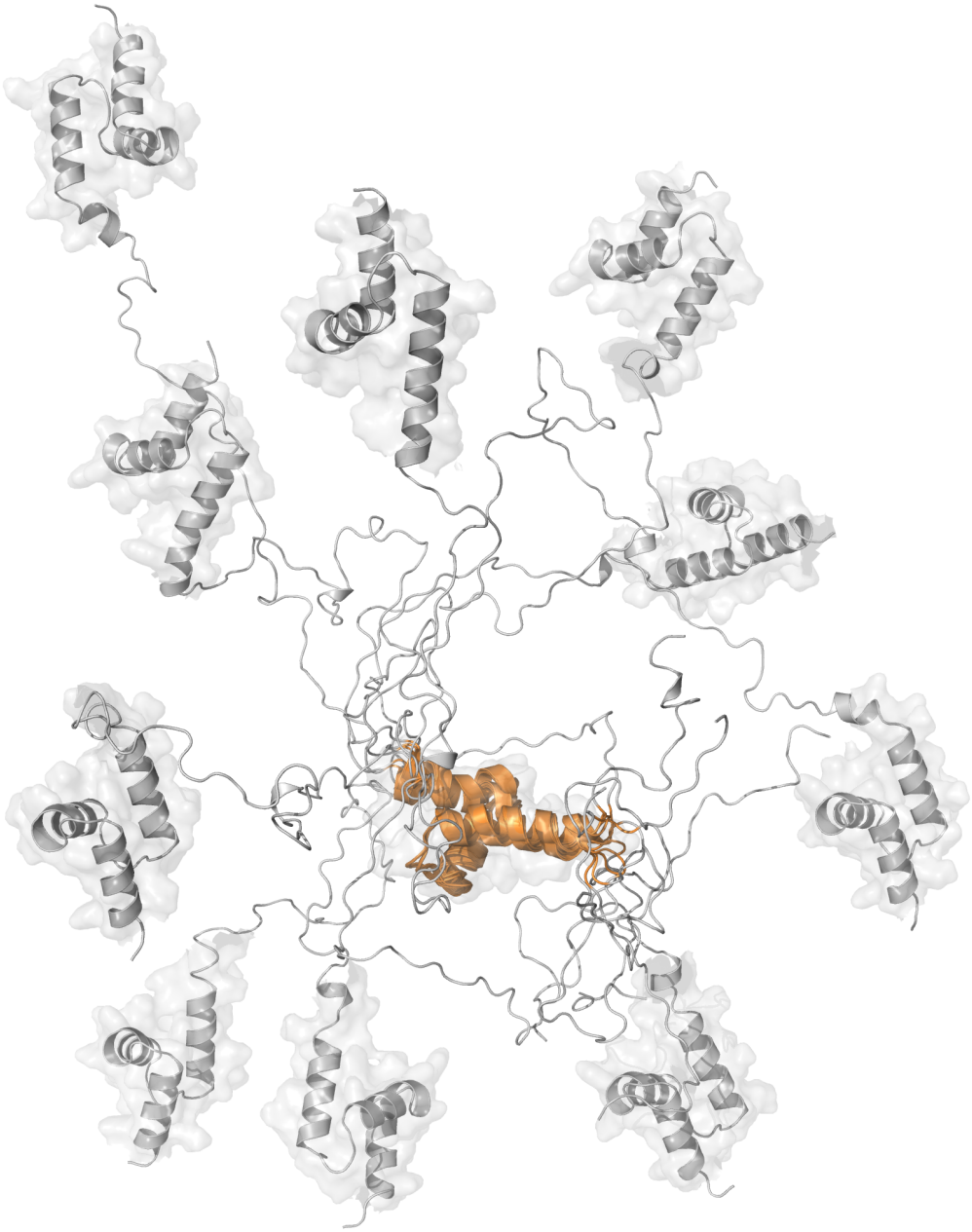


Figure 5.32: Superposition of 10 low-energy dCUE Structure - Structures superimposed on CUE-1 position (orange). Linker is flexible and the CUE-2 (grey) domain is not in a fixed position relative to CUE-2. It samples a wide range of conformational space.

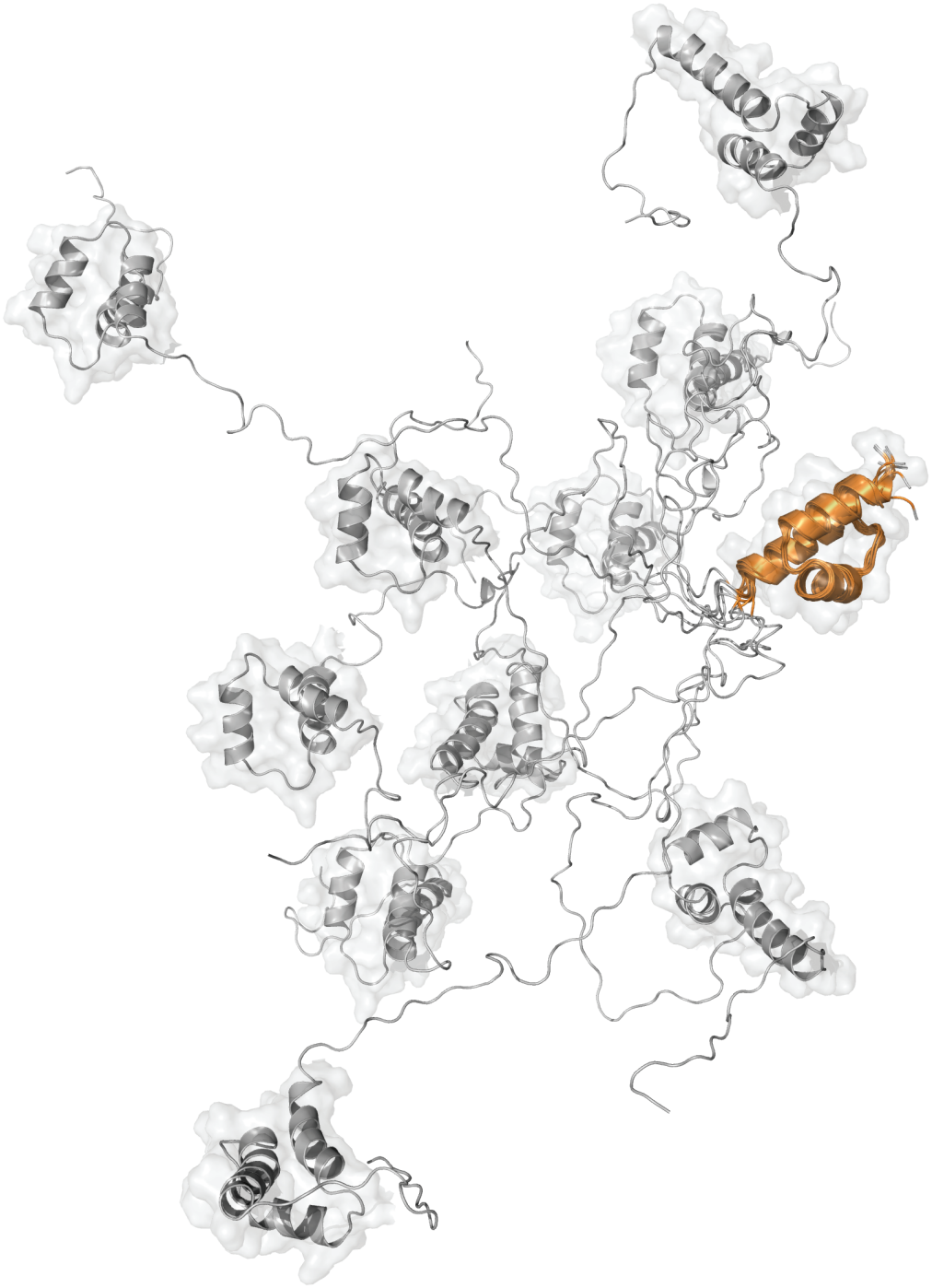


Figure 5.33: Superposition of 10 low-energy dCUE Structure - Structures superimposed on CUE-2 position (orange).

Chapter 6

Functional Studies of SMARCAD1 Double CUE Domain

6.1 Introduction

As previously discussed, CUE domains are generally classified as ubiquitin binding domains (section 1.4.3). In the context of SMARCAD1, a chromatin remodeller without any other apparent accessory domains (section 1.3), it was hypothesised that the double CUE domain was important for conferring functional specificity of the remodeller by acting as a substrate binder.

SMARCAD1 has been shown to bind both proliferating cell nuclear antigen protein (PCNA) and Küppel-associated protein 1 (KAP1) (section 1.3.2). A series of binding experiments was carried out to investigate the binding activity of the double CUE domain and elucidate its function.

6.2 Ubiquitin Binding

As CUE domains are classified as ubiquitin binding domains, a key objective was to establish the binding ability of both CUE domains in SMARCAD1.

6.2.1 NMR Titrations

The use of NMR titration experiments to investigate protein interactions has two significant advantages: it is highly sensitive and can provide very specific information. This means it can be used to detect very weak binding and identify the particular residues involved. The basic principle utilises the observed phenomenon of chemical shift changes with ligand interactions. Unlabelled ligand is titrated into an isotopically labelled protein solution. The labelled protein is detected by NMR spectroscopy and changes in peak chemical shifts / intensity are a result of changes in chemical environment due to ligand interactions (269).

6.2.1.1 Binding Sites

Two titration experiments were carried out to investigate the binding of ubiquitin to dCUE. Figure 6.1 shows an overlay of dCUE ^{1}H, ^{15}N-HSQC spectra after different concentrations of unlabelled mono-ubiquitin had been added. A number of peaks shifted considerably, signifying an interaction. Based on the backbone assignments of dCUE (section 4.3), these predominantly correspond to residues in the first CUE domain and

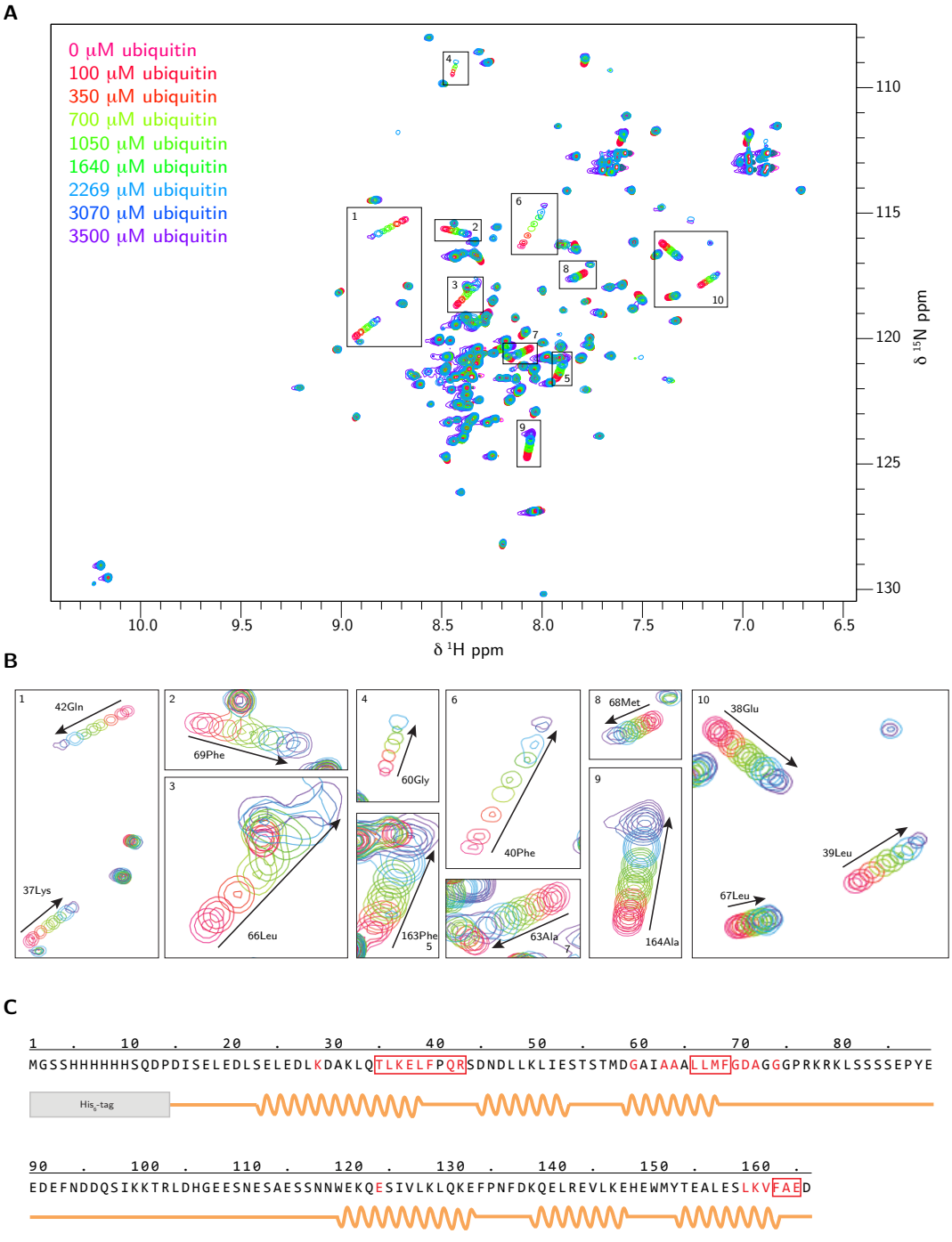


Figure 6.1: NMR Titration Experiment of ^{15}N-dCUE Shifts with Ubiquitin - A Overlay of nine ^{1}H, ^{15}N-HSQC specta for dCUE, each with different ubiquitin concentrations. **B** Selection of magnified ^{1}H, ^{15}N-HSQC peaks that shifted considerably. The sign and magnitude of shift is consistent with increasing ligand concentration. **C** Sequence and secondary structure of dCUE construct. Residues with significant shifts are shown in red, boxes signify a cluster of shifted peaks around a residue with a particularly large change in chemical shift. Clusters of shifts are at the end of helices and correspond to conserved CUE motif residues.

specifically the C-terminal ends of the first and third helices. These sites are the highly conserved residues in CUE motifs (section 3.2.1). Interaction at these conserved sites makes sense given that CUE domains are ubiquitin binding domains and the highly conserved CUE residues are involved with ubiquitin binding. Furthermore, structurally these sites are close in space, providing a single interaction surface.

Fewer peaks from the second CUE domain and none from the assigned parts of the linker move significantly upon titration with ubiquitin. F163 and A164 towards the C-terminus of the dCUE construct have large changes in chemical shift (figure 6.1), which also suggests some ubiquitin affinity. These correspond to the residues immediately after the C-terminal end of helix 3 in CUE-2 and are roughly equivalent residues to those at the end of helix 3 in CUE-1 which bind ubiquitin. This result suggests a situation where free mono-ubiquitin interacts to a larger extent with CUE-1 than with CUE-2, despite some evidence of CUE-2 C-terminal binding.

A complimentary experiment, titrating unlabelled dCUE into ^{15}N-ubiquitin was carried out to identify the respective binding sites on ubiquitin. A ^{1}H, ^{15}N-HSQC spectrum was collected and assigned based on published ubiquitin shifts (figure 6.2) (196). As experimental conditions were slightly different to those used in the published assignments, some resonances had slightly different chemical shifts. Ambiguity was addressed by collecting a 3D ^{15}N-edited TOCSY to give relevant side chain shifts,

which allowed the peaks to be correctly assigned.

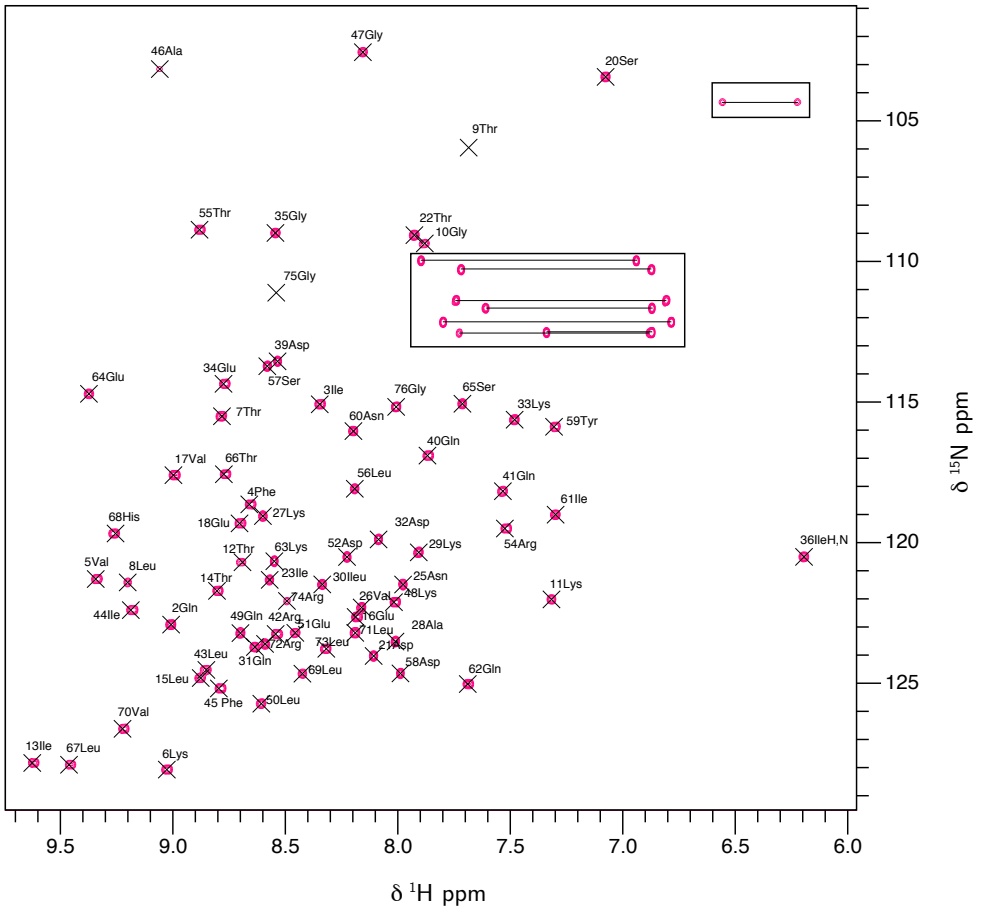


Figure 6.2: Assigned ^{1}H, ^{15}N-HSQC of ^{15}N-labelled Ubiquitin - An HSQC spectrum was assigned based on published assignments and a 3D ^{15}N-edited TOCSY (196). Boxes correspond to asparagine and glutamine side chains.

As in the case of titrating unlabelled ubiquitin into ^{15}N-labelled dCUE, a change in chemical shifts of some peaks was observed when titrating unlabelled dCUE into the ^{15}N-labelled ubiquitin sample (figure 6.3) indicating a role in the dCUE interaction.

In order to quantify which peaks undergo the largest change in position upon ligand

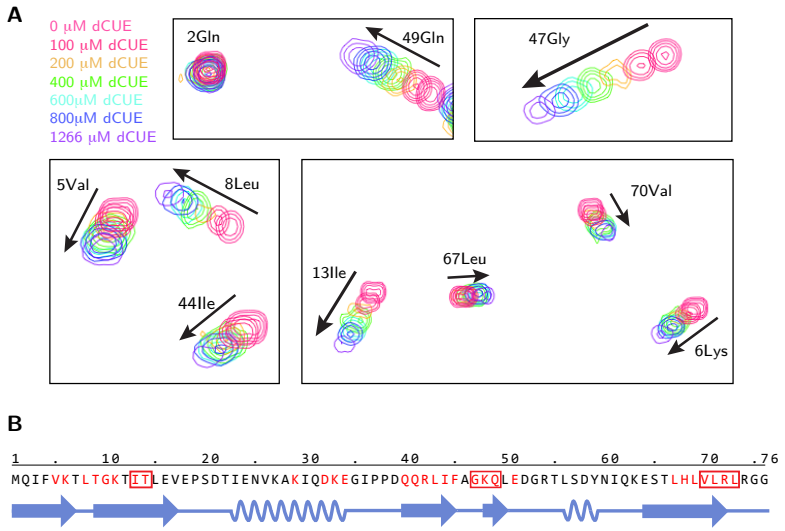


Figure 6.3: NMR Titration Experiment of ^{15}N Ubiquitin with Unlabelled dCUE. - **A** Selected peaks from ^{15}N-labelled ubiquitin ^{1}H, ^{15}N-HSQC that show considerable shifts due to increasing dCUE concentration. **B** Sequence and secondary structure of ubiquitin. Residues with significant shifts are shown in red, boxes signify a cluster of shifted peaks around a residue with a large change in chemical shift.

binding, a weighted change in chemical shift, $\Delta\delta$, was calculated for all residues in both ubiquitin and dCUE based on their initial chemical shift and their final peak chemical shift using the following formula:

$$\Delta\delta = \left[(\Delta\delta_{\mathrm{H}})^2 + (0.2 \cdot \Delta\delta_{\mathrm{N}})^2\right]^{1/2}$$

$\Delta\delta_{\mathrm{H}}$ and $\Delta\delta_{\mathrm{N}}$ are the observed chemical shift differences in ppm between the initial state (no ligand) and the end of the titration for the ^{1}H and ^{15}N dimensions respectively. As peaks shift in both dimensions in the HSQC, $\Delta\delta$ is the sum of the magnitude of change in both dimensions, ^{1}H and ^{15}N, with the $\Delta\delta_{\mathrm{N}}$ value weighted relative to $\Delta\delta_{\mathrm{H}}$ ($\times$ 0.2) as the ^{15}N dimension covers a wider range of chemical shifts (~30 ppm) than

the $^{1}H^{N}$ dimension (~4 ppm). (270).

A large change in chemical shift is indicative of a change in chemical environment and may suggest a binding event. Comparison of the relative differences in observed chemical shift was useful for determining the binding sites and key residues within a given protein. Figure 6.4 shows the calculated $\Delta\delta$ for each assigned residue in both dCUE and ubiquitin. Residues were either classified as *small* ($\Delta\delta < 0.15$ ppm), *medium* ($0.15 \leq \Delta\delta < 0.3$) or *large* ($\Delta\delta \geq 0.3$) difference (table 6.1). The secondary chemical shifts were mapped onto 3D structures of ubiquitin (figure 6.5) and CUE-1 (figure 6.6).

Residue	**Relative $\Delta\delta$**
dCUE	
Lys 37	medium
Glu 38	medium
Leu 39	medium
Phe 40	large
Gln 42	large
Gly 60	medium
Leu 66	medium
Asp 69	medium
Phe 163	medium
Ala 164	medium
Ubiquitin	
Ile 13	medium
Gly 47	medium
Lys 48	large
Gln 49	medium
Leu 71	medium

Table 6.1: Significant Shifts on dCUE:Ub Binding

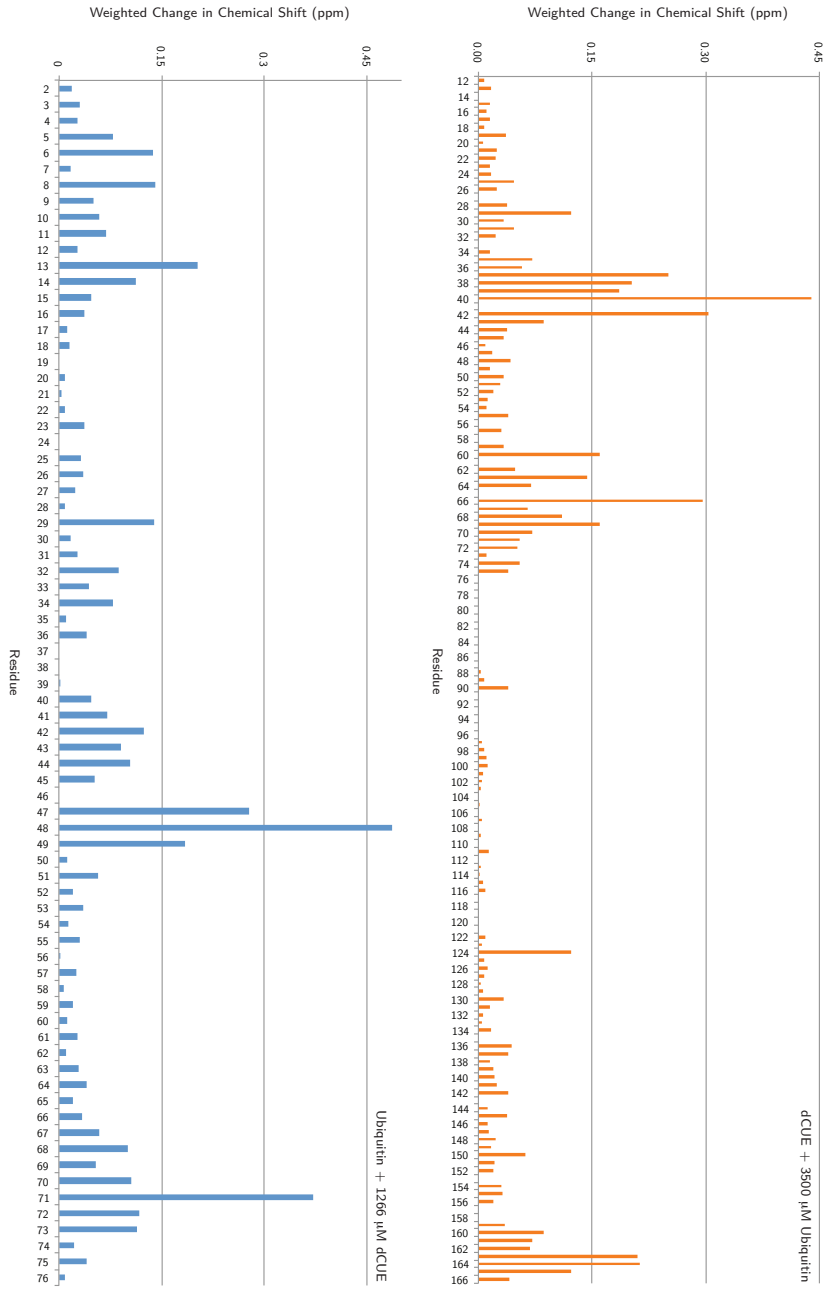


Figure 6.4: Chemical Shift Differences Caused by Ubiquitin:dCUE Binding - **Top** dCUE residues with magnitude of chemical shift differences plotted. **Bottom** Ubiquitin residues with observed chemical shift changes. Larger bars represent a greater change in peak position in the HSQC during the titration and thus are indicative of residues that interact (174).

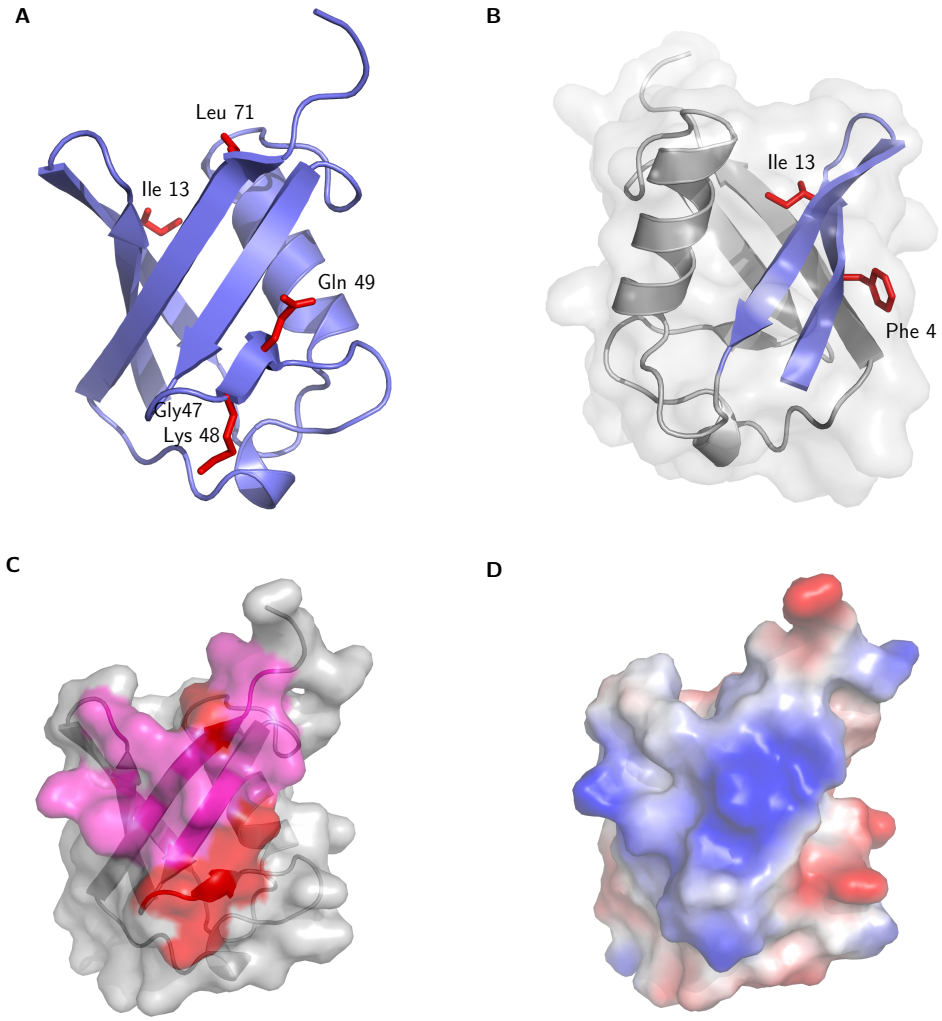


Figure 6.5: Ubiquitin Residues Affected by dCUE Binding - **A** and **C** Residues in red show medium or large chemical shift differences ($\Delta\delta \geq 0.15$) when dCUE is added. Magenta shaded residues (**C**) have a small chemical shift difference ($0.08 \leq \Delta\delta \leq 0.15$). **B** Ile 13, which has a large secondary chemical shift (figure 6.4), maps to the far side of the ubiquitin molecule relative to the patch of interacting residues. In the ubiquitin structure, residue Ile13 is close to aromatic residue Phe4. Due to a ring current effect, aromatic residues have a large effect on neighbouring chemical shifts. If the binding of dCUE to ubiquitin does not directly occur at the site of Ile13 but causes a relatively small knock-on conformational change which alters an interaction between Ile13 and Phe4 it may have a disproportionately large change in observed chemical shift and thus not indicate a direct interaction with dCUE (271). **D** The surface plot of electrostatic potentials shows that the interacting residues align with positively charged residues (blue)

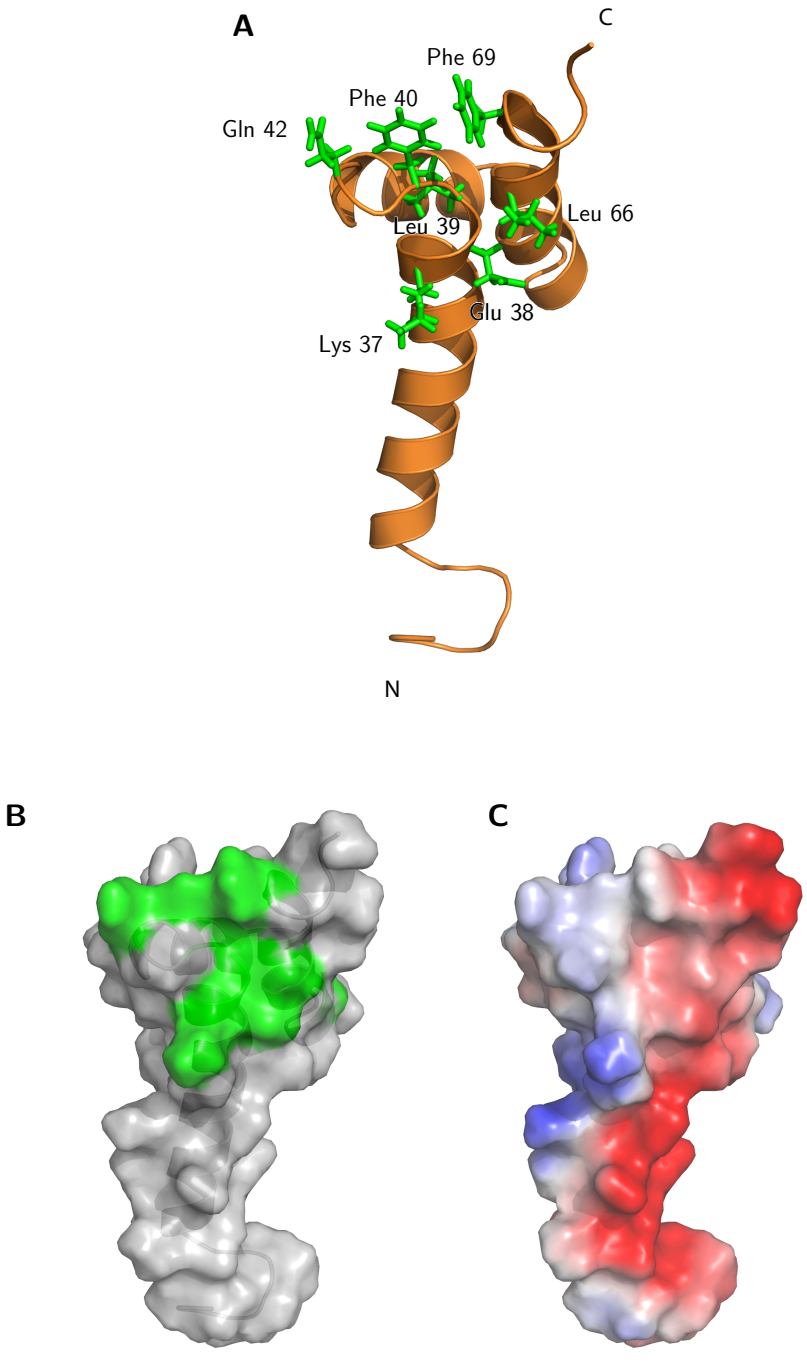


Figure 6.6: SMARCAD1 CUE Domain 1 Residues Affected by Ubiquitin Binding - **A** and **B** Residues in green show medium or large chemical shift differences when ubiquitin is added: $\Delta\delta \geq 0.15$. **C** Some of the residues involved with ubiquitin binding show negative electrostatic potentials (red), which complements the positive binding site on ubiquitin (figure 6.5).

6.2.1.2 Dissociation Constant

The binding affinity of the interaction between ubiquitin and dCUE was investigated using the results of NMR titration experiments. The dissociation constant, K_d, of interacting residues can be determined by plotting the degree of saturation ($^{\Delta_{obs}}/_{\Delta_{max}}$) with free ligand concentration (figure 6.7) (272).

$$y = \frac{\Delta_{obs}}{\Delta_{max}} = \frac{[Ligand]_{free}}{[Ligand]_{free} + K_d}$$

An *in silico* iterative approach was used to fit the curve non-linearly using the rearranged function:

$$\Delta_{obs} = \Delta_{max} \left(\frac{[Ligand]_{free}}{[Ligand]_{free} + K_d} \right)$$

where Δ_{obs} is the observed chemical shift difference , Δ_{max} the theoretical maximum chemical shift upon total saturation and $[Ligand]$ the ligand concentration, where 'ligand' is the unlabelled protein being titrated into a fixed concentration of labelled protein. Free ligand concentration was determined iteratively, based the percentage saturation and an initial assumption of 1:1 binding.

$$[Ligand]_{free} = [Ligand]_{total} - [Ligand]_{bound}$$

$[Ligand]_{total}$ is the total concentration of ligand protein added at any particular titration step and $[Ligand]_{bound}$ is the concentration of ligand bound, a function of target protein concentration, $[Protein]$, and its proportion of saturation. The target protein is the isotopically labelled protein of fixed concentration.

$$[Ligand]_{bound} = \frac{\Delta_{obs}}{\Delta_{max}}[Protein]$$

Figure 6.7 shows the chemical shift differences for selected residues as a function of added ubiquitin. Residues with the largest chemical shift difference were plotted (i.e. those involved in the binding event; section 6.2.1.1). From the plot of changes in dCUE peak position upon titration with unlabelled ubiquitin, two distinct families are seen. These correspond to each of the two CUE domains and have different dissociation constants, suggesting two distinct interactions. The residues from CUE-1 (Phe 40, Gln 42, Met 68 and Phe 69) give a set of binding interactions with K_d = 1800 μM ± 156 and residues Ala 163 and Phe 164 from CUE-2 have an average K_d of 6928 μM ± 458.

In contrast, the titration experiment with ^{15}N-labelled ubiquitin and unlabelled dCUE suggests a single binding surface on ubiquitin. All residues affected significantly by addition of dCUE have a very similar dissociation constant: K_d = 592 μM ± 61. This apparent dissociation constant suggests a stronger affinity for dCUE than either

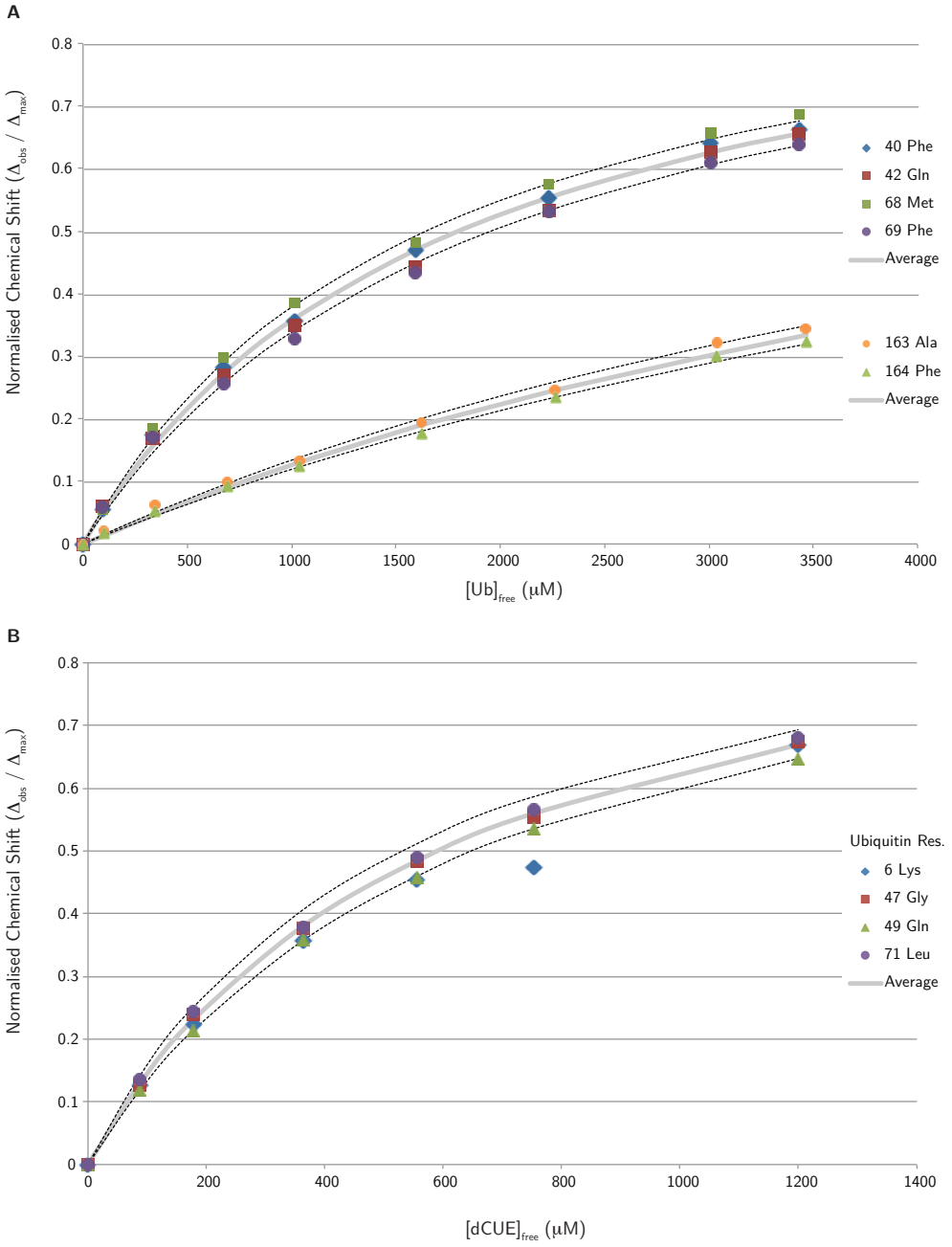


Figure 6.7: Chemical Shift Changes from Double CUE and Ubiquitin Interactions - **A** Relative chemical shift changes of six dCUE residues upon titration with ubiquitin. Residues from two different helices (both within the first CUE domain) have a dissociation constant (K_d) calculated as 1800 μM ± 156. Residues Ala 163 and Phe 164 from the dCUE construct had an average K_d of 6928 μM ± 458, representing a separate, weaker binding interaction. **B** Chemical shift changes of selected ubiquitin residues upon titration with unlabelled dCUE. A single K_d was calculated as 592 μM ± 61. In both plots, chemical shift changes have been normalised against estimated maximum shift. Dashed lines show error range of ± 1 standard deviation from the respective mean.

of the individual CUE domains but as ubiquitin only has a single binding surface that interacts with both CUE domains, which have different affinities, the apparent dissociation constant from ubiquitin is a complex function of the two binding events. It is not readily possible to conclude whether there is any degree of cooperativity involved or exact stoichiometries (269).

From the dissociation constant results, it can be concluded that ubiquitin binds dCUE very weakly but with higher affinity for CUE-1 than CUE-2 on a specific, structurally discreet, site. The observed dissociation constant for mono-ubiquitin was approximately 1.8 mM, which is generally considered too weak to be biologically meaningful (273). This is not surprising given the abundance of ubiquitin in the cell and its myriad of functions and binding. A tight interaction with free monoubiquitin would not be target or function specific. This experiment shows the interaction between free mono-ubiquitin and dCUE but, in a biological context, it seems more likely that other components would be involved to produce a stronger affinity and more tightly controlled function. It is conceivable that the true biological interaction involves a covalently bound, ubiquitinated protein and / or the action of flanking regions.

6.2.2 Docking Model

Based on chemical shift changes observed in the ubiquitin-CUE-1 titration experiments, a docking model was calculated using the HADDOCK (High Ambiguity Driven biomolecular DOCKing) software package (figure 6.8)(199). Ubiquitin residue Ile13 and dCUE residue Gln42, which both have significant shift changes (section 6.2.1.1), were excluded from the HADDOCK calculations. The large chemical shift changes seen for these residues may be a result of their proximity to aromatic residues as opposed to being at the binding interface. Ring current effects from aromatic residues produce a large effect on chemical shift of nearby nuclei. If a relatively small conformational change occurs to a residue, not as a result of a direct protein-protein interaction but as a consequence of local reordering, it may alter the residue's proximity to the aromatic residue and thus affect the ring current effect. In the case of these residues, the chemical environment would be changed significantly and we would see a larger shift change (271).

6.2.2.1 Comparison with Other Helical Ubiquitin-binding Domains

It is important to appreciate that, while based on experimentally determined restraints, HADDOCK models do not serve as a robust alternative to a full structural solution of a protein complex such as ubiquitin-CUE-1. Regardless, the HADDOCK model and

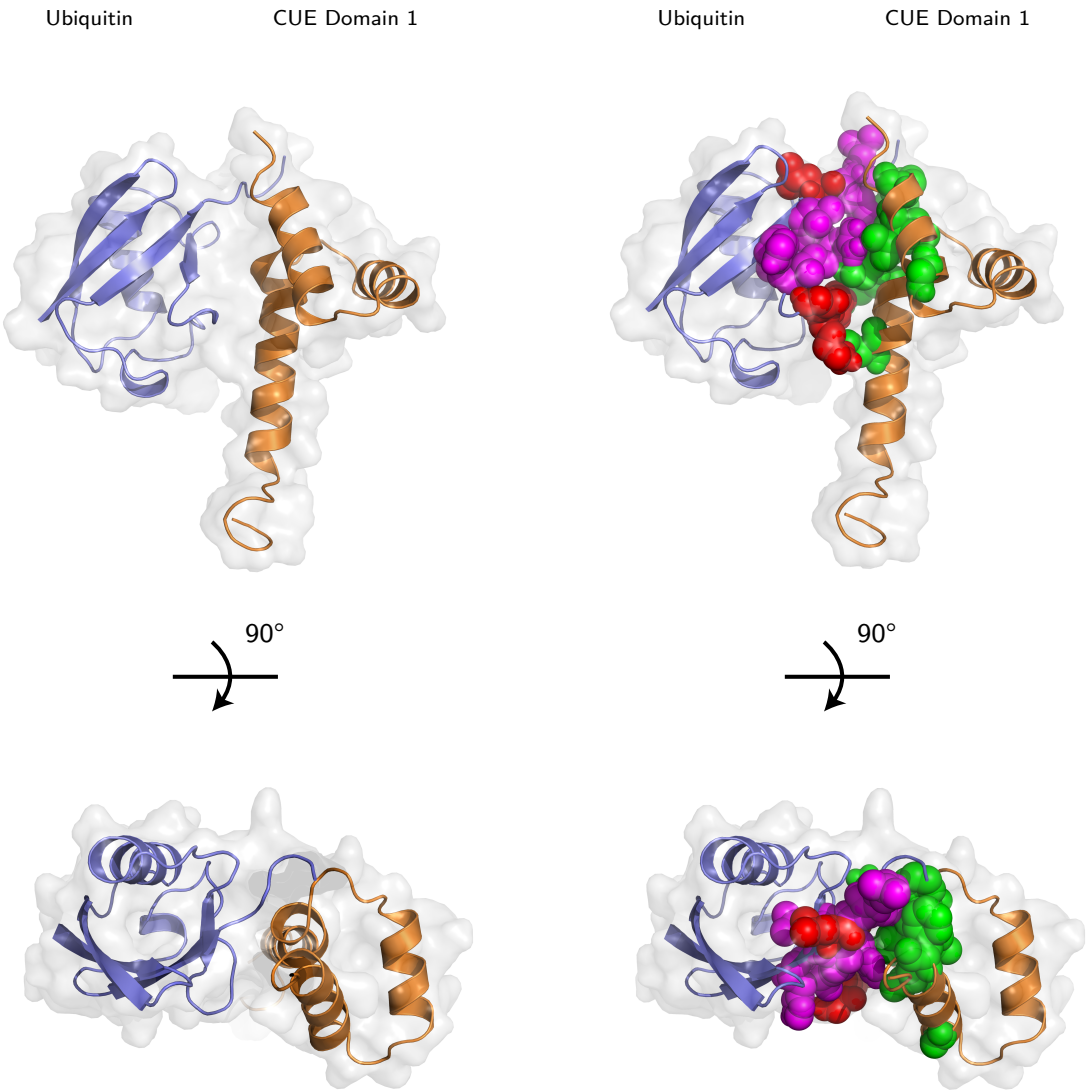


Figure 6.8: SMARCAD1 CUE-1-Monoubiquitin Binding Model - Best scoring HADDOCK CUE-1-ubiquitin docking model based on ambiguous interactions (199). Right image shows interacting residues used in HADDOCK calculations drawn as spheres based on large chemical shift changes. Ubiquitin residues with large shifts shown in red, smaller shifts shown in magenta. Green spheres correspond to dCUE residues with large chemical shift changes.

titration results do show that SMARCAD1 CUE-1 binds to the same site on ubiquitin as other CUE proteins and helical ubiquitin binders (figures 6.9 and 6.10) (168). Some ubiquitin binding domains (UBDs) are a single helix with hydrophobic residues at the ubiquitin-binding interface. The model of SMARCAD1 CUE-1 binding presented here has a protruding N-terminal helix.

6.3 SUMO

SUMO (Small Ubiquitin-like MOdifier) is a family of proteins related to ubiquitin. SUMO proteins have been identified as a covalent modification in several suspected binding partners of SMARCAD1, such as PCNA, KAP1 and histones (50, 113, 122, 124, 278). Given the similarity to ubiquitin and the known modification targets, the binding interaction with SMARCAD1 dCUE was investigated.

6.3.1 NMR Titrations

6.3.1.1 Binding Site

In a similar manner to the ubiquitin titration experiment, a titration was carried out to investigate the binding of SUMO-1 to dCUE. Figure 6.11 shows an overlay of dCUE ^{1}H, ^{15}N-HSQC spectra after different concentrations of unlabelled SUMO-1 had been added. Compared to the ubiquitin titration experiments (section 6.2.1.1), a more limited concentration range was used with SUMO-1 titrations due to material constraints.

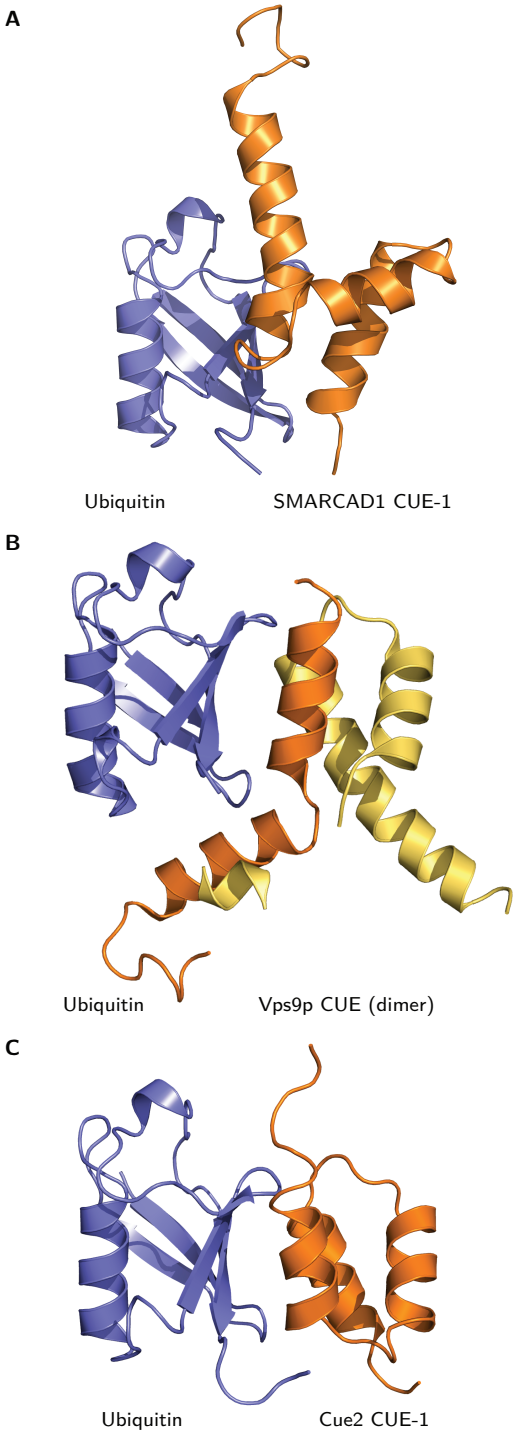


Figure 6.9: CUE-Ubiquitin Binding - Comparison of HADDOCK model for SMARCAD1 CUE-1-ubiquitin binding with published structures. All shows a 'side-on' packing with ubiquitin and the helix bundle. **B** Vps9p CUE domain forms a dimer (PDB: 1P3Q) (169) whereas **C** Cue2 CUE-1 (PDB: 1OTR) (167) is a small helix bundle similar to another ubiquitin binding domain, the UBA domain (figure 6.10)

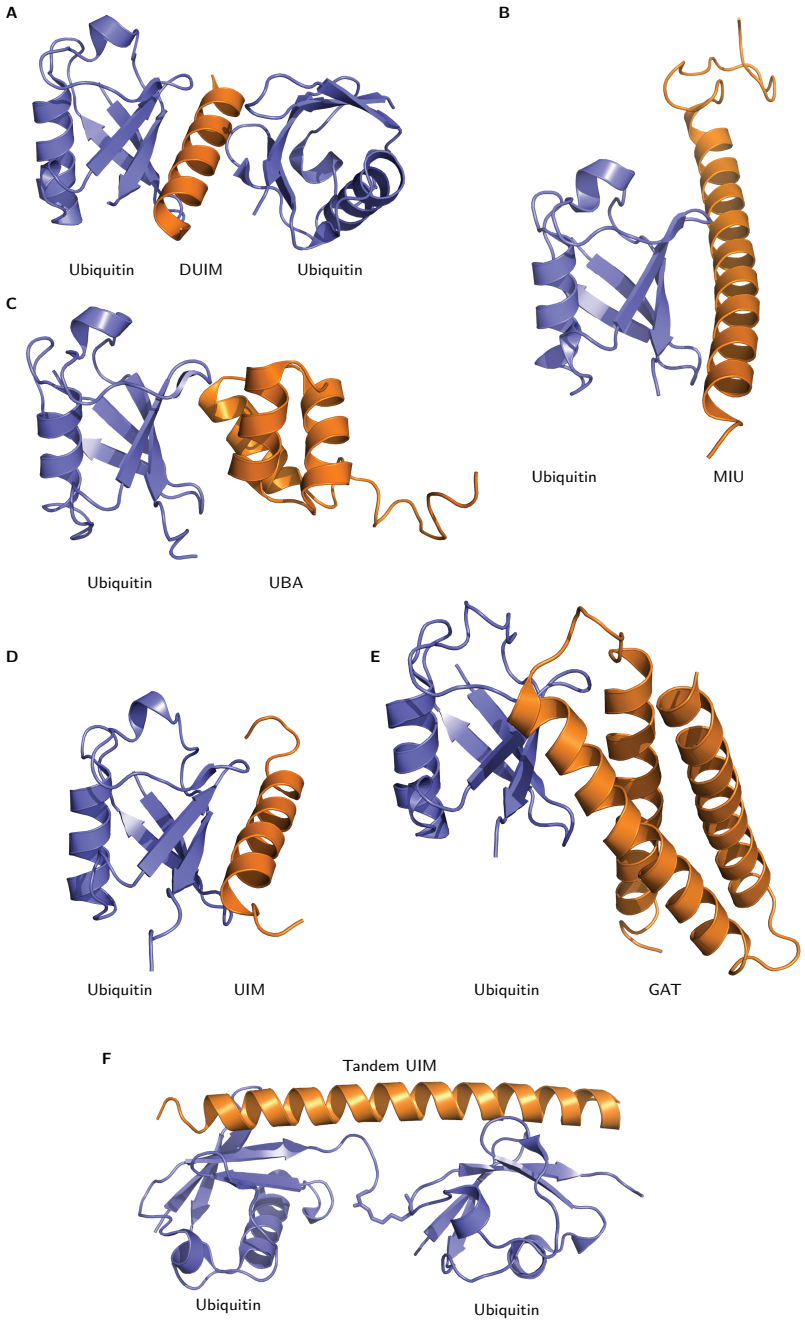


Figure 6.10: Comparison of Non-CUE Helical Ubiquitin Binding Domains (UBDs) - Many ubiquitin binding domains are weak-binding helical proteins with a range of structures and binding modes, including binding of more than one ubiquitin molecule. Most helical UDBs bind to the same surface on ubiquitin, centred around Ile44. The ubiquitin-associated (UBA) domain of Dsk2p (PDB: 1WR1) (274, 275). Vps27 ubiquitin interacting motif (UIM) (PDB: 1Q0W) (276). Double-sided UIM (DUIM) from Hrs protein (PDB: 2D3G) (277). Rabex-5 motif interacting with ubiquitin (MIU) (PDB: 2FIF)(181). GAT domain from ADP-ribosylation factor binding protein GGA3 (PDB: 1YD8) (175). Tandem UIM from Rap80 (PDB: 3A1Q) (167, 184).

A number of peaks shifted in a concentration dependent manner, suggesting a binding interaction. Based on the backbone assignments of dCUE (section 4.3), these mainly correspond to residues at the N-terminus of the first CUE domain, predominantly in an unstructured section prior to the helix in the solved structure.

The sequence of dCUE from residue 16-28 is **I**SE**L**ED**L**SE**L**ED**L**, which has a number of evenly-spaced hydrophobic residues (bold) with polar residues in between, forming a hydrophobic surface along one side of the amphipathic helix. As discussed in section 5.7.1.4, the N-terminal portion of α-helix 1 is interesting and possibly more dynamic than the rest of the core helix bundle. It is possible that this section, including the residues before residue 21, form a protein-interaction surface which induces structure upon binding.

As the SUMO titration range was smaller than that used with ubiquitin, difference in peak position over the course of the titration experiment was notably smaller. Figure 6.12 shows the difference in peak chemical shift for each dCUE residue after the final SUMO titration step, 819 μM, compared with residue chemical shift differences from a comparable step of the ubiquitin titration, 700 μM.

The residues that exhibit the greatest shift differences are the most N-terminal residues of the construct that belong to SMARCAD1 sequence (i.e. not the tag). Although most of the tag sequence is unassigned, two residues, Gln 12 and Asp 13 (Pro

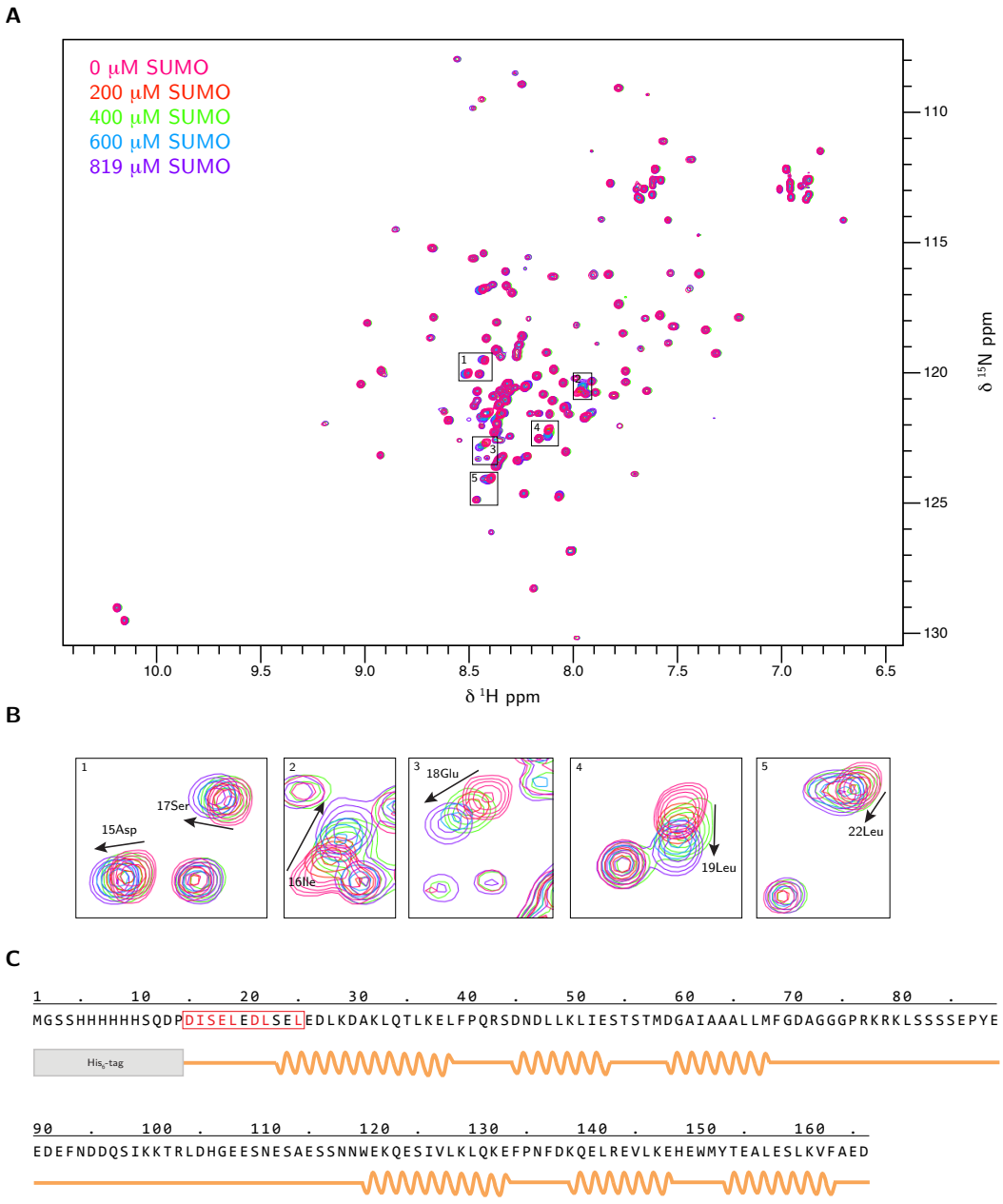


Figure 6.11: NMR Titration Experiment of ^{15}N-dCUE Shifts with SUMO-1 - **A** Overlay of five ^{1}H, ^{15}N-HSQC spectra for dCUE, each with different SUMO-1 concentrations. **B** Selection of magnified ^{1}H, ^{15}N-HSQC peaks that shifted considerably (boxed in A). The sign and magnitude of shift is consistent with increasing ligand concentration. **C** Sequence and secondary structure of dCUE construct with significant shifts shown in red. Most shifts occur at the N-terminal end of the dCUE sequence but none of the peaks from residues in the His_6-tag shift. Compared to the ubiquitin titration, the size of the shifts seen are smaller but this is partly due to affinity and the smaller concentration range, with a lower maximum ligand concentration.

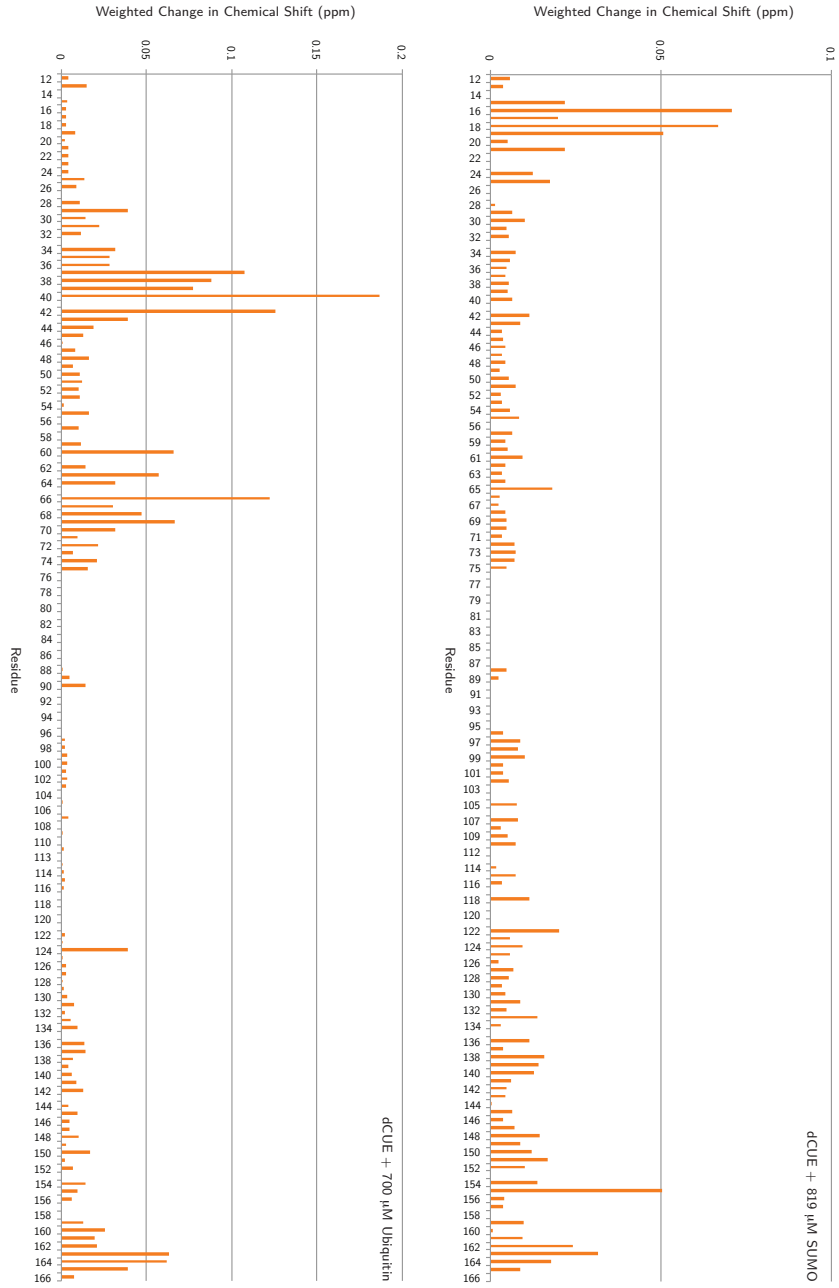


Figure 6.12: Comparison of Chemical Shift Differences Caused by SUMO:dCUE Binding and Ubiquitin:dCUE - Top Size of chemical shift differences with 819 µM SUMO-1. **Bottom** Chemical shift differences with a comparable concentration of ubiquitin, 700 µM. Larger bars represent a greater change in peak position on the HSQC during the titration and are indicative of residues, which interact with the ligand.

14 is also part of the tag but doesn't give a peak in the ^{1}H, ^{15}N-HSQC), are assigned and directly precede the SMARCAD1 residues that have large chemical shift differences ($\Delta\delta > 0.05$). The residues from the tag do not have a large change in chemical shift, which indicates that they do not interact significantly with SUMO-1. This suggests that the effect of SUMO-1 on N-terminal dCUE residues is a real result of site-specific dCUE binding.

Comparing the relative magnitude of residue chemical shift differences observed for dCUE + SUMO-1 (819 μM) and dCUE + ubiquitin (700 μM) gives an indication of the importance of specific interactions relative to the background 'noise'. The peaks which have the largest chemical shift difference in the dCUE + SUMO don't change as much as the largest differences in the dCUE + ubiquitin experiment. SMARCAD1 dCUE residues Ile 16, Glu 18, Leu 19 and Glu 155 have a moderate shift difference upon SUMO-1 titration ($\Delta\delta > 0.05-0.75$ ppm) but not as many or as large as for the ubiquitin titration. Given that the major residues that shift are clustered directly at the start of the SMARCAD1 dCUE sequence, it is conceivable that some of the SMARCAD1 sequence not included in the construct are also involved in SUMO-1 binding.

A previously reported SUMO binding motif (SBM) is present in the full-length SMARCAD1 protein but not in the dCUE construct. Approximately 20 residues upstream of the dCUE construct in SMARCAD1 there is a sequence which fits the V/I-x-

V/I-V/I SBM (figure 6.13) (279). Recognition of this motif by SUMO is very different than the helical binding motifs seen in ubiquitin binding domains. The SBM creates an extended structure by binding between the α-helix and β-strand of SUMO (280). In the case of dCUE, there is evidence for weak SUMO binding in a region that may be susceptible to induced helical structure and therefore distinct from the SBM. One possibility is that the SBM seen in full-length SMARCAD1 does interact with SUMO but also that the N-terminal, putatively helical, region of CUE-1 binds in a manner more similar to that seen with ubiquitin binding (helical).

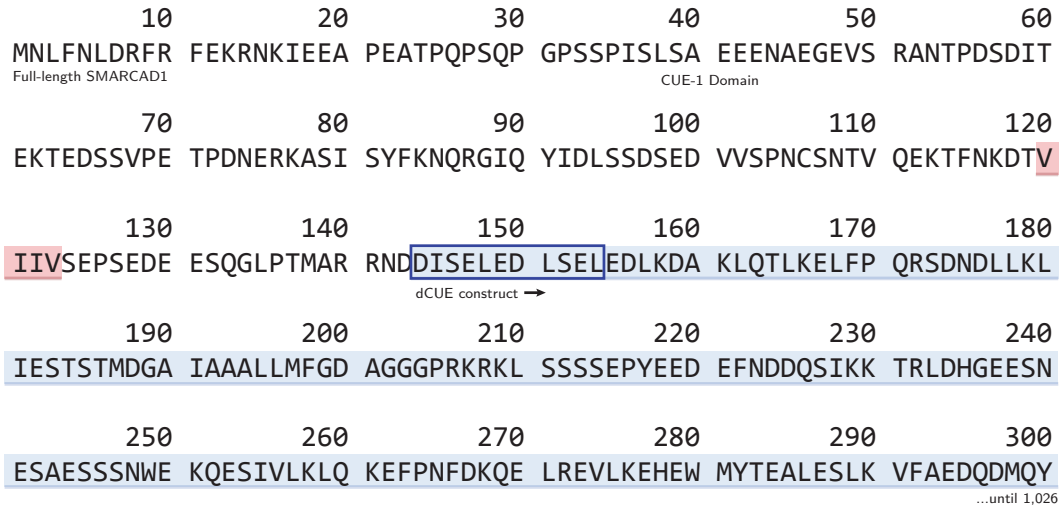


Figure 6.13: Potential SUMO Binding Motif Upstream of dCUE - Sequence of the first 300 residues of full-length SMARCAD1. The dCUE construct used in the NMR experiments begins at SMARCAD1 position 144 (blue). A SUMO binding motif was identified with consensus sequence V/I-X-V/I-V/I (red) (279, 280), which is located 20 residues upstream of the start of the dCUE sequence. Residues in the region shown to be affected by SUMO-1 titration are boxed.

6.3.1.2 Dissociation Constant

As with ubiquitin binding, the dissociation constant, K_d, for the dCUE:SUMO-1 interaction was determined from the rate at which peak position changes for those residues with the largest chemical shift differences (figure 6.14). Only N-terminal residues where the corresponding peak was resolved at all titration points were used (i.e. if peaks overlapped for intermediary positions, it was not possible to accurately plot these points). The K_d for the interaction was determined to be 1674 μM ± 56, which is similar to that of ubiquitin although the residues involved are different. Like that found for the dCUE interaction with ubiquitin, this dissociation constant suggests too weak an interaction to be biologically meaningful but may be indicative of SUMO binding in a different context (e.g. SUMO-1 covalently linked to a specific protein and / or possibly a different member of the SUMO family). As mentioned above, it may be that some of the flanking SMARCAD1 sequence is missing from the dCUE construct, which may also affect the binding affinity.

6.4 PCNA

As mentioned in section 1.3.2.2, PCNA (proliferating cell nuclear antigen) has been identified as a major binding partner of full-length SMARCAD1. PCNA is a homotrimer of *ca* 30 kDa subunits and at least three lysine residues per subunit are

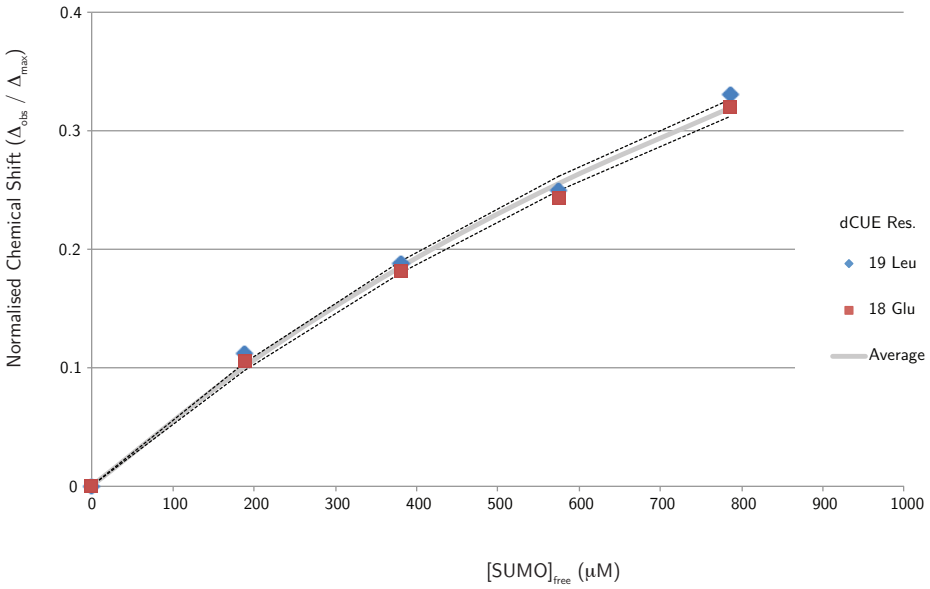


Figure 6.14: Chemical Shift Change from Double CUE and SUMO Interactions - Relative chemical shift changes of two N-terminal dCUE residues upon titration with SUMO. A dissociation constant (K_d) was calculated as 1674 μM ± 56 using an iterative process of non-linear curve fitting. Chemical shift changes are normalised relative to projected maximum shift (fitted using xvgr and Grace; section 2.5.4.1 (197)). Dashed lines show error range of ± 1 standard deviation from the mean. The concentration range (0-819 μM) of SUMO-1 used was less than that in the ubiquitin titration.

capable of being covalently linked with SUMO and/or ubiquitin modifications (either mono or polyubiquitin chains) (121, 123, 124).

Previous studies have shown pull-down and co-localisation of nuclear PCNA with SMARCAD1. Cell-based experiments have shown a clear interaction with PCNA and excluded the possibility that only PCNA interacting protein (PIP) boxes (PCNA-binding motifs) found within SMARCAD1 are responsible for PCNA-SMARCAD1 interaction (108, 110). It was hypothesised that, given the range of PCNA modifications with ubiquitin / SUMO, that the SMARCAD1 double CUE domain may be involved with PCNA binding and that this may be dependent on a particular set of modification states.

NMR titration experiments looking at binding interactions between PCNA and dCUE were performed using recombinant (unmodified) PCNA. Western blot analysis of pull-down experiments using both recombinant PCNA and nuclear extract protein mixed with either dCUE or full-length SMARCAD1 was carried out to better understand the role of the dCUE domain and post-translational modifications in PCNA binding.

6.4.1 NMR Titrations

A titration experiment with ^{15}N-labelled dCUE and unlabelled PCNA was used to investigate whether the double CUE domain of SMARCAD1 was involved in the binding interaction. Figure 6.15 shows an overlay of dCUE ^{1}H, ^{15}N-HSQC spectra after different concentrations of unlabelled PCNA had been added. Very few peaks showed any significant and consistent shift although C-terminal residues Phe 163, Ala 164 and Glu 165 do show small shifts. The titration used a very broad range of PCNA concentration (0-3050 μM) with a constant concentration of ^{15}N dCUE (100 μM). As the few observed shifts are very small, especially given the large excess of PCNA, and the residues exhibiting a shift are at the C-terminus of the construct, it is plausible that the shifts are due to a non-specific interaction and rather a result of their relative accessibility within the labelled protein.

Although the titration experiment didn't produce significant concentration-dependent changes in peak position, one distinct effect was an apparent broadening of specific peaks. In other words, the peak intensity of certain peaks lessened as concentration of PCNA increased. This resulted in some peaks disappearing altogether. The effect is best observed when individual ^{1}H, ^{15}N-HSQC spectra are compared alongside one another (figure 6.16). To consider whether the broadening effect was specific, for par-

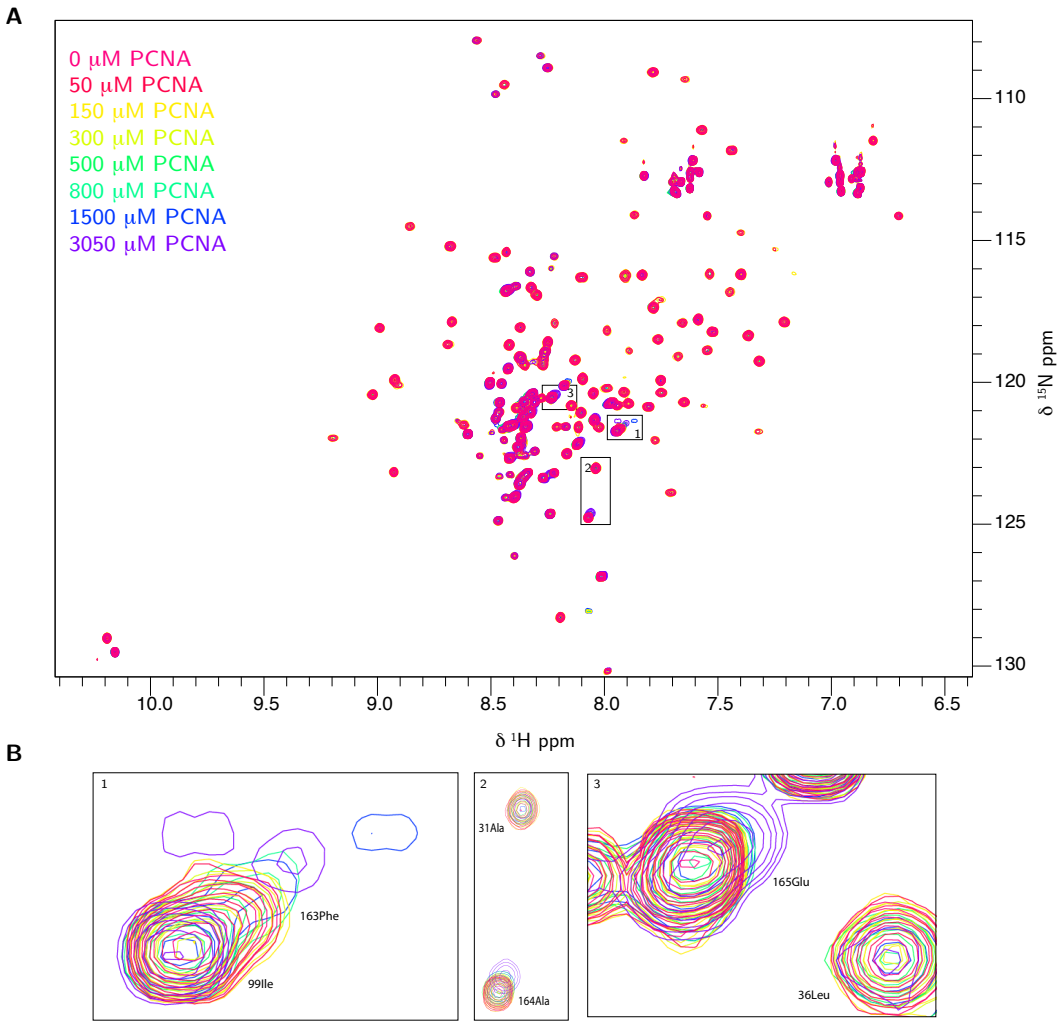


Figure 6.15: NMR Titration Experiment of ^{15}N-dCUE with PCNA - At very high concentrations of PCNA, some shifts in the C-terminal tail region are observed but are unlikely to be significant.

ticular residues, which may be indicative of residues that bind PCNA, a plot of relative decrease in intensity for each assigned residue position was made (figure 6.16).

The rate of chemical exchange affects the way in which binding events can be interpreted in an NMR experiment. Fast, sub-millisecond, exchange will typically present an averaged chemical shift position between two states (bound / unbound) whereas slow exchange exhibits line broadening and the appearance of a peak with chemical shift in the position corresponding to the bound species (236, 281). The affect of PCNA titration, however, doesn't appear to be a result of slow chemical exchange and therefore doesn't suggest a specific binding interaction. Figure 6.17 shows that the line broadening effect correlates with structured regions and shows no sign of residue-specificity beyond that. By contrast, the ubiquitin titration shows line broadening of structured residues but also shows site-specific additional broadening at the sites of interaction. One possible explanation for this general peak broadening may be due to increased viscosity from protein concentration, which may affect the way in which the ordered CUE domains tumble in solution.

6.4.2 Pull-Down Experiments and Western Blots

In addition to NMR titrations, pull-down experiments were performed to investigate dCUE binding to both recombinant untagged PCNA and PCNA from HeLa cell nuclear

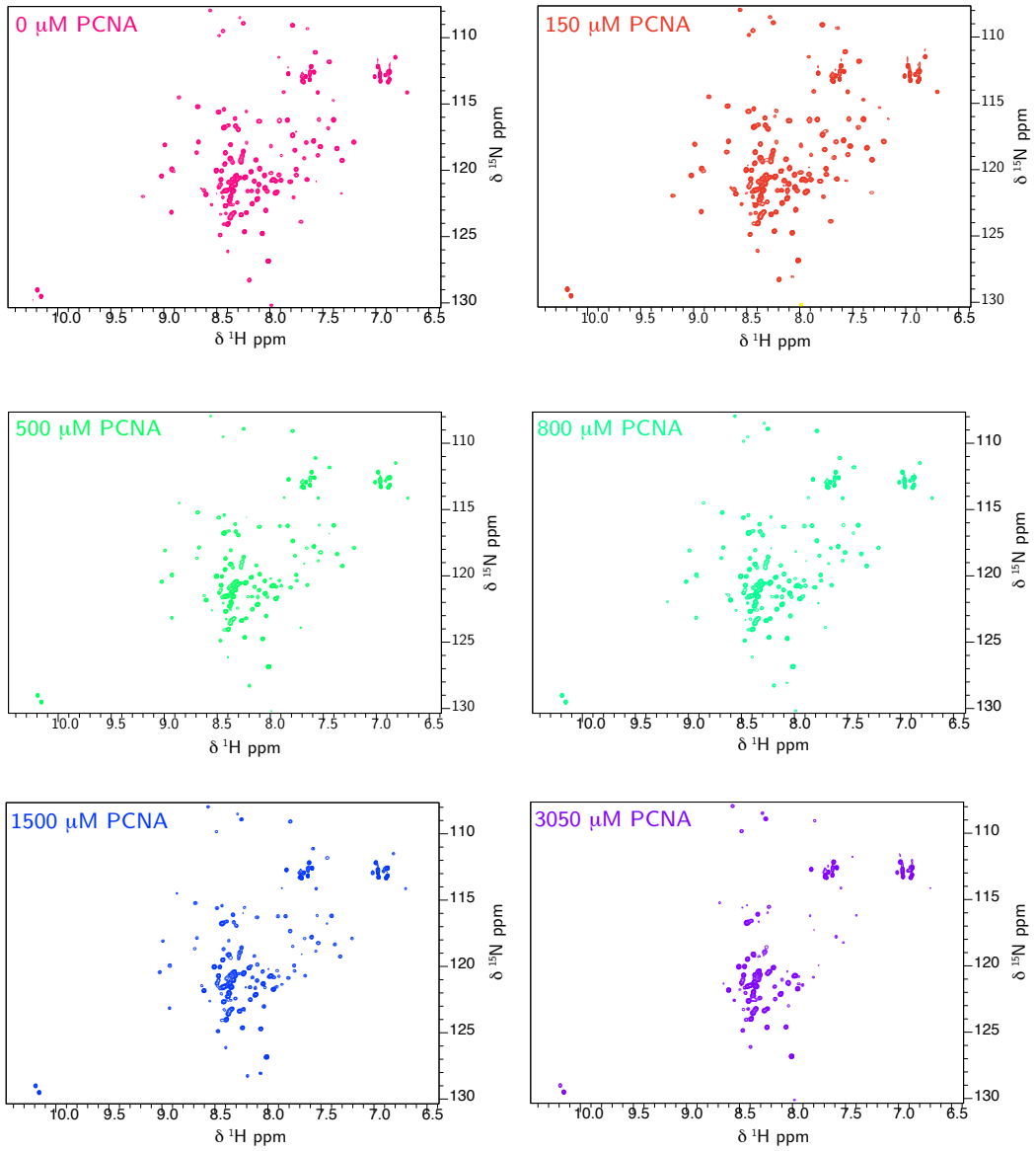


Figure 6.16: At High Concentrations of PCNA, Peak Intensities Vary - dCUE titration with PCNA shows line broadening and disappearance of some peaks. This is particularly evident when comparing dCUE spectra without PCNA (top left) with that of dCUE with high concentration (> 3 mM) PCNA (bottom right).



Figure 6.17: Protein Concentration-Dependent, Non-Specific Peak Broadening Occurs in Structured Regions - Plot of ^{1}H, ^{15}N-HSQC peak intensities and residue positions shows that titration of high concentration PCNA produces a relative drop in peak intensity (1 = peak intensity with no ligand). The drop in intensity is greatest with higher total protein concentration and correlates with the structured regions but is not specific to particular residues. Comparison with ubiquitin titration, which shows a similar pattern of drops in peak intensity with protein concentration, suggests an effect dependent on total protein concentration. This may be a result of increased sample viscosity affecting tumbling. In the case of ubiquitin, the relative drop in intensity is greatest in the first CUE domain and at the C-terminus. This may be because ubiquitin binds specifically at these residues and further affects line broadening. As the relative peak intensity changes for the dCUE with PCNA titration are consistent for all structured residues, it suggests that an unspecific ligand concentration effect is being observed as opposed to that of a specific binding event. The gap in the plots corresponds with a stretch of unassigned residues in the linker region.

extract. It has previously been reported that SMARCAD1 directly interacts with PCNA although it was not clear whether this interaction involved the double CUE domain or post-translationally modifications such as ubiquitination or sumoylation on PCNA (108). Pull-down experiments of (figures 6.18 and 6.19) show that full-length SMARCAD1 is capable of binding to both recombinant PCNA and PCNA from nuclear extract whereas the double CUE construct does not pull-down PCNA. These results and those from the NMR titration support the conclusion that the SMARCAD1-PCNA interaction is not mediated by the double CUE domain. It has recently been shown that the flanking regions either side of the double CUE domain can co-localise SMARCAD1 with PCNA to a limited extent (110) and therefore sites contained in those regions may be responsible for SMARCAD1-PCNA interaction.

6.5 KAP1

As mentioned in section 1.3.2.1, KAP1 (KRAB (Krüppel associated box)-associated protein 1) has been identified as a *bona fide* binding partner of SMARCAD1. KAP1 is a ubiquitously expressed 88.5 kDa protein involved in mediating gene silencing and response to DNA damage. Furthermore, KAP1 can be SUMOylated at multiple sites and mediates SUMOylation of other proteins (108, 113, 282, 283).

Based on mass spectrometry analysis of nuclear extract pull-down experiments using

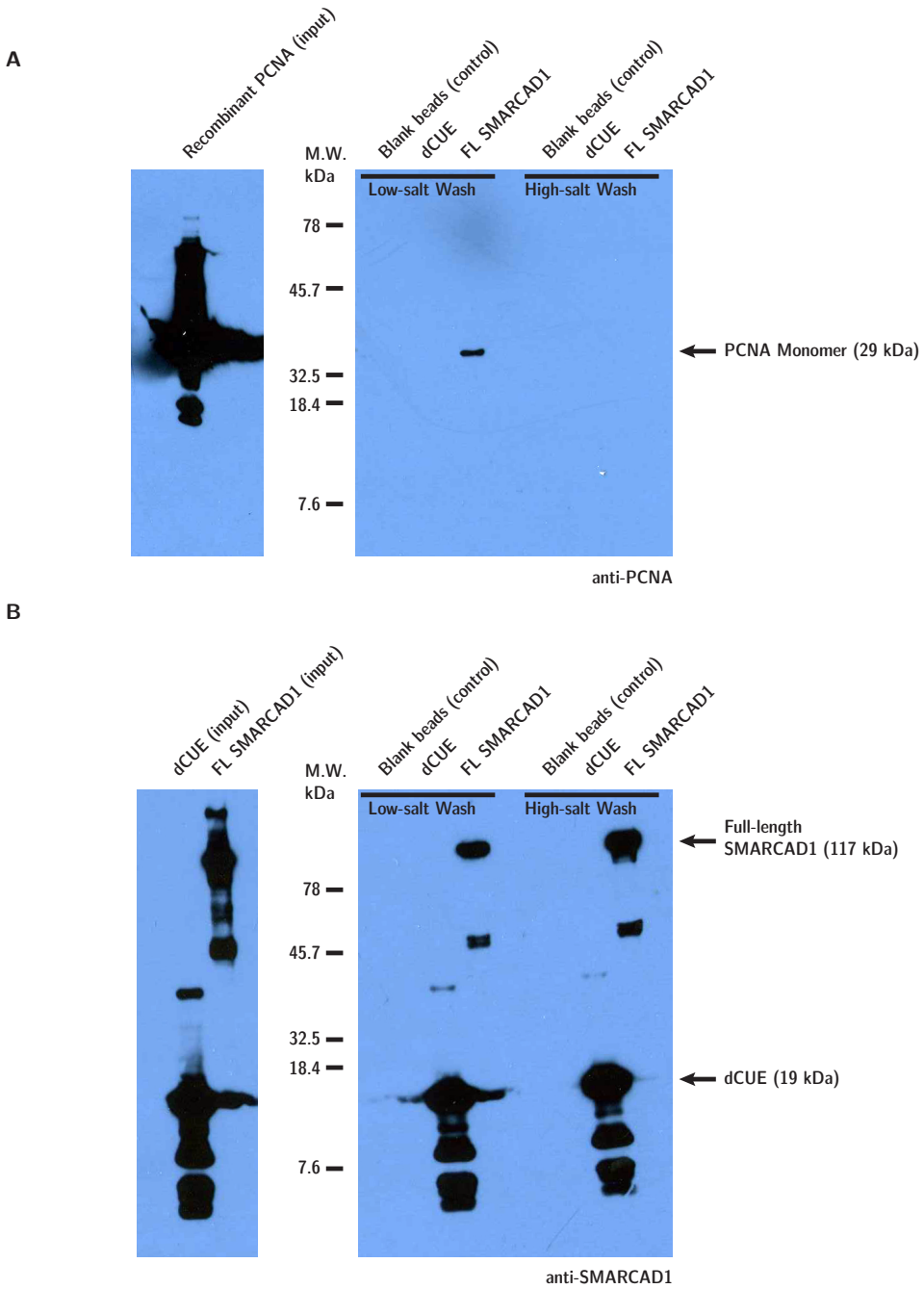


Figure 6.18: Pull-down Experiment Using Recombinant PCNA - Recombinant PCNA was mixed with either blank beads of dCUE / full-length SMARCAD1-coated beads and washed. Western blots of bead elution with **A** anti-PCNA shows that only full-length SMARCAD1 pulls-down PCNA and retains it during washing steps. A high-salt wash appears to be too stringent and no PCNA is retained. dCUE does not pull-down recombinant PCNA. Recombinant PCNA (highly concentrated) was also run on the same gel and blotted for size comparison (left). **B** Anti-SMARCAD1 western blot identifies both dCUE and SMARCAD1 proteins. SMARCAD1 and dCUE remain attached to the beads during both washing conditions. Some degradation products, which attach to the beads, are seen.

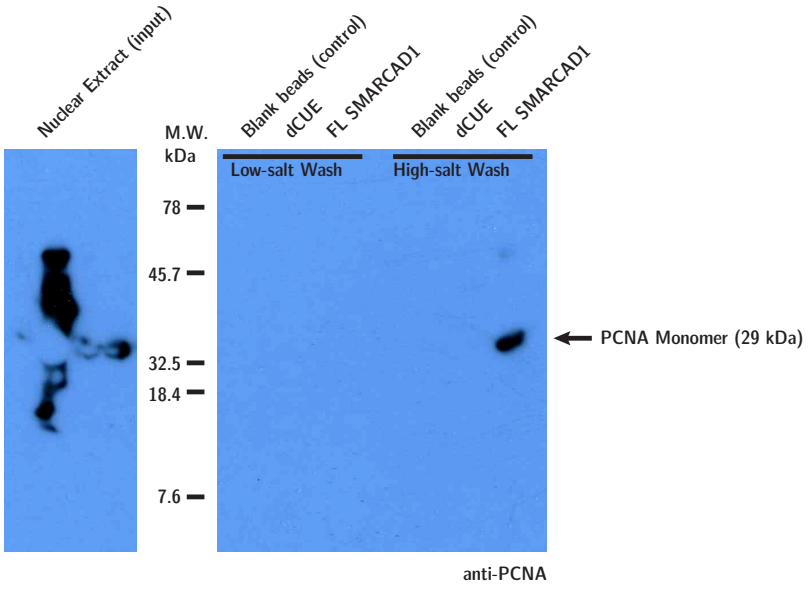


Figure 6.19: Pull-down Experiment Using Nuclear Extract and anti-PCNA Primary Antibody - Immobilised full-length SMARCAD1 is capable of pulling-down PCNA from HeLa cell nuclear extract but the double CUE domain doesn't. A control lane of nuclear extract was run (left), which shows up strongly with a saturated band of PCNA. Interestingly, the low-salt wash doesn't appear to retain PCNA from nuclear extract.

full-length SMARCAD1, Rowbotham *et al* (108) have shown that KAP1 is the most abundant binding partner of SMARCAD1 and is present in stoichiometric amounts. SMARCAD1 was also shown to be required to recruit KAP1 to heterochromatin.

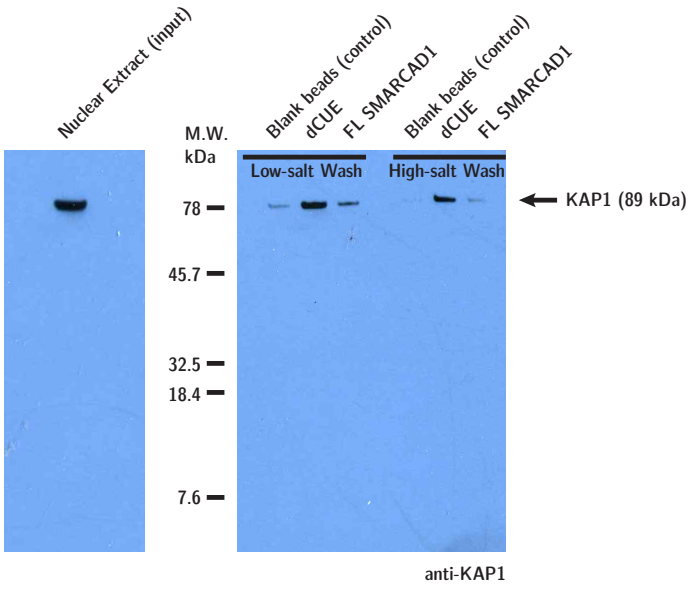


Figure 6.20: Pull-down Experiment Using Nuclear Extract and anti-KAP1 Primary Antibody - Both full-length SMARCAD1 and dCUE pull-down KAP1 from nuclear extract. This suggests that dCUE plays a key role in facilitating SMARCAD1-KAP1 binding; the major interaction of SMARCAD1.

A similar pull-down experiment using the double CUE construct and HeLa cell nuclear extract was carried out to investigate whether dCUE interacts with KAP1. Figure 6.20 shows the anti-KAP1 western blot of dCUE and SMARCAD1 nuclear extract pull-downs. Both full-length SMARCAD1 and dCUE pull-down KAP1 specifically with a strong band from dCUE suggesting that the double CUE domain from SMARCAD1 is responsible for the SMARCAD1-KAP1 interaction. As KAP1 is a stoichiometric

component of the SMARCAD1 complex, this suggests that the main function of dCUE may be in recruiting and binding KAP1. It remains unclear whether the dCUE-KAP1 interface occurs via a post-translational modification, such as a SUMO tag.

6.6 Additional Pull-down Experiments

A screen of potential dCUE binding partners was carried out using nuclear extract pull-down experiments and blotting against the proteins of interest (figure 6.21). Compared to a control, the only significant bands seen are high molecular weight bands in the anti-ubiquitin blot. Two bands appear at approximately 110 and 120 kDa (arrows), which suggests that dCUE binds to a ubiquitinated protein. Ubiquitin is approximately 8.5 kDa in mass and can form chains. It is conceivable that the two bands correspond to the same high molecular weight protein with differing numbers of post-translational subunit modifications. KAP1 is a *ca* 90 kDa protein which has been shown to be SUMOylated (a 12 kDa modification) although no evidence has been reported so far that shows it can be ubiquitinated. Other significant known SMARCAD1 binders with approximately the corresponding molecular weight include PARP1 (Poly [ADP-ribose] polymerase 1; 113 kDa), a protein involved with DNA repair, and GTF3C4 (General transcription factor 3C polypeptide 4; 92 kDa), a transcription factor for RNA polymerase III (108) both of which have been identified as ubiquitination targets (284).

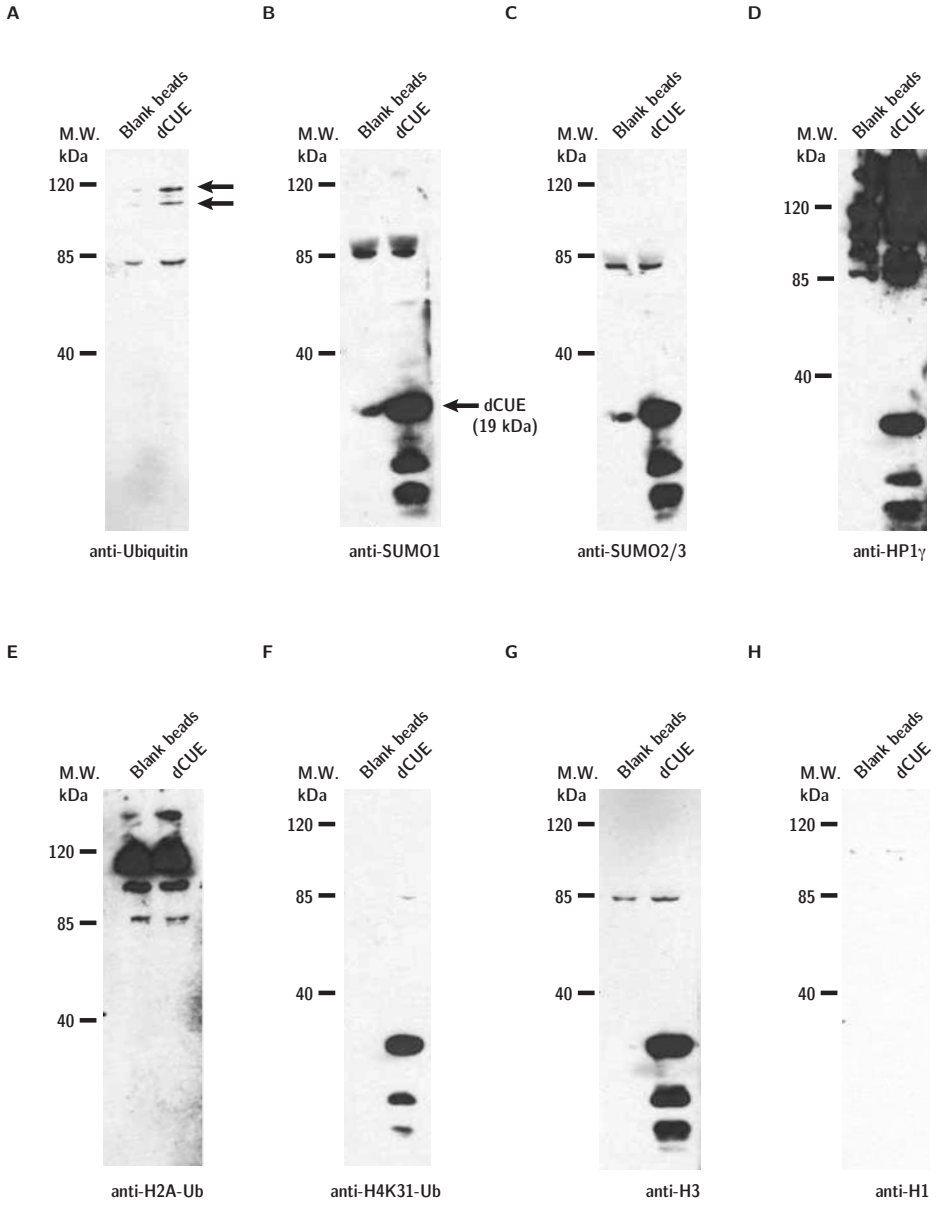


Figure 6.21: Pull-down Experiments Using Nuclear Extract - Initial screening results using a variety of antibodies against SMARCAD1 associated proteins, ubiquitin / SUMO and histone proteins (including ubiquitinated histones). Nuclear extract was applied to blank beads or beads coated with His_6-tagged dCUE. Any additional protein in the dCUE beads relative to the control would suggest some binding affinity. Five of the blots have three characteristic large bands in the column corresponding to the dCUE sample. This correspond to dCUE and likely dCUE degradation products. Primary antibodies raised against His-tagged proteins are likely to cross-react with the His-tagged dCUE construct and thus dCUE will also show up on the western blot. **A** anti-ubiquitin **B** anti-SUMO1 **C** anti-SUMO2/3 **D** anti-HP1γ **E** anti-ubiquitinated-H2A **F** anti-ubiquitinated-H4K31 **G** anti-Histone H3 **H** anti-Histone H1.

Chapter 7

Discussion

SMARCAD1 is a poorly understood chromatin remodeller with emergent roles in the fundamental nuclear processes of DNA replication and repair. The absence of any identified histone-modification recognition domains and the presence of ubiquitin-binding CUE domains mark SMARCAD1 out as unique among remodellers. The double CUE has been shown to be involved in coordinating SMARCAD1 association with its major binding partner, KAP1, suggesting an essential role in SMARCAD1 action. The work presented in this thesis has looked at the structure and function of the tandem CUE domains. This chapter will discuss the conclusions of this work in a broader context.

7.1 Structure of the Double CUE Domain

The solved structure of the SMARCAD1 double CUE domain addresses various questions and offers a suggestion for why initial crystallisation trials failed (section 3.5). From NMR experiments (chapter 5), it is clear that the dCUE domain is not a rigid

structure that adopts a discreet conformation. It consists of two, independent globular CUE domains with a long, dynamic, and largely unstructured linker. The flexible linker between each CUE domain of the dCUE construct is likely responsible for the failed crystallisation.

Previously, it was unknown whether the two SMARCAD1 CUE domains were structured relative to - or interacted with - one another. An unresolved question was whether the dCUE domain functioned as a single unit with cooperative substrate binding or acted independently with discreet functions. A beneficial outcome of the approach taken in this study, as opposed to further dissecting SMARCAD1 into individual CUE domains, has been to investigate how the CUE domains function together. This is the first study of the structure and function of a complete tandem CUE domain.

Structural analysis of the two CUE domains show a significant conformational difference between them. CUE-1 appears to have an extended N-terminal helix, which is dynamically different to the rest of the CUE domain (section 5.7.1.4). The significance of this remains unclear although it is directly adjacent to the ubiquitin binding site of CUE-1 so it is conceivable that the N-terminal section has a functional role associated with that of CUE-1 (section 7.3.1).

Surface electrostatic potentials of both SMARCAD1 CUE domains were calculated (22) and compared with those from two other CUE domains (figure 7.1).

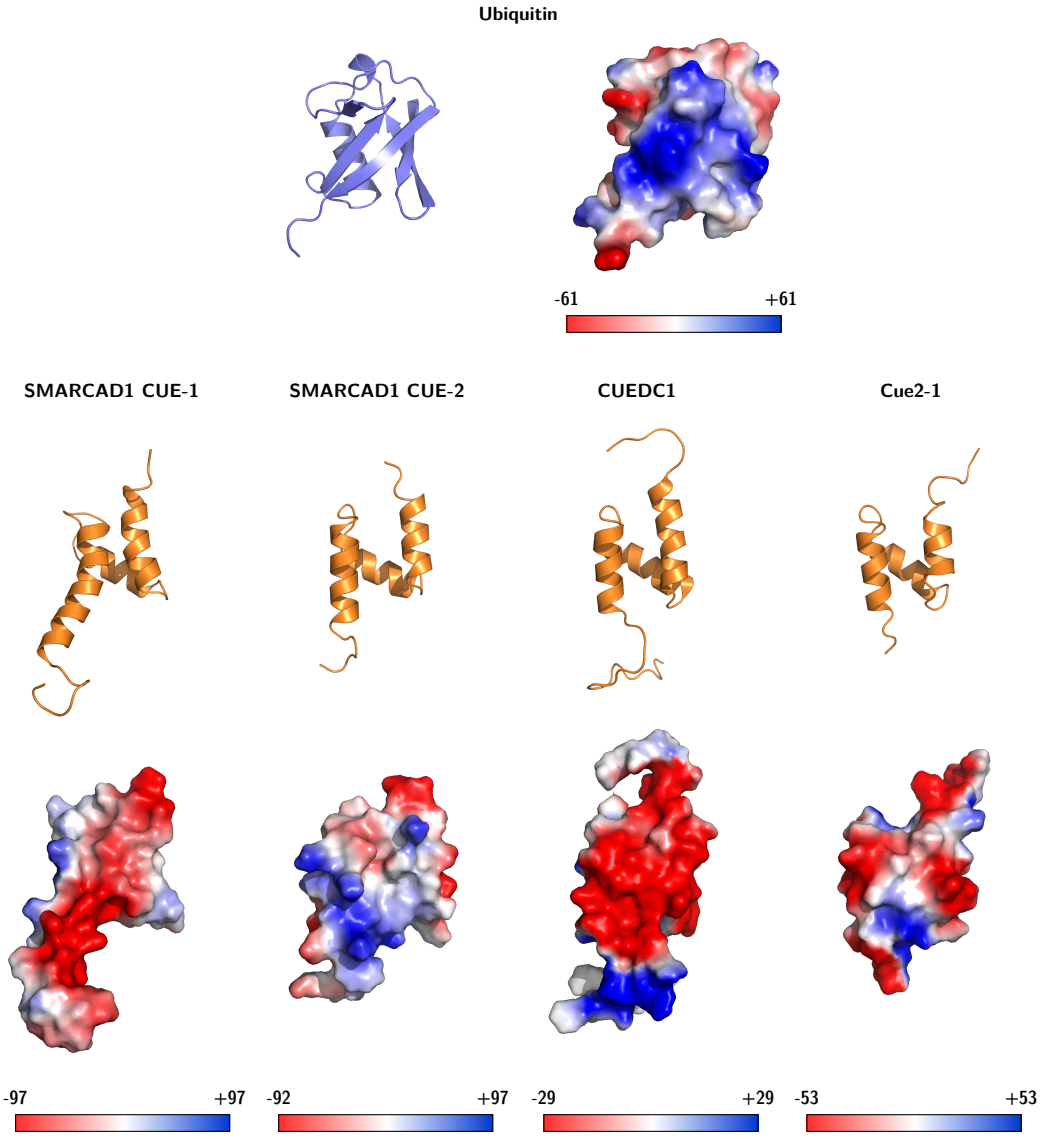


Figure 7.1: Electrostatic Potentials of CUE Domains - Electrostatic potentials of ubiquitin and CUE domains orientated to show binding surfaces. CUE-2 does not show ubiquitin binding surface but has been orientated as for other CUE domainsg (i.e. α-helix 1 and 3 towards the viewer). CUE-1 and other ubiquitin-binding CUE domains (CUEDC1 CUE and Cue2-1) have a negatively charged (red) surface which complements the positively charged ubiquitin binding surface (blue). CUE-2 is more positively charged than other CUE domains and does not complement ubiquitin. CUEDC1 CUE doman (PDB: 2DHY) and Cue2-1 (PDB: 1OTR) are shown for comparison.

7.2 Conserved CUE Sequence

CUE domains share a consensus sequence fingerprint which is important for structure and ubiquitin binding. Using the PROSITE CUE motif profile (section 3.2.1) (221), 81 CUE domains were found from the complete UniProt database (161). These include proteins from a wide variety of species (yeast to human). Sequences were aligned using ClustalW (figure 7.2) and residue similarity scores generated using ESPript (163, 285, 286). Similarity scores identify regions of conserved sequence homology among CUE domains. These were plotted onto the structure for CUE-1, identifying a highly conserved patch located around the C-terminal end of helix 1 (figure 7.3).

Highly conserved residues in α-helix 2 are hydrophobic residues (I/L/V) which pack into the core of the helix bundle. Highly conserved residues from α-helices 1 and 3 correlate closely with the residues in SMARCAD1 CUE-1 that interact with ubiquitin (section 6.2.1 and figure 7.2 **A**). Ubiquitin is the common binding partner of CUE domains and therefore it is not surprising that conserved residues are involved with ubiquitin recognition. Interestingly the SMARCAD1 CUE-1 domain is distinct in having a homologous 'AL' dipeptide sequence at the position of the highly conserved LL in helix 3 (green boxes, figure 7.4). In the case of CUE-1, the AL sequence is immediately followed by another leucine (which does produce an LL sequence run) but



Figure 7.2: Aligned CUE Domain Sequences from UniProt with CUE-1 Ubiquitin Interacting Residues - A SMARCAD1 CUE-1 residues that interact with ubiquitin (section 6.2.1) are highly conserved residues. Chemical shift changes from NMR titrations are not observed for proline residues (marked P) as they do not have a backbone NH thus a shift difference of 'zero' doesn't indicate that there is no interaction. **B** Complete list of all 81 CUE domain sequences found in UniProt using PROSITE CUE motif, highlighting conserved residues (red). Numbering corresponds to CUE domain profile not dCUE sequence.

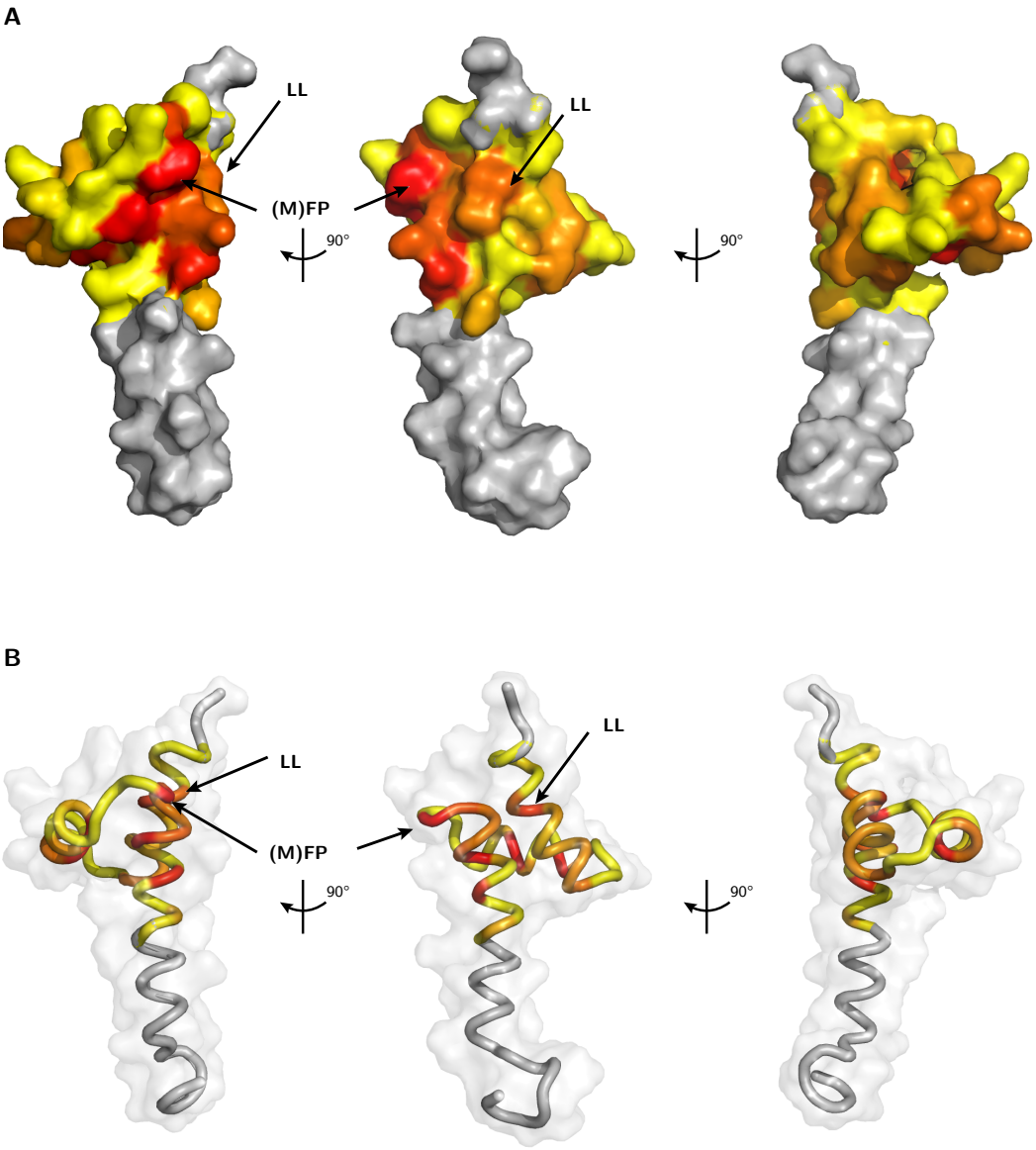


Figure 7.3: CUE Domain Sequence Homology - Surface (**A**) and cartoon (**B**) representations of ESPript similarity score (285, 286) mapped to CUE-1 domain structure. Residues shaded red represent highly conserved residues, yellow low level of sequence conservation and grey corresponds to residues not included in the CUE motif profile (PROSITE). It is important to note that the structured region of CUE-1 is bigger than the CUE motif profile.

crucially, this residue is located one-third helix turn from the position of the homologous residue and thus does not appear to have the same structural role as in other CUE domains.

SMARCAD1 CUE-1 is particularly striking compared to other CUE domains (figures 5.30) as helices 1 and 3 are much closer than in other CUE domains. This may be because the presence of a small alanine side-chain doesn't create bulky side-chain interactions between the two helices. In the case of most CUE domains, the highly conserved phenylalanine residue from helix 1 is involved in bridging the helices 1 and 3. In the case of SMARCAD1 CUE-2, as with other CUE domains, the phenylalanine packs between the helices. Due to a ring current effect from the aromatic ring, nearby nuclei display an upfield shifted chemical shift. In the case of CUE-1, this effect is not seen, adding additional evidence that the Phe residue (Phe40 in dCUE) is not packed towards the core of the helix bundle.

7.3 Binding Partners of the Double CUE Domain

7.3.1 CUE-1 Ubiquitin and SUMO binding

It has been shown that CUE-1 binds monoubiquitin specifically and very weakly (section 6.2). A dissociation constant of 1.8 mM is not biologically significant but does hint at an binding parter for CUE-1 which involves ubiquitin in some way. It is possible

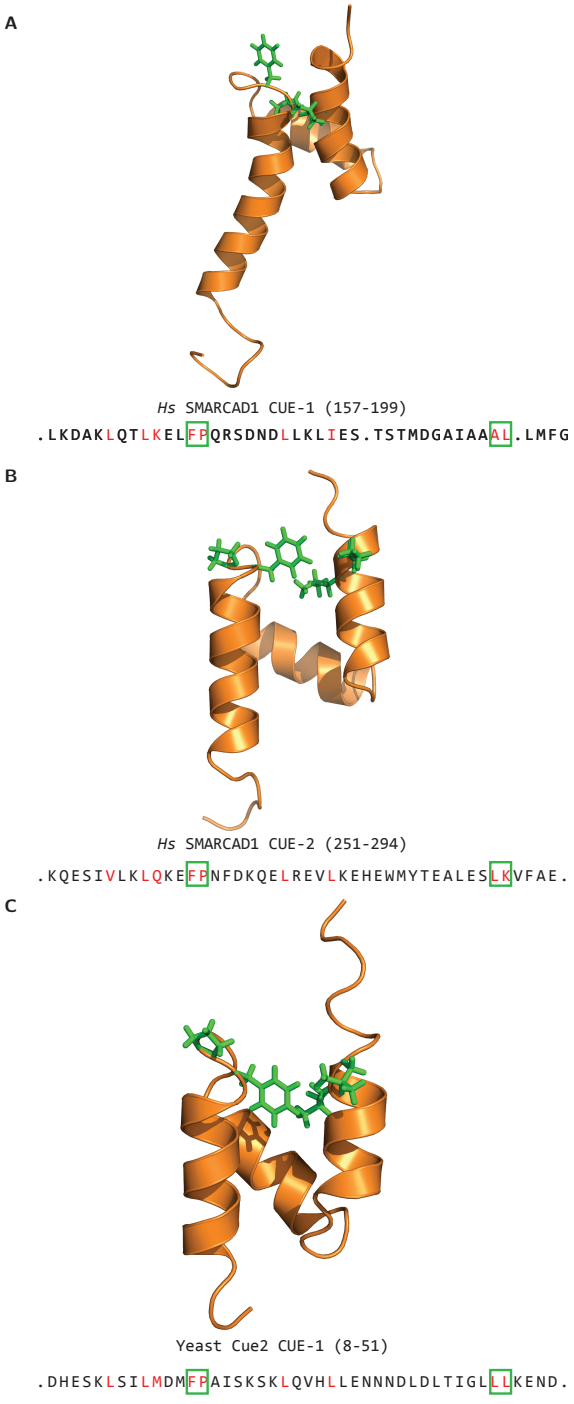


Figure 7.4: Conserved Residues bridge CUE Helices 1 and 3 - Two highly conserved regions (green boxes) are typically found at the C-terminal ends of helices 1 and 3. SMARCAD1 CUE-1 (**A**) is unusual among CUE domains in having an AL motif instead of LL (or similar with large, typically hydrophobic, side-chain residues). Only one other example, *Schizosaccharomyces pombe* protein Cue4 (CUE domain-containing protein 4), has an AL in this position (287). SMARCAD1 CUE-2 (**B**) and Cue2 CUE-1 (**C**) share the structure seen in other solved CUE domains. The presence of a short residue unable to interact with Phe breaks the bridging connection between helices, which collapses the centre of the core as Phe unusually orientates out of the centre.

that the endogenous, high affinity, binding target of CUE-1 is a ubiquitinated protein, poly-ubiquitin chain or some other protein.

Similarly, the protruding N-terminal helix in CUE-1 and a section of sequence upstream of that show very weak binding to SUMO-1 (sections 5.7.1.4 and 6.3.1.1). It may be the case that the CUE-1 domain / N-terminal SUMO-recognition site bind the same protein, which has been modified with ubiquitin and SUMO. One intriguing possibility is that the high-affinity recognition target for CUE-1 maybe be ubiquitin-SUMO heterologous chain, atypical and poorly understood linked chains of ubiquitin and SUMO (133) (figure 7.5).

7.3.2 CUE-2 Doesn't Bind Ubiquitin

One significant result from NMR titration studies of dCUE is the fact that CUE-2 is unable to bind ubiquitin (section 6.2). Structural analysis suggests that this is due to a positively charged lysine residue in the location of a typically highly conserved LL motif (figure 7.6). This leaves an open question regarding the function and binding partner of CUE-2.

7.3.3 KAP1

KAP1 has previously been identified as the major stoichiometric binding partner in the SMARCAD1 complex (108). Based on western blot results (section 6.5), dCUE was

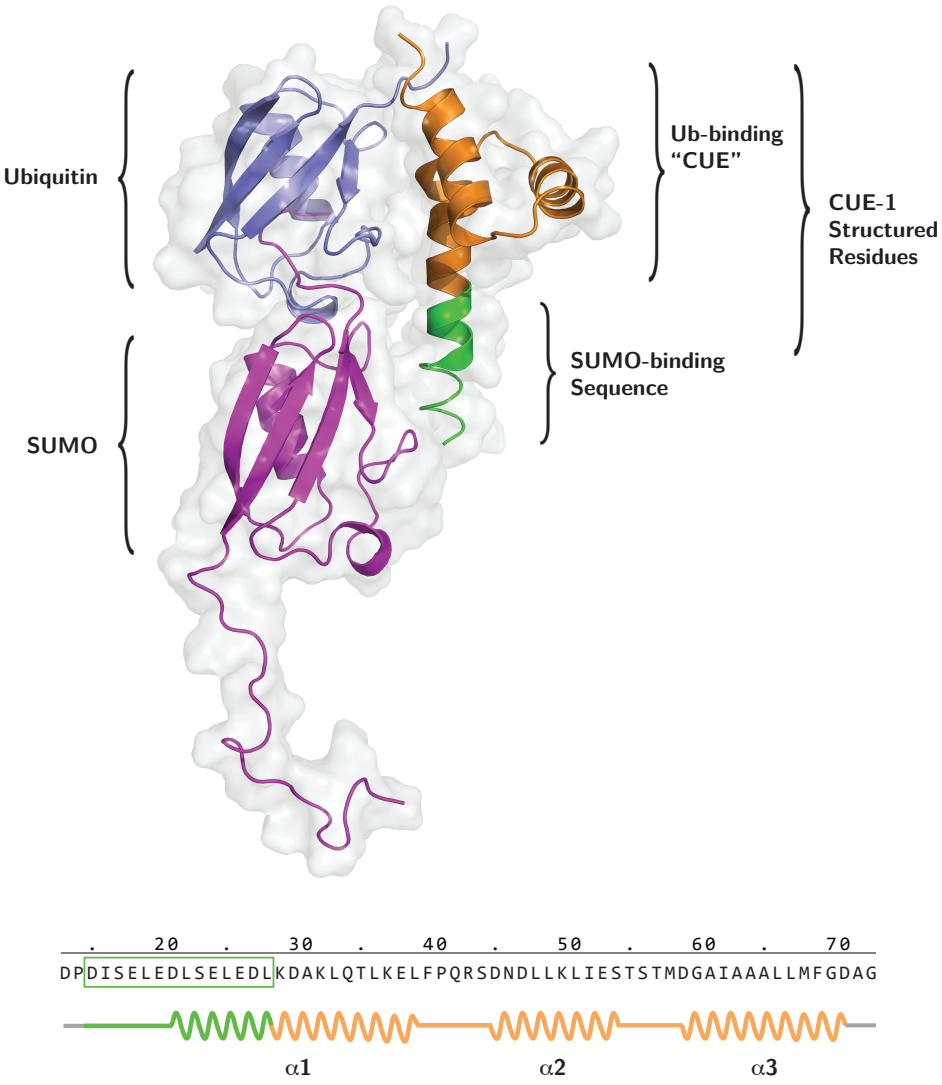


Figure 7.5: Theoretical Model of CUE-1 Binding to Atypical Ubiquitin-SUMO Heterologous Chain - CUE-1 domain and the SUMO binding site (N-terminal to CUE-1, green; section 6.3.1.1) bind ubiquitin (blue) and SUMO-1 (magenta) respectively. This model proposes a binding target for the CUE-1 as a ubiquitin-SUMO linked heterologous chain (133). The solved structure of CUE-1 corresponds to all structured residues around the first CUE domain in SMARCAD1. This structure has a protruding N-terminal helix, which may represent a separate, but adjacent, SUMO binding motif. The SUMO binding site is a repeat of sequence **(I)SELEDL**, which has hydrophobic residues every third residue, conceivably providing a hydrophobic face to an even longer, more extended, helix than seen in CUE-1. This protruding section of helix is more dynamic in solution and a fully extended helix may only be induced upon ligand (SUMO) binding and thus not seen in the solution structure. The **SELEDL** motif is only found in SMARCAD1 proteins (PROSITE (221)) although various single-helix motifs with a hydrophobic surface bind ubiquitin. It is possible that this may represent a comparable, single helix SUMO binding motif. Excluded from the dCUE construct is a potential upstream SUMO binding motif (figure 6.13) (281). It is plausible that this would be spatially able to interact with SUMO by extending the SUMO β-sheet of SUMO as seen with V/I-x-V/I-V/I SUMO binding motifs (280).

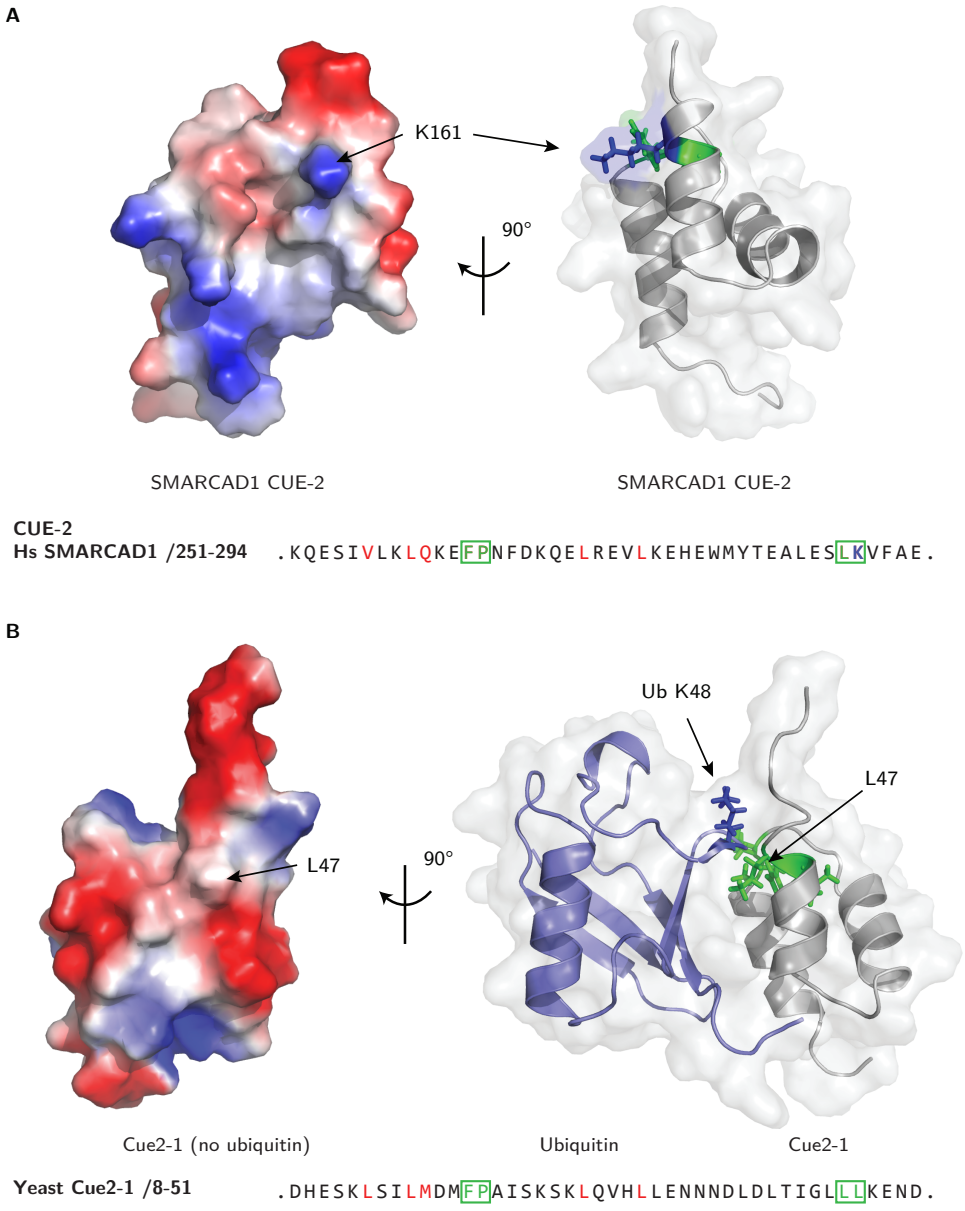


Figure 7.6: Positively Charged Lysine in Conserved Position Responsible for CUE-2 Inability to Bind Ubiquitin - CUE domains typically interact with ubiquitin through hydrophobic packing. Four highly conserved hydrophobic residues FP and LL line the C-terminal ends of helices 1 and 3 (green). **A** In the case of SMARCAD1 CUE-2 domain, a positively charged lysine residue (blue) has replaced the second leucine in the C-terminal LL motif. Electrostatics show this as blue surface mark. **B** Comparison with Yeast Cue2-1 domain bound to ubiquitin (ubiquitin cartoon in blue) (167). The LL motif from yeast Cue2-1 packs very closely to ubiquitin K48. Electrostatic surface potentials shown for just the Cue2-1 domain, without ubiquitin. Leucine 47 is neutrally charged and thus doesn't interfere with ubiquitin binding. SMARCAD1 CUE-2 is not able to bind ubiquitin due to positive charge repulsions that would occur between ubiquitin K48 and CUE-2 K161.

shown to be a strong binder of KAP1 from nuclear extract and, given the relative pull-down affinities of dCUE and full-length SMARCAD1 for KAP1, quite possibly the only part of SMARCAD1 that binds KAP1. SMARCAD1-KAP appears to represent a core complex with dCUE playing an important role in mediating the interaction between the two.

It remains unknown which part of dCUE is involved with the binding of KAP1 and whether any post-translational modifications mediate the interaction. KAP1 can be SUMOylated across a range of lysine residues in its PHD-Bromodomain (Lysines 554, 575, 676, 750, 779 and 804) (113) and has been found to be autoubiquitinated (288). Recruitment partners of KAP1, such as the histone methyltransferase SETDB1 (SET domain, bifurcated 1) (113), bind to SUMOylated bromodomains. SUMOylated KAP1 represents the repressive form of KAP1, mediating chromatin silencing (section 7.5) (289, 290, 291). As the dCUE domain has been shown to bind ubiquitin / SUMO, it suggests that KAP1-SMARCAD1 interaction may be through post-translational modification states.

KAP1 is recruited to numerous sites throughout the genome by binding to site-specific DNA binding KRAB-ZNF (Krüppel-associated box - Zinc Finger) domains where it is involved with transcription repression (113). This interaction utilises a protein-protein interface at its RBCC domain which could alternatively be the site of

dCUE-KAP1 interactions.

7.4 Linker

The linker region between the two CUE domains is long (*ca.* 50 residues) and shows little sign of structure with the exception of a dynamic helical turn in the middle. Due to the flexible nature of the linker, it was challenging to make assignments as chemical shift values were often poorly resolved and not significantly different from random coil values. A 10 residue stretch of the linker has remained intractable to backbone assignment (section 4.3.5). This includes a SSSS repeat, which was unassignable due to overlap. Likewise a RKRK motif could not be assigned. It is possible that the SSSS might be observed at lower sample pH.

The function of the linker remains unknown. None of the linker residues interacted with ligands during the NMR titration experiments. It is likely that it serves as a spacer between CUE domains but may also have a signalling role. The SSSS repeat is phosphorylated on DNA damage. This is thought to be by the action of protein kinases ATM (ataxia telangiectasia mutated) or ATR (ATM and Rad3-related respectively) (117, 292, 293).

7.5 Role of Double CUE Domain in SMARCAD1

Unlike other chromatin remodellers, SMARCAD1 doesn't show signs of being targeted to histone post-translational modifications. Instead it appears to respond to modified and unmodified states of other proteins. The double CUE domain has emerged as an important protein binding domain as it associates SMARCAD1 with one of its main binding parters, KAP1.

7.5.1 Silencing of Newly Synthesised Chromatin

The best understood function for SMARCAD1 is its role in restoring chromatin silencing at the sites of DNA replication. Figure 7.7 presents a model for SMARCAD1 recruitment to PCNA via non-dCUE interactions. SMARCAD1 itself may bind histone acetyltransferases (HDAC1/2) (108) and histone methyltransferases such as G9a/GLP (108, 110, 294) although the regions of SMARCAD1 involved are unknown. The double CUE domain of SMARCAD1 binds KAP1, which associates with histone modifying enzymes (acetyltransferases and methyltransferases) and heterochromatin protein 1 (HP1) to confer a heterochromatic, silenced state on newly produced chromatin. SMARCAD1 ATPase activity is required for histone deacetylation and likely involved with nucleosome reorganisation and histone accessibility (108).

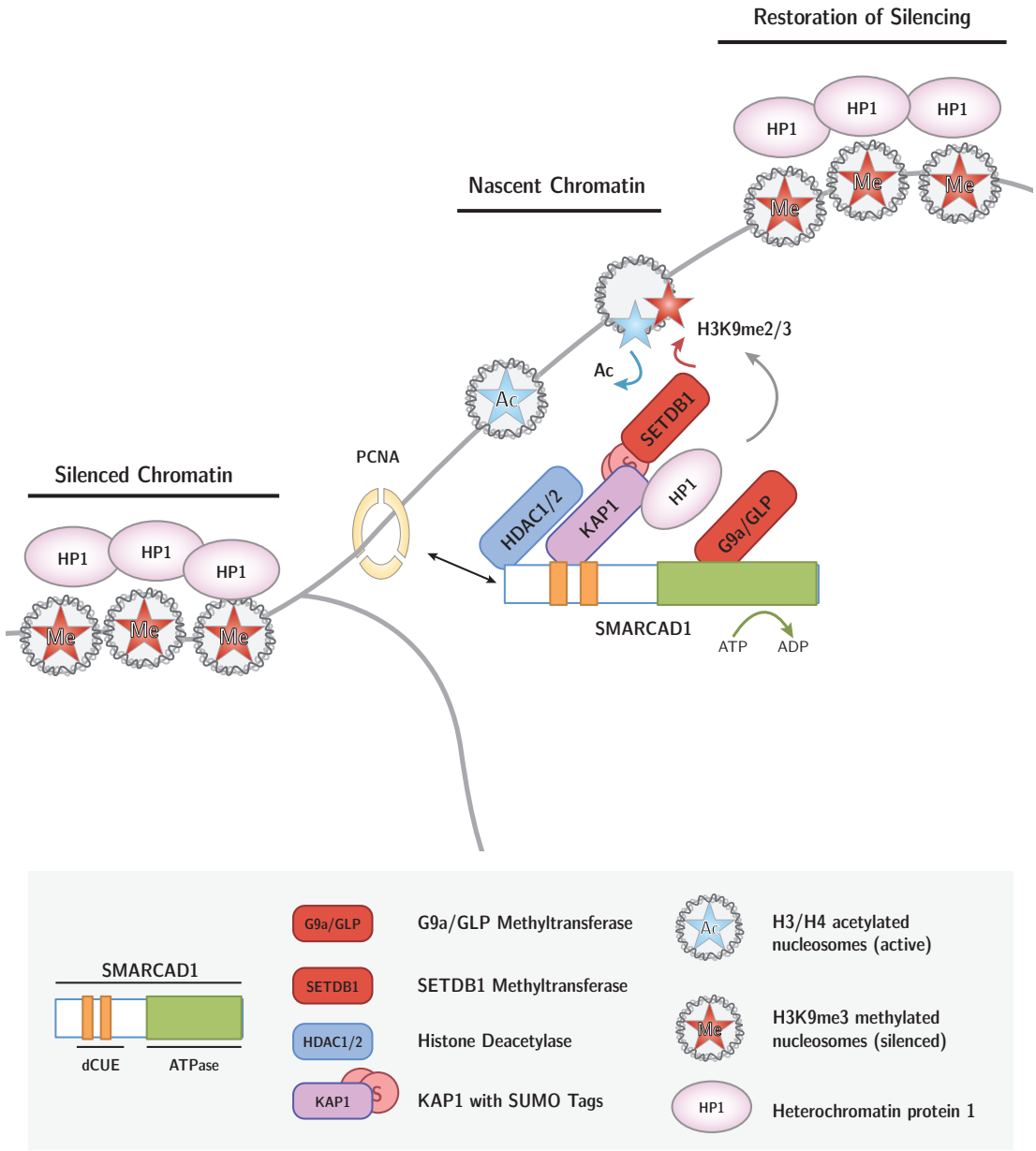


Figure 7.7: Model of SMARCAD1 Recruitment and Function During DNA Replication - Newly produced chromatin in 'active' acetylated state is silenced by the actions of the SMARCAD1 complex working at the site of PCNA near the DNA replication fork. SMARCAD1 is targeted to PCNA, binding directly through dCUE flanking regions. SMARCAD1 recruits KAP1 via a direct interaction with the dCUE domain. Associated factors bound to SMARCAD1 or KAP1 alter histone post-translational modifications and KAP1-associated HP1 is transfered to trimethylated H3K9. HP1 helps maintain heterochromatic packaging. Expanded from Mermoud *et al* (110).

7.5.2 DNA Repair

One little studied function of SMARCAD1 is in repair of DNA double-stranded breaks. SMARCAD1 and its yeast homologue FUN30 have both been shown to be directly involved with DNA damage response. Fun30 associates with double-stranded break ends and, through a mechanism involving its ATPase activity, promotes both Exo1- and Sgs1-dependent end resection (104).

Figure 7.8 propose a model of SMARCAD1 recruitment to double-stranded DNA breaks by phosphorylation of the linker region by ATM / ATR. SMARCAD1 may recruit a DNA double-stranded break end resection complex. Whether phosphorylated KAP1 and phosphorylated SMARCAD1 remain a bound complex is unclear.

7.6 Final Conclusions and Future Work

The work presented in this thesis has investigated the tandem CUE domain region of chromatin remodelling factor SMARCAD1. This region is characterised by two globular domains with a long flexible linker. The N-terminal CUE domain, CUE-1, has an unusual fold in comparison to other CUE domains. The core has collapsed slightly due to a substitution of a conserved leucine residue and the N-terminal α-helix is extended and protrudes from the conserved CUE helical core bundle, which may represent an adjacent SUMO binding site.

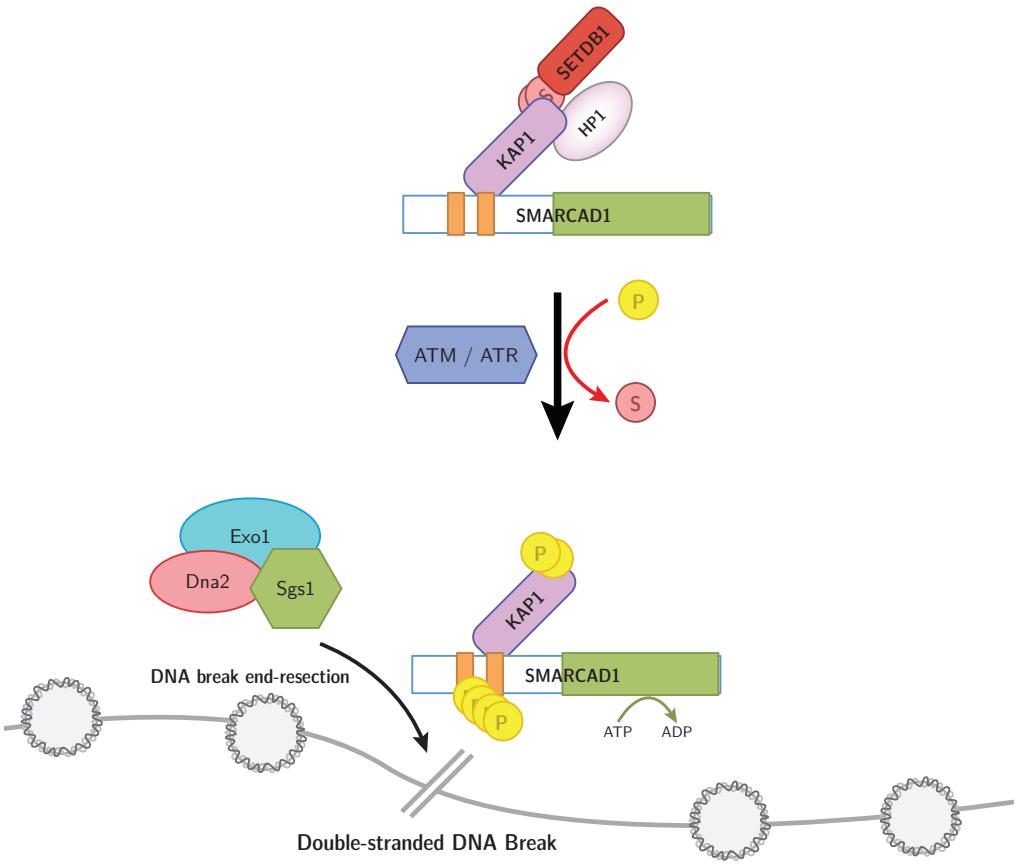


Figure 7.8: Model of SMARCAD1-KAP1 Response to DNA Damage - ATM / ATR kinases coordinate a response to DNA damage through phosphorylation of the dCUE in the linker region and KAP1 (section 1.3.2.1) (295, 296). SUMO modifications are replaced with phosphorylation. SMARCAD1 is recruited to sites of double-stranded DNA breaks and may recruit the helicase Sgs1 and two nucleases Dna2 and Exo1, which resect DNA double-strand break ends (104, 297).

CUE-1 has been demonstrated to bind ubiquitin with weak affinity. Similarly, the region immediately upstream of this binding site has weak affinity for ubiquitin-like SUMO protein. Although both affinities are too weak to be biologically meaningful, it suggests that the endogenous binding partner(s) may be modified proteins. It is possible that binding affinity is enhanced due to cooperativity with both binding sites occupied.

Interestingly, the C-terminal CUE domain, CUE-2, has a fold highly similar to other reported CUE domains but is unable to bind ubiquitin. This may be due to a substitution of a conserved leucine residue with a charged lysine that may inhibit ubiquitin binding.

The double CUE domain binds KAP1, a key binding partner of SMARCAD1 and scaffold protein for histone modification enzymes. SMARCAD1 directly interacts with PCNA but the double CUE domain is not involved in this interaction.

7.6.1 Structure

There remains some ambiguity over the exact conformation of the structure for CUE-1. The ambiguity lies in the turns between helices and has a consequent effect on helix orientation. It may be that additional NOE or RDC restraints could be determined from either further analysis of the existing data or collection of experiments that investigate

the side-chains of aromatic residues. These additional restraints may clarify the exact CUE-1 conformation.

7.6.2 KAP1 Binding

A pull-down experiment of recombinant full-length KAP1 vs probing for KAP1 pull-down from nuclear extract may help our understanding of whether the interaction is mediated by a modification such as SUMO or ubiquitin. If pull-down experiments show binding with recombinant, unmodified KAP1, NMR experiments with unlabelled, recombinant KAP1 and KAP1 fragments may elucidate the residues or at least the region(s) of dCUE (CUE-1, linker and / or CUE-2) that interact with KAP1.

7.6.3 SUMO Binding

The observation of a consensus SUMO binding motif N-terminal of both the start of the dCUE construct sequence and the region of dCUE that shows weak SUMO binding suggests that this may be a genuine, higher affinity, SUMO binding site. Design and use of a new protein construct for CUE-1 that includes additional N-terminal SMARCAD1 sequence would allow for the SUMO binding properties to be reanalysed with NMR.

7.6.4 Additional dCUE Binding

The dCUE pull-down experiment using nuclear extract identifies a high molecular weight protein which reacts with anti-ubiquitin antibody. This suggests that dCUE

may bind to a protein that is ubiquitinated (either through the ubiquitin modification or otherwise). A repeat of this experiment could be coupled to mass spectrometry of trypsin-digest fragments to identify the protein involved. Mass spectrometry of all dCUE binding proteins could also be carried out and compared with the list of full-length SMARCAD1 binding proteins produced by Rowbotham *et al* (108). In addition to understanding the proteins that bind dCUE, it would also give an indication of proteins that bind elsewhere on SMARCAD1.

7.6.5 dCUE Phosphorylation and KAP1 Binding

Very little is known about the linker region of dCUE and the significance of the phosphorylation states of the tetra-serine motif. Both SMARCAD1 and KAP1 appear to be targets of ATM/ATR phosphorylation in response to DNA damage. In the case of KAP1, the SUMO tag is replaced by phosphorylation. Investigating phosphorylation states and dCUE-KAP1 binding under different cell conditions (such as hypoxia; DNA damage) may help explain the nature of the SMARCAD1-KAP1 interaction. It is conceivable that dCUE and KAP1 interact via a SUMO modification on KAP1 which is lost to phosphorylation or that the two proteins interact via another interface and remain in complex.

Appendices

Appendix A

NMR Data

A.1 Chemical Shift Tables

Table A.1 contains the chemical shift values for all assigned resonances.

Table A.1: Chemical Shifts.

Residue	HN	N	Hα	Hβ	Hγ	Hδ	Hϵ	C'	Cα	Cβ	Cγ	Cδ	Cϵ
1 Met	-	-	-	-	-	-	-	-	-	-	-	-	-
2 Gly	-	-	-	-	-	-	-	-	-	-	-	-	-
3 Ser	-	-	-	-	-	-	-	-	-	-	-	-	-
4 Ser	-	-	-	-	-	-	-	-	-	-	-	-	-
5 His	-	-	-	-	-	-	-	-	-	-	-	-	-
6 His	-	-	-	-	-	-	-	-	-	-	-	-	-
7 His	-	-	-	-	-	-	-	-	-	-	-	-	-
8 His	-	-	-	-	-	-	-	-	-	-	-	-	-
9 His	-	-	-	-	-	-	-	-	-	-	-	-	-
10 His	-	-	-	-	-	-	-	-	-	-	-	-	-
11 Ser	-	-	4.35	3.80,3.80	-	-	-	-	58.72	63.89	-	-	-
12 Gln	8.48	121.08	4.32	1.93,2.09	2.30,2.26	-	-	175.33	55.86	29.65	33.93	-	-
13 Asp	8.3	122.54	4.81	2.47,2.73	-	-	-	174.45	52.69	40.92	-	-	-
14 Pro	-	-	4.35	1.88,2.20	-	-	-	176.7	63.54	32.27	27.32	-	-
15 Asp	8.48	119.37	4.56	2.67,2.57	-	-	-	176.51	54.66	41.33	-	-	-
16 Ile	7.93	120.07	4.16	1.9	0.88,1.41,1.16	0.83	-	176.36	61.49	38.72	17.51,27.22	13.1	-
17 Ser	8.39	118.88	4.38	3.90,3.84	-	-	-	174.77	59.11	63.85	-	-	-
18 Glu	8.38	122.03	4.29	1.91,2.05	2.24	-	-	176.43	56.77	30.23	36.33	-	-
19 Leu	8.09	121.49	4.28	1.63,1.58	1.61	0.82,0.88	-	177.47	55.49	42.44	27.44	23.59,25.03	-
20 Glu	8.32	120.7	-	1.94	2.64	-	-	176.05	56.96	30.55	-	-	-
21 Asp	8.37	120.84	4.59	2.58,2.75	-	-	-	176.46	54.28	41.29	-	-	-
22 Leu	8.36	123.34	4.33	1.57,1.63	-	0.89,0.83	-	178.13	55.7	42.07	27.2	25.19,23.45	-
23 Ser	8.4	116.16	4.29	3.88,3.89	-	-	-	175.25	60	63.4	-	-	-
24 Glu	8.35	121.62	4.27	1.96,2.09	-	-	-	177.27	57.4	29.98	-	-	-
25 Leu	8.01	120.7	4.21	1.71,1.73	1.58	-	-	178.55	56.5	42.23	-	-	-
26 Glu	8.3	120.4	4.12	2.01	-	-	-	177.18	58.01	30.28	-	-	-
27 Asp	8.3	120.37	4.51	2.69,2.69	-	-	-	177.66	55.64	40.9	-	-	-
28 Leu	8.13	121.89	4.25	1.67,1.73	-	0.94,0.87	-	178.37	56.63	42.16	27.1	24.99,23.82	-
29 Lys	8.07	120.47	3.87	1.87,1.86	1.37,1.37	1.68,1.68	2.91,2.91	177.83	60	32.42	24.84	29.72	41.76
30 Asp	8.21	117.99	4.4	2.66	-	-	-	178.06	57.02	41.18	-	-	-
31 Ala	8.01	122.4	4.14	1.48	-	-	-	180.85	55.12	18.42	-	-	-
32 Lys	8.1	118.61	4.03	1.64,2.00	-	-	2.76,2.76	178.25	60.01	33.32	-	-	42.21
33 Leu	-	-	3.9	1.39,2.04	1.48	0.91,0.90	-	178.06	58.46	41.18	27.3	25.90,24.30	-
34 Gln	8.27	116.32	3.88	2.12	2.43,2.51	-	-	178.81	58.95	27.81	33.5	-	-
35 Thr	7.8	115.61	3.95	4.3	1.22	-	-	176.59	66.92	68.77	21.5	-	-
36 Leu	8.11	120.25	4.16	1.73,1.95	1.87	0.91	-	178.45	58.32	42.59	27.75	26.29	-
37 Lys	8.88	119.32	3.77	1.68,1.83	-	-	-	178.5	59.01	32.13	24.99	29.28	-
38 Glu	7.36	115.61	3.92	2.02,2.13	2.44,2.21	-	-	178.5	58.57	29.7	36.46	-	-
39 Leu	7.18	117.28	3.9	0.74,1.66	-	0.84,0.70	-	176.7	56.56	44.17	-	22.70,25.71	-
40 Phe	8.07	115.7	4.9	2.78	-	-	-	173.09	55.58	39.08	-	-	-
41 Pro	-	-	4.57	1.92,2.33	-	3.22	-	178.09	64.65	32.41	-	-	-
42 Gln	8.65	114.6	4.28	1.99,2.20	2.37,2.28	-	-	176.28	56.25	28.25	34.35	-	-
43 Arg	7.49	117.63	4.43	1.75,1.53	0.71,1.27	2.27,2.61	-	175.54	53	30.12	25.77	41.69	-
44 Ser	8.95	117.46	4.31	3.96,3.96	-	-	-	174.82	57.57	64.96	-	-	-
45 Asp	8.89	122.53	4.16	2.56,2.57	-	-	-	178.13	58.11	39.51	-	-	-

Continued on next page

Table A.1 – *Continued from previous page*

Residue	HN	N	Hα	Hβ	Hγ	Hδ	Hϵ	C'	Cα	Cβ	Cγ	Cδ	Cϵ
46 Asn	8.4	114.8	4.34	2.67,2.68	-	-	-	177.63	56.77	38.42	-	-	-
47 Asp	7.72	119.34	4.36	2.51,2.87	-	-	-	178.54	57.44	40.9	-	-	-
48 Leu	7.73	117.89	4.08	1.91	-	0.85,0.78	-	179.13	57.64	41.45	-	21.96,25.85	-
49 Leu	8.98	119.83	3.97	1.82,1.54	-	-	-	179.27	58.3	41.97	-	-	-
50 Lys	7.55	117.2	4.05	1.92,1.92	-	-	-	179.81	59.39	32.18	-	-	-
51 Leu	7.28	118.63	3.99	2.04,1.36	1.8	0.68,0.74	-	179.41	57.61	41.59	26.44	23.17,25.90	-
52 Ile	7.86	120.11	3.54	1.99	0.86,1.70,0.87	0.78	-	178.09	64.68	38.33	30.34,17.49	13.71	-
53 Glu	8.29	116.1	4.13	2.06,2.08	2.43,2.25	-	-	177.51	58.4	30.03	36.77	-	-
54 Ser	7.54	110.52	4.53	3.91,3.91	-	-	-	173.58	58.91	64.54	-	-	-
55 Thr	7.4	111.19	4.72	4.21	1.08	-	-	173.95	60.54	72.57	21.71	-	-
56 Ser	-	-	4.63	3.92,3.92	-	-	-	175.1	59.52	64.83	-	-	-
57 Thr	7.62	108.77	4.82	4.63	1.23	-	-	175.15	59.09	73.06	21.86	-	-
58 Met	-	-	4.36	2.03,2.26	2.48,2.77	-	2.15	177.67	57.88	33.07	32.09	-	18.05
59 Asp	8.64	117.27	4.35	2.53,2.59	-	-	-	179.43	57.56	40.24	-	-	-
60 Gly	8.4	108.84	3.84,3.94	-	-	-	-	176.42	46.66	-	-	-	-
61 Ala	8.15	127.6	3.98	1.48	-	-	-	178.86	56.19	17.76	-	-	-
62 Ile	8.33	117.4	3.53	1.88	0.98,1.91,0.86	0.86	-	177.43	66.23	38.65	30.51,17.12	13.88	-
63 Ala	8.02	119.82	4.06	1.46	-	-	-	180.7	55.29	18.13	-	-	-
64 Ala	8.06	119.22	4	1.37	-	-	-	179.8	54.71	18.65	-	-	-
65 Ala	8.32	122.65	3.34	0.97	-	-	-	178.78	54.87	18.52	-	-	-
66 Leu	8.38	118.06	4.09	1.54,1.90	1.95	0.96,1.04	-	180.25	57.93	42.05	27.62	23.42,25.62	-
67 Leu	7.34	117.72	4.08	1.54,1.66	1.61	0.82,0.83	-	178.59	57.31	42.35	27.03	24.03,24.51	-
68 Met	7.75	116.73	3.99	1.15,1.40	2.07,1.90	-	1.85	177.72	57.8	33.19	31.63	-	16.97
69 Phe	8.44	115.04	4.93	2.75,3.28	-	-	-	177.08	55.87	39.42	-	-	-
70 Gly	7.75	108.46	4.15,4.06	-	-	-	-	173.15	45.61	-	-	-	-
71 Asp	8.24	118.38	4.63	2.66,2.66	-	-	-	176.54	54.23	41.92	-	-	-
72 Ala	8.44	124.25	4.29	1.38	-	-	-	178.5	52.91	19.31	-	-	-
73 Gly	8.53	107.34	3.91,3.91	-	-	-	-	174.83	45.59	-	-	-	-
74 Gly	8.25	107.91	3.97,3.97	-	-	-	-	174.4	45.4	-	-	-	-
75 Gly	8.22	108.3	4.03,4.03	-	-	-	-	171.83	44.97	-	-	-	-
87 Pro	-	-	4.56	2.24,2.05	-	-	-	175.73	62.53	34.39	24.7	-	-
88 Tyr	8.52	121.99	4.53	2.94,3.02	-	7.11	6.82	175.77	58.07	38.63	-	133.21	118.25
89 Glu	8.39	122.65	4.25	1.83,1.97	2.16,2.16	-	-	-	56.06	31.03	36.3	-	-
90 Glu	8.32	121.97	-	-	-	-	-	-	57.56	33.19	-	-	-
92 Glu	-	-	4.14	1.81,1.81	2.05,1.96	-	-	-	56.93	30.51	36	-	-
93 Phe	8.29	119.84	4.57	2.97,3.12	-	-	-	175.46	57.91	39.58	-	-	-
94 Asn	8.21	119.99	4.64	2.76,2.68	-	-	-	174.65	53.16	39.65	-	-	-
95 Asp	-	-	4.51	2.69,2.60	-	-	-	176.09	54.58	41.33	-	-	-
96 Asp	8.28	120.08	4.5	2.68,2.68	-	-	-	176.92	58.53	42.88	-	-	-
97 Gln	8.43	119.48	4.22	1.98,2.18	2.35,2.35	-	-	176.56	56.57	28.87	33.97	-	-
98 Ser	8.3	115.48	4.29	3.87,3.87	-	-	-	174.98	59.64	63.81	-	-	-
99 Ile	7.92	121.14	4.1	1.85	1.15,1.43,0.88	0.83	-	176.41	61.63	38.42	27.67,17.47	13.09	-
100 Lys	8.21	123.98	4.23	-	-	-	-	176.64	56.68	33.02	-	-	-
101 Lys	8.27	121.81	4.31	-	-	-	-	176.75	56.65	33.31	-	-	-
102 Thr	8.19	114.94	4.27	4.16	1.17	-	-	174.41	62.28	70.01	21.61	-	-
103 Arg	8.34	122.96	-	-	-	-	-	177	56.27	31.07	-	-	-

Continued on next page

Table A.1 – *Continued from previous page*

Residue	HN	N	Hα	Hβ	Hγ	Hδ	Hϵ	C'	Cα	Cβ	Cγ	Cδ	Cϵ
104 Leu	8.34	122.99	4.3	1.57,1.48	1.61	0.83,0.89	-	177.03	55.26	42.34	27.44	23.39,25.05	-
105 Asp	8.32	120.11	4.51	2.59,2.59	-	-	-	175.98	54.25	41.35	-	-	-
106 His	8.32	118.45	4.61	3.08,3.21	-	-	-	175.54	56.12	30.04	-	-	-
107 Gly	8.47	109.22	3.92,3.92	-	-	-	-	174.19	45.56	-	-	-	-
108 Glu	8.45	120.14	4.27	1.91,2.05	2.24,2.24	-	-	176.72	56.78	30.55	36.27	-	-
109 Glu	8.57	121.2	4.27	1.92,2.04	2.25,2.25	-	-	176.56	57.56	30.36	-	-	-
110 Ser	-	-	4.42	3.83	-	-	-	174.34	58.36	63.97	-	-	-
111 Asn	8.53	120.71	4.72	2.74,2.79	-	-	-	175.38	53.53	39.16	-	-	-
112 Glu	8.45	120.67	4.26	2.03,1.91	-	-	-	-	56.97	30.26	36.35	-	-
113 Ser	8.37	116.1	4.39	3.88	-	-	-	174.78	58.53	63.94	-	-	-
114 Ala	8.37	125.49	4.26	1.39	-	-	-	178.22	53.21	19.15	-	-	-
115 Glu	8.34	118.5	4.18	1.92,2.00	2.22,2.22	-	-	177.03	57.22	29.97	-	-	-
116 Ser	8.21	115.38	4.38	3.91,3.84	-	-	-	175.13	59.06	63.86	-	-	-
117 Ser	-	-	-	-	-	-	-	-	59.22	63.79	-	-	-
118 Asn	8.35	119.63	4.6	2.72,2.72	-	-	-	-	53.77	38.69	-	-	-
119 Asn	8.17	118.43	-	-	-	-	-	-	54.33	38.83	-	-	-
120 Trp	-	128.82	-	-	-	-	10.13	-	-	-	-	-	[Nϵ1: 128.89]
121 Glu	-	-	3.74	1.98,1.98	-	-	-	179.27	59.82	29.52	-	-	-
122 Lys	7.92	120.17	3.99	1.77,1.77	-	-	-	179.57	58.84	32.29	-	-	-
123 Gln	7.86	118.25	3.55	-	-	-	-	178.07	58.78	28	-	-	-
124 Glu	8.31	118.77	3.92	1.75,1.49	-	-	-	178.73	58.63	29.09	-	-	-
125 Ser	7.83	113.47	4.14	3.96,3.96	-	-	-	-	61.95	63.02	-	-	-
126 Ile	7.67	123.26	3.81	2.33	1.28,1.22,1.93	1.04	-	177.76	65	37.83	19.84,28.54	13.37	-
127 Val	8.17	120.95	3.37	2.33	1.11,0.94	-	-	177.85	67.3	31.63	21.91,23.84	-	-
128 Leu	8.19	117.31	4.01	1.55,1.82	1.8	0.89,0.92	-	179.66	58.06	41.68	27.15	23.39,25.04	-
129 Lys	7.61	120.08	4	1.90,1.90	-	-	-	179.49	60.18	32.91	-	-	-
130 Leu	8.35	120.32	4.49	1.78,1.92	1.78	0.77,0.63	-	178.79	57.62	42.4	26.76	24.33,26.15	-
131 Gln	8.87	119.5	3.87	2.24,1.87	2.47,2.44	-	-	177.16	58.26	29.23	35.21	-	-
132 Lys	7.5	115.53	3.97	1.94,1.94	-	-	-	178.71	58.9	32.28	-	-	-
133 Glu	7.13	115.54	3.98	2.00,2.00	-	-	-	175.9	57.27	30.36	-	-	-
134 Phe	7.72	116.49	4.84	3.18,2.66	-	-	-	172.99	56.3	40.05	-	-	-
135 Pro	-	-	4.66	2.31,1.94	-	-	-	177.67	64.68	32.28	-	-	-
136 Asn	8.82	113.86	4.63	2.70,2.69	-	-	-	174.88	53.7	37.67	-	-	-
137 Phe	7.78	120.29	4.68	2.91,3.07	-	-	-	175.13	57.09	41.1	-	-	-
138 Asp	9.16	121.33	4.59	2.68,2.89	-	-	-	176.93	53.54	42.91	-	-	-
139 Lys	8.6	120.9	3.73	1.85,1.86	-	-	-	178.33	61.04	32.75	-	-	-
140 Gln	8.66	118.03	4.08	2.23,2.14	-	-	-	178.67	59.29	27.95	-	-	-
141 Glu	7.95	119.64	4.04	2.34,2.00	-	-	-	179.6	59.28	29.98	-	-	-
142 Leu	7.95	117.56	3.81	1.98,1.18	1.82	0.85,0.49	-	178.43	57.57	42.12	26.34	25.82,22.83	-
143 Arg	8.24	118.71	3.66	2.00,2.00	1.78,1.62	3.46,3.22	-	177.57	60.43	30.21	29.62	43.2	-
144 Glu	7.62	117.31	3.92	2.10,2.10	-	-	-	178.98	59.42	28.98	-	-	-
145 Val	7.72	119.69	3.71	1.92	0.90,0.67	-	-	179.27	66.14	31.81	21.13,22.36	-	-
146 Leu	7.98	120.98	3.25	1.06,0.06	1.07	0.59,0.76	-	179.1	59.08	40.64	28.51	26.29,23.97	-
147 Lys	8.09	121.1	4.29	1.74,1.70	1.55	1.6	2.98,2.98	180.28	59.97	32.32	-	-	-
148 Glu	7.89	119.63	3.92	1.85,2.00	-	-	-	176.86	58.44	29.34	-	-	-
149 His	7.22	114.6	4.51	2.43,3.61	-	-	-	173.73	56.06	28.33	-	-	-

Continued on next page

Table A.1 – *Continued from previous page*

Residue	HN	N	Hα	Hβ	Hγ	Hδ	Hϵ	C'	Cα	Cβ	Cγ	Cδ	Cϵ
150 Glu	7.87	115.51	3.87	2.09,2.32	2.22,2.17	-	-	174.64	57.98	27.2	36.72	-	-
151 Trp	8.55	108.67	3.2	3.31,3.58	-	-	10.16	174.21	59.56	24.88	-	-	[Nϵ1: 128.37]
152 Met	7.52	118.16	4.68	2.21,2.37	2.79,2.63	-	-	175.39	54.72	31.66	32.18	-	-
153 Tyr	-	-	3.69	3.04,3.15	-	6.96	6.67	175.93	63.18	39.38	-	132.66	118.36
154 Thr	8.69	111.23	3.58	4.16	1.29	-	-	176.23	66.56	68.05	22.42	-	-
155 Glu	7.49	120.12	3.9	2.03,2.03	-	-	-	178.84	59.25	29.79	-	-	-
156 Ala	7.76	121.43	2.88	0.96	-	-	-	178.42	54.7	18.59	-	-	-
157 Leu	8.24	118.61	3.56	1.60,1.02	1.45	0.87,0.89	-	177.57	58.95	41.89	27.13	25.30,25.19	-
158 Glu	-	-	3.8	1.93,1.93	-	-	-	-	59.13	29.4	-	-	-
159 Ser	7.36	114.05	4.12	3.70,3.58	-	-	-	175.65	61.27	62.82	-	-	-
160 Leu	7.28	121.14	3.94	1.07,0.27	1.17	0.11,-0.04	-	178.05	56.02	41.73	26.28	22.40,25.87	-
161 Lys	7.63	118.45	4.15	1.88,1.88	1.51,1.69	1.69,1.69	2.88,2.88	177.6	58.59	32.95	25.32	29.57	42.07
162 Val	7.41	116.17	3.97	1.99	0.84,0.75	-	-	176.48	63.35	32.15	20.63,20.92	-	-
163 Phe	7.89	121.02	4.58	3.24,3.05	-	-	-	175.65	58.27	39.46	-	-	-
164 Ala	8.04	124.16	4.34	1.44	-	-	-	177.11	52.58	19.89	-	-	-
165 Glu	8.2	119.89	4.28	1.91,2.09	2.30,2.30	-	-	175.4	56.5	30.81	-	-	-
166 Asp	7.98	126.19	4.35	2.54,2.64	-	-	-	180.94	56.02	42.4	-	-	-

A.2 Dihedral Angles Used in Structure Calculations

Table A.2 contains the dihedral angle values used in structure calculations. Prediction error of ϕ and ψ is indicated by the standard deviation, σ, although all calculations used a minimum error of at least 20 °. Dihedral angles were predicted based on chemical shift values (section 5.3.2).

Table A.2: Dihedral Angles

Residue	ϕ	ψ	σ ϕ	σ ψ
12	-98.658	-12.504	12.729	26.346
13	-87.878	131.342	25.487	25.478
14	-64.287	-22.692	9.904	12.91
15	-81.739	-16.496	24.237	20.485
16	-92.922	-14.14	20.719	21.96
17	-87.808	139.995	79.741	26.269
18	-78.808	-22.78	19.24	17.324
19	-87.009	138.527	62.236	20.584
20	-94.949	131.274	64.761	23.426
21	-116.939	156.605	69.897	52.911
22	-68.084	-28.455	18.098	19.81
23	-62.011	-32.116	6.031	11.034
24	-65.772	-42.358	6.192	7.072
25	-62.765	-37.785	4.914	9.717
26	-61.596	-40.867	5.394	11.108
27	-65.974	-36.187	4.481	9.985
28	-67.138	-38.555	18.357	11.642
29	-62.67	-41.193	7.52	7.921
30	-61.541	-42.608	3.921	4.796
31	-65.022	-39.382	4.329	1.745
32	-63.542	-43.455	3.938	3.05
33	-64.421	-43.133	3.943	4.41
34	-61.584	-39.303	4.014	3.93
35	-64.937	-44.791	5.262	4.536
36	-63.004	-40.974	4.658	3.231
37	-62.889	-43.062	5.361	7.164
38	-69.247	-37.113	11.033	6.538

Continued on next page

Table A.2 – *Continued from previous page*

Residue	ϕ	ψ	σ ϕ	σ ψ
39	-84.79	-19.403	12.481	24.058
40	-118.6	90.057	21.926	29.868
41	-62.473	-22.523	9.089	7.202
42	-86.932	-7.988	12.725	24.159
43	-106.205	140.933	34.787	20.558
44	-94.82	159.538	32.89	12.169
45	-60.077	-38.316	2.997	6.162
46	-60.814	-42.065	6.325	5.855
47	-65.248	-45.005	3.79	5.585
48	-64.908	-39.096	5.639	5.75
49	-64.65	-43.324	5.022	4.084
50	-61.95	-40.425	4.073	5.63
51	-65.36	-41.434	5.43	8.394
52	-66.917	-41.992	9.406	5.865
53	-69.816	-29.094	13.6	11.682
54	-102.515	-2.058	12.871	21.21
55	-118.74	150.132	27.968	18.201
56	-87.663	-20.929	25.462	20.658
57	-132.408	161.413	25.069	10.026
58	-64.855	-37.086	7.661	9.284
59	-63.348	-42.524	5.911	4.207
60	-64.318	-39.052	2.471	3.047
61	-63.474	-42.471	2.797	2.937
62	-63.772	-43.826	2.869	7.424
63	-58.401	-41.526	4.138	4.804
64	-65.991	-40.368	3.773	7.025
65	-63.08	-42.615	5.282	4.422
66	-66.11	-38.945	4.81	8.197
67	-64.94	-32.29	6.618	10.823
68	-73.302	-36.266	12.915	10.474
69	-99.632	-3.601	12.015	13.511
70	94.841	-3.516	10.157	11.134
71	-90.162	134.25	17.525	25.185
72	-82.265	-17.554	24.973	21.936
73	77.205	18.552	14.991	19.656
74	71.815	18.922	16.646	10.45
75	-134.437	173.75	70.469	51.823
87	-73.974	152.253	6.078	9.696
88	-102.055	130.525	18.301	21.571
89	-86.656	122.618	61.119	25.683
90	-128.369	145.08	29.841	16.931

Continued on next page

Table A.2 – *Continued from previous page*

Residue	ϕ	ψ	σ ϕ	σ ψ
91	-111.809	148.283	25.366	21.39
92	-78.138	-15.378	17.805	17.613
93	-110.945	132.006	83.085	28.734
94	-100.935	128.35	65.42	42.731
95	-79.785	-11.471	24.117	20.194
96	-61.099	-38.868	16.575	13.061
97	-66.945	-35.902	7.733	17.647
98	-67.632	-32.791	10.175	14.146
99	-99.466	-7.011	16.264	17.867
100	-78.057	141.649	71.884	39.34
101	-75.945	118.855	70.433	46.471
102	-96.77	126.35	67.1	26.198
103	-79.8	133.132	74.88	39.256
104	-88.418	124.11	78.577	42.044
105	-70.81	-22.472	21.193	22.933
106	-98.102	-6.249	15.493	20.722
107	86.792	12.37	15.74	22.025
108	-81.047	136.728	55.574	30.151
109	-93.791	129.478	71.706	21.152
110	-88.159	-14.746	26.336	28.615
111	-99.336	-6.703	18.512	17.274
112	-81.109	133.646	56.748	31.94
113	-66.808	-27.924	21.16	23.017
114	-65.012	-38.129	12.244	4.975
115	-65.77	-37.735	8.002	10.194
116	-70.117	-35.141	11.287	10.835
117	-73.661	-28.505	16.678	19.39
118	-68.885	-31.124	13.611	16.027
119	-71.051	-29.023	18.81	13.954
120	-63.775	-42.402	5.694	7.399
121	-58.1	-39.347	5.948	11.761
122	-63.78	-42.523	8.121	6.36
123	-63.894	-41.877	5.335	7.272
124	-64.855	-42.954	4.974	7.71
125	-65.321	-37.411	5.994	10.364
126	-64.769	-45.858	3.083	4.396
127	-59.522	-43.117	6.391	8.403
128	-58.995	-40.567	4.203	3.949
129	-61.977	-43.775	4.188	4.34
130	-63.661	-39.94	4.236	7.477
131	-66.033	-41.93	4.565	5.168

Continued on next page

Table A.2 – *Continued from previous page*

Residue	ϕ	ψ	σ ϕ	σ ψ
132	-65.29	-33.976	6.447	7.03
133	-85.808	-10.352	12.22	21.355
134	-114.498	86.706	19.927	24.317
135	-64.291	-19.38	9.093	4.523
136	-84.508	-4.964	19.579	14.828
137	-117.25	141.917	62.834	27.253
138	-85.176	164.212	26.734	6.854
139	-59.997	-39.544	3.214	4.943
140	-62.632	-42.721	5.116	6.947
141	-67.533	-40.394	3.062	5.898
142	-66.932	-33.698	5.378	8.972
143	-60.947	-46.017	9.308	6.303
144	-61.536	-43.783	5.972	5.893
145	-63.934	-45.901	3.74	3.954
146	-61.675	-39.057	5.697	7.002
147	-62.843	-38.509	3.664	6.209
148	-65.556	-29.29	5.04	7.635
149	-94.78	9.982	10.181	11.225
150	-95.184	14.601	12.523	14.959
151	60.326	30.736	6.228	11.615
152	-90.905	117.333	21.766	46.608
153	-52.872	-51.947	4.964	9.21
154	-61.551	-43.11	5.968	4.293
155	-61.447	-45.266	5.062	7.671
156	-65.497	-35.586	5.377	9.132
157	-60.254	-45.115	9.334	7.779
158	-59.328	-39.729	6.226	7.275
159	-63.536	-31.7	6.04	15.027
160	-63.353	-39.724	5.403	9.348
161	-66.662	-28.793	7.649	11.525
162	-78.243	-29.779	21.537	19.934
163	-100.981	-4.274	15.13	22.587
164	-70.662	145.73	67.876	38.342
165	-109.779	116.947	40.811	36.772

A.3 Measured Residual Dipolar Coupling Values

Table A.3 shows the residual dipolar coupling (RDC) values used in the structure calculations. RDC values could not be measured for unassigned residues and some assigned residues. The majority of RDC values are for backbone NH resonance. Figure 5.15 shows a plot of this data, indicating more clearly how RDC values change along the sequence. RDCs for residues 23-66 (30-66 when restraints prior to residue 30 were excluded) and 123-162 were used in structure calculations for the first and second CUE domains respectively.

Table A.3: Residual Dipolar Coupling Values

Residue	Residual Dipolar Coupling (Hz)
12	-0.122
15	-1.395
16	-0.425
18	-0.303
19	-0.485
20	-0.728
21	-1.395
22	-0.364
23	3.457
24	6.005
25	3.942
28	6.368
29	7.763
30	12.069
31	14.677
32	12.858
34	15.466
35	14.799
36	11.341
38	15.224

Continued on next page

Table A.3 – *Continued from previous page*

Residue	Residual Dipolar Coupling (Hz)
39	14.617
40	11.706
42	-12.251
43	3.457
44	0.909
45	-12.008
46	-3.578
48	-11.463
49	-10.554
50	4.548
51	-1.759
55	-6.975
59	6.672
60	7.642
61	8.552
62	2.486
66	4.306
69	2.79
70	2.729
71	4.852
74	0.061
75	-0.728
89	-0.424
90	-1.516
97	0.971
98	1.213
99	0.122
100	-0.546
101	-2.062
102	-1.274
107	-0.061
108	-0.545
109	-0.91
111	0.061
113	0.121
114	0.607
116	1.334
123	8.369
125	11.888
126	6.005
127	12.433

Continued on next page

Table A.3 – *Continued from previous page*

Residue	Residual Dipolar Coupling (Hz)
128	16.557
129	8.248
130	9.279
132	13.464
133	6.004
134	14.253
136	-4.367
137	-4.125
138	-1.213
140	-2.062
141	-4.67
142	-7.4
144	-2.487
145	-6.369
147	-3.033
148	-1.213
149	-6.429
150	11.342
151 [Nϵ1-Hϵ1]	-3.785
152	-4.549
155	13.586
156	11.038
159	12.798
160	7.096
161	3.518
162	6.55
164	-1.031
165	-1.152
166	-0.061

Appendix B

Principles of Protein NMR

B.1 Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a phenomenon in which atomic nuclei in a magnetic field absorb and re-emit electromagnetic radiation at specific frequencies, depending on the strength of both the magnetic field and the properties of the isotopes involved. Nuclear magnetism is a consequence of spin angular momentum (spin), a feature of quantum mechanics characterised by the nuclear spin quantum number, I, which is itself related to the make-up of the nucleus (236).

B.1.1 Nuclear Spin

Nuclei with even mass and atomic numbers, such as ^{12}C, have $I = 0$ and are NMR inactive. Nuclei with odd mass numbers have half-integral spin quantum numbers (e.g. $I = 1/2$ or $I = 5/2$). The lifetime of magnetic states for nuclei with $I = 1/2$ is significantly longer than for nuclei with I greater than $1/2$, which makes these nuclei easier to study

as NMR resonance lines are less broad. Particularly important nuclei in biomolecules with $I = 1/2$ are ^{1}H, ^{13}C, ^{15}N and ^{31}P. Nuclei with an even mass number and an odd atomic number, such as the deuteron (^{2}H), have $I = 1$ (236).

B.1.2 Nuclear Magnetism

Nuclear spin angular momentum is a quantised vector quantity, adopting discreet values. Nuclei with a non-zero spin angular momentum have magnetic moments, μ, which are also quantised and proportional to the spin angular momentum vector. This essentially means that the nucleus generates a small magnetic field. When an external magnetic field, B_0, is applied, there is an interaction between the two magnetic fields. In a magnetic field, nuclei with spin can have a number of different energy states, depending on the spin quantum number. Specifically, $2I + 1$ energy levels can be adopted. Over time, a small net alignment of magnetic moments in lowest energy states occurs. This produces a *bulk magnetisation* along the applied field (figure B.1) (236, 245, 251).

B.1.3 Larmor Precession

The Larmor precession refers to a precession of individual magnetic moments around the external field, which is fixed and, by convention, aligned along the z axis (figure B.2). The frequency of this precession, the Larmor frequency, is dependent on the strength of magnetic field, B_0, and gyromagnetic ratio of the nucleus, γ, (the ratio of

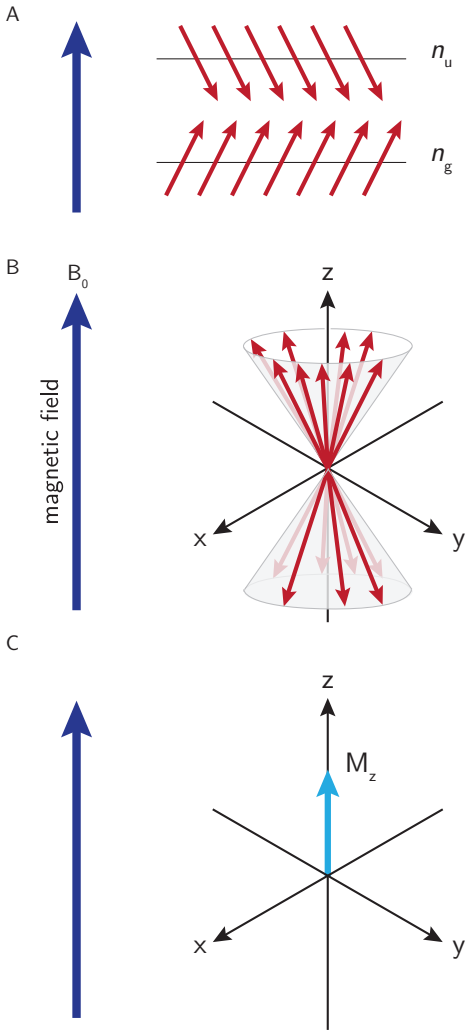


Figure B.1: Bulk Magnetisation - **A** Spin 1/2 nuclei have two possible energy states when interacting with an applied field (blue). Magnetic moments occupy these states with a slight preference for the lower energy ground state, n_g compared with the upper state n_u due to the Bolzmann distribution. **B** If considered with the same origin, these magnetic moments are randomly distributed around the z axis with different xy components. **C** In the presence of an external magnetic field, the xy components will average out to 0 (i.e. in line with the z axis) but the energetically favourable population will favour an alignment in the direction of the magnetic field. Over time, there is an equilibrium in which magnetic moments align to produce a bulk magnetisation (light blue), M_z along the z axis. Figures based on Campbell, 2012 and Keeler, 2005 (245, 251).

the magnetic moment to its angular momentum, which differs for different nuclei) and given by $\omega_0 = \gamma B_0$ (251).

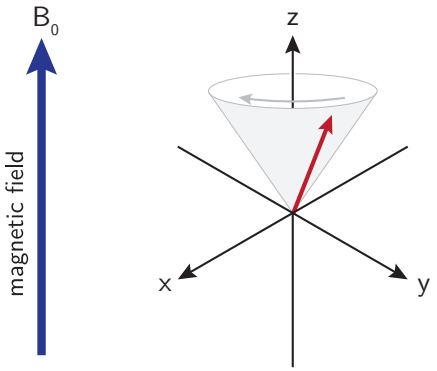


Figure B.2: Larmor Precession - The magnetic moment (red) rotates around the axis of the field at a constant angle. Figure based on Keeler, 2005 (251).

B.2 Nuclear Magnetic Resonance Spectroscopy

Radio frequency (RF) pulses perpendicular to the external magnetic field, at the correct frequency, produce a field, B_1, that bring multiple distributed magnetic dipoles into phase in the x-y plane neutralising the averaging effect. In NMR experiments, this can be used to bring the bulk nuclear magnetisation vector away from the equilibrium position along the z axis and perpendicular to the magnetic field (a 90° pulse).

Similar to individual magnetic moments, the bulk magnetisation, M_0, rotates at an angular rate of the Larmor frequency, producing a signal that can be detected. The precessing magnetisation generates a current, which gives a signal called the free induction decay (FID). This is detected using a coil aligned in the xy plane (figure B.3)

(251). This FID signal can be converted to a frequency-based spectrum by Fourier transform (figure B.4) (245).

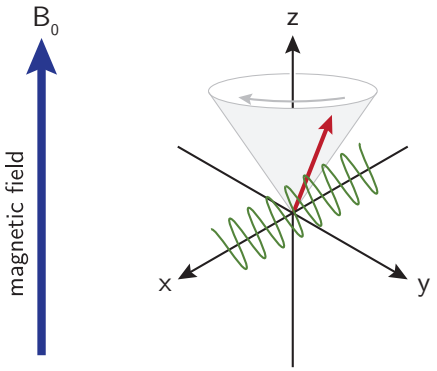


Figure B.3: NMR Detection - Precessing magnetisation will induce a current in a coil wound around the x-axis (green). Figure based on Keeler, 2005 (251).

A pulse sequence is the series of radio frequency pulses and delays used in an NMR experiment to generate a free induction decay signal, which, depending on the pulses used, gives a range of information. Pulses are generated along the x axis by applying field B_1, which then induces precession of the equilibrium magnetisation. Different pulses have a different effect on the bulk equilibrium magnetisation; e.g. if applying a pulse for time t can rotate the magnetisation 90°(perpendicular to +z axis), a pulse for $2t$ would rotate it 180° along the -z axis. Combining pulses in specific ways can be used to obtain a range of information (245, 298).

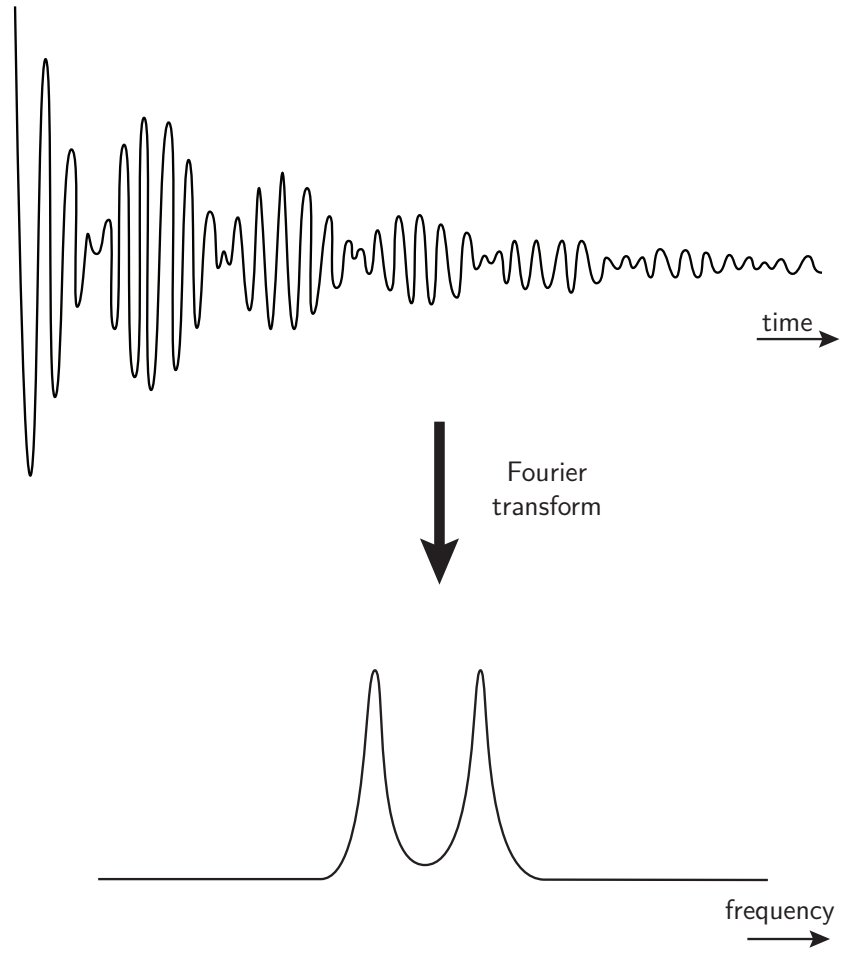


Figure B.4: NMR Fourier Transform - Diagram showing conversion of a complex signal to a frequency-dependent spectrum by fourier transform. The two lines (peaks) in the frequency-dependent plot have the same decay rate and intensity but different frequencies. Figure based on Campbell, 2012 (245).

B.2.1 Chemical Shift

The applied magnetic field, B_0, in an NMR experiment gives rise to a current in the electrons surrounding nuclei, which produces a shielding field proportional to the applied magnetic field. This means the effective magnetic field, B_{eff}, is dependent on the shielding constant, σ, which is determined by the nature of the electrons around the nucleus: $B_{eff} = B_0(1 - \sigma)$ (245).

In NMR, the resonant frequency which is proportional to the magnetic field, is given as chemical shift, a measure of frequency relative to a reference compound and independent of the applied magnetic field. Chemical shift, $\delta = (\nu - \nu_{reference})/\nu_0$ where $(\nu - \nu_{reference})$ is the difference in frequency from that measured and the reference, in Hz and ν_0 is the applied spin resonant frequency in MHz, which depends on the applied magnetic field and nuclei. Chemical shift is therefore a very small number, given as parts per million (ppm) (236, 245).

Particular chemical shifts are characteristic of chemical groups or influences from environmental effects through space. This allows resonances of groups with similar chemistry to be assigned, which is important for structure determination and interaction studies.

B.3 Protein NMR

NMR is a particularly powerful tool for investigating the structure and dynamics of protein molecules. NMR can be carried out in solution at near physiological conditions, providing a crucial alternative for obtaining atomic-resolution structural information of proteins that cannot be crystallised. Furthermore, as data is collected from proteins in solution, NMR experiments can also be used to investigate protein dynamics and binding interactions with ligands (232). In order to carry out meaningful protein NMR experiments, it is normally necessary to isotopically label recombinant protein with ^{15}N and / or ^{13}C (299).

B.3.1 Multidimensional NMR

One important development for the study of proteins by NMR has been the emergence of NMR techniques that produce multi-dimensional NMR specta. These typically use complicated pulse sequences to yield additional information in the extra dimensions. Two important additions to the pulse sequences are evolution (t_1) and mixing periods after the initial preparation stage and prior to acquisition. In two-dimensional spectra, the evolution time can be altered in increments, with the resultant spectra plotted along the second dimension (236). A great deal of different NMR experiments have been developed that allow highly specific information to be generated.

A very important experiment used in protein NMR is the HSQC (heteronuclear single-quantum coherence), first reported by Bodenhausen and Ruben in 1980 (237), which gives a peak for every ^{15}N-^{1}H group (and therefore a peak for every amino acid residue except proline). The process underlying the HSQC experiment is the transfer of proton magnetisation to a second nucleus (either ^{15}N or ^{13}C) through a bond, using an INEPT (insensitive nuclei enhanced by polarization transfer) method, which utilises J-coupling and distributions of spin populations. After an evolution period (t_1), the now frequency-labelled coherence is transferred back to the proton by a second INEPT step and the signal is then detected. HSQC experiments record a series of experiments where the time delay t_1 is systematically increased. The ^{1}H signal is acquired in the directly measured dimension in each experiment, whereas the chemical shift of the less sensitive isotope, ^{15}N or ^{13}C, is recorded in the indirect dimension, gathered over the series of experiments (figure B.5) (236).

Three-dimensional experiments used in protein NMR typically build onto an HSQC experiment by adding an additional dimension that contains information from other nuclei, either connected through space or through bond. Triple-resonance experiments (i.e. those using three types of nuclei; ^{1}H, ^{15}N and ^{13}C) can be particularly useful in sequential assignment of protein backbone residues (section 4.3.3) (236, 245).

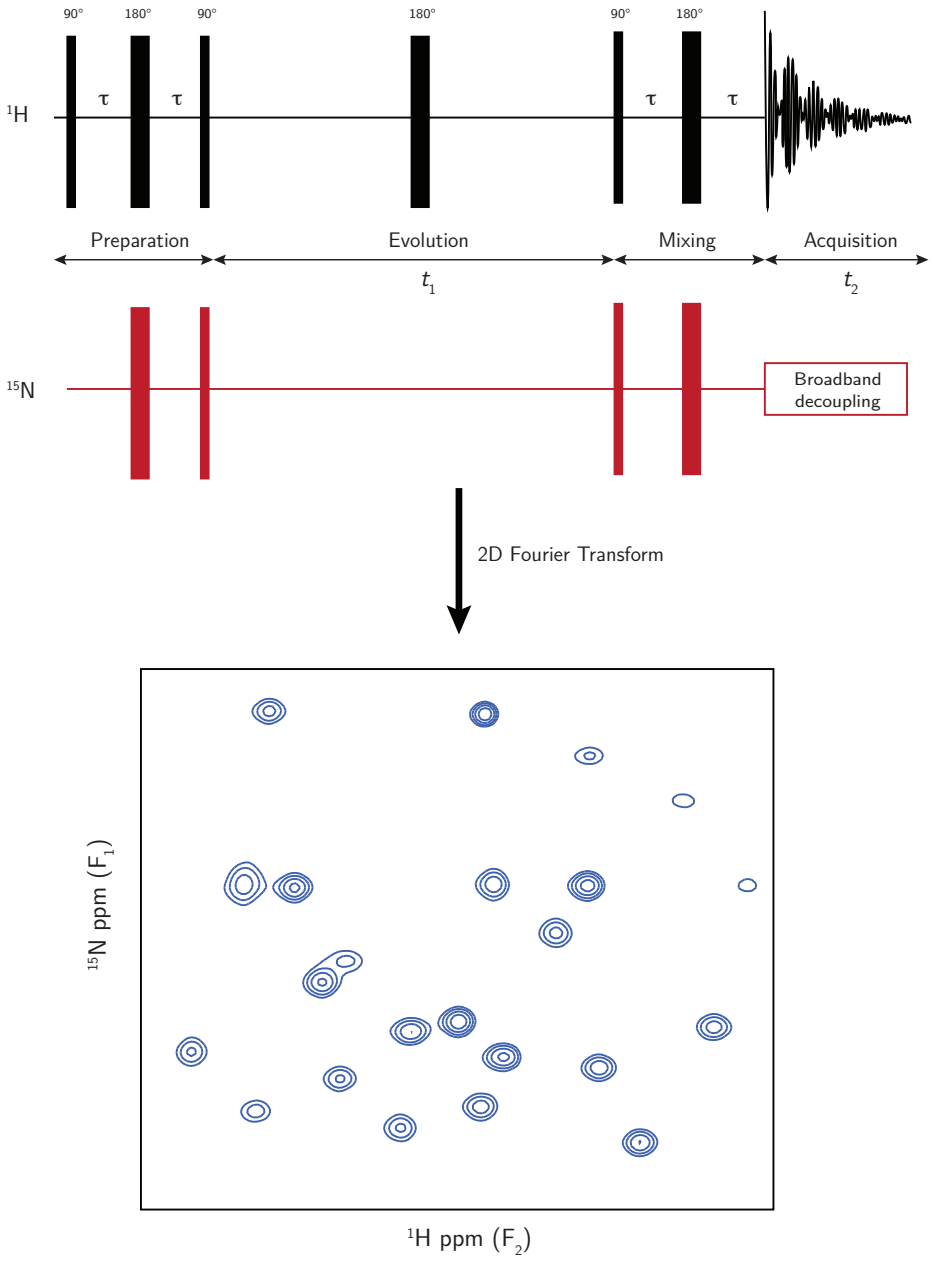


Figure B.5: HSQC Pulse Sequence and Spectrum - A ^{1}H, ^{15}N-HSQC spectrum (bottom) shows peaks separated in both the ^{1}H and ^{15}N dimensions. HSQC spectra are generated using the pulse sequence shown (top). Radio frequency pulses are applied to the ^{1}H (black) and ^{15}N (red) nuclei as indicated. Narrow and wide bars represent 90° and 180° pulses respectively. In the ^{1}H, ^{15}N-HSQC experiment, $\tau = 1/4\ J_{NH}$ (2.5 ms; J_{NH} *ca.* 100 Hz). In NH_2 selective HSQC experiments, $\tau = 1/8\ J_{NH}$ (1.275 ms). Figure based on Campbell, 2012; Cavanagh, 2007 and experimental data (section 4.3.1) (236, 237, 245).

B.3.2 Protein Structure Determination

Protein structure determination using NMR relies on a restrained molecular dynamics approach (section B.3.2.4). In addition to chemical shift values, several additional phenomena can be exploited to help solve the structure. The nuclear Overhauser effect is a through-space phenomenom that can be used to give distance restraints. Residual dipolar couplings give information about relative orientations of NH bond vectors (235, 245, 300, 301).

B.3.2.1 The Nuclear Overhauser Effect

The nuclear Overhauser effect (NOE) is caused by dipole-dipole induced relaxation where intensity changes are observed due to a change in the relaxation effect from a nearby dipole (302). The strength of this effect is dependant on distance ($^1/r^6$) and is important in structure determination as it allows for identification of nuclei that are sequentially far away but structurally close (figure B.6). Furthermore, medium-range NOEs can help identify secondary structure regions, such as α-helices, based on characteristic distances of nuclei (245).

B.3.2.2 {^{1}H} ^{15}N Heteronuclear steady-state NOE

The backbone$^1H^N$-^{15}N bond can be thought of as a vector with two spins, P and Q. The magnetic field generated by the magnetic moment of P interacts with Q. In a $^1H^N$

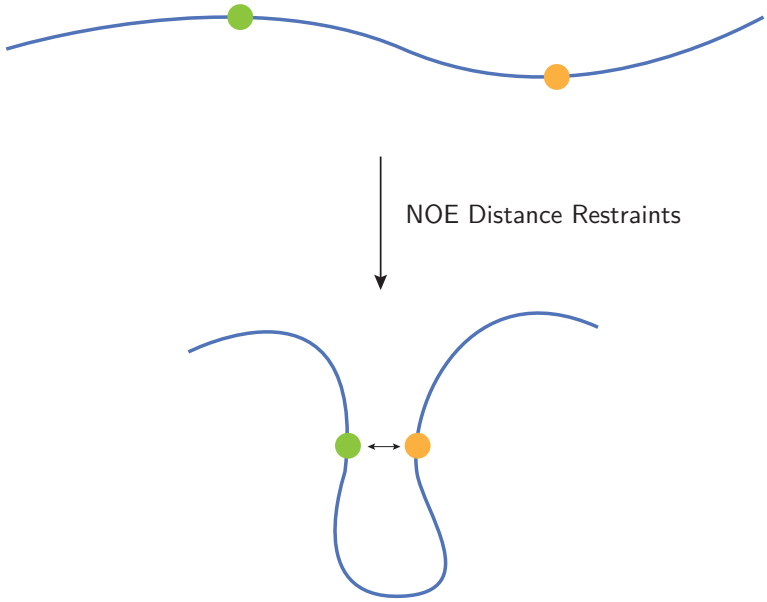


Figure B.6: Long-Range NOE Distance Restraints - Identification of long-range NOEs are important for structure calculations to ensure that the protein 'folds' correctly. Spatially close, but sequentially distant, spins are key to understanding how the the protein is structured. The intensity of NOE peaks is inversely proportional ($1/r^6$) to the distance between the nuclei involved (245). This diagram illustrates the principle that the backbone (blue) must adopt a limited subset of conformations based on NOE distance restraints between spatially close resonances (green and orange). In structural calculations, models which satisfy measured NOE intensities are produced.

steady-state NOE experiment, the proton spin is irradiated with radiofrequency pulses, which saturates its spin, neutralising its magnetisation. The effect of this saturation on the connected ^{15}N spin is determined by comparing the resultant spectra with a reference (in which ^{1}H is not saturated)..

Two spectra are recorded, one with proton saturation (NOE) and one without, a reference (NONOE). The NOE ratio is given by a peak intensity from the saturated spectra / intensity from reference spectra. As ^{15}N relaxation is related to mobility of the bond, this ratio gives an indication of residue flexibility. Higher ratios around 0.75 (i.e. where peak intensity in the saturated spectra is similar to the spectra without proton saturation) indicate rigid bond vectors from structured residues. Lower ratio values signify a greater degree of backbone flexibility. (245, 251, 261).

B.3.2.3 Residual Dipolar Couplings

Dipolar couplings are the direct interactions between nuclei (figure B.7). Residual dipolar couplings (RDCs) give long-range orientational information. In an isotropic solution, the effect of these dipole-dipole interactions is averaged to zero as all directions relative to the external magnetic field are sampled uniformly due to random tumbling. In protein solutions, dipolar couplings are therefore not normally observed. In the case of solution NMR and liquid crystals, the following equation gives a value for the dipolar coupling, D_{PQ}, in Hertz for two nuclei, P and Q, with spin 1/2 :

$$D_{PQ} = -\frac{\mu_0 \gamma_P \gamma_Q h}{(2\pi r_{PQ})^3} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle$$

γ_P and γ_Q are the magnetogyric ratios for nuclei P and Q respectively. μ_0 is the permittivity of space, r is the distance between the nuclei, h is Planck's constant and θ is the angle between the nuclear vectors (303, 304).

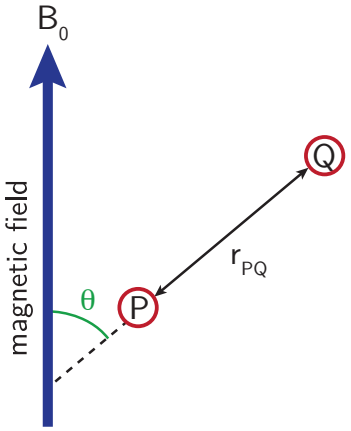


Figure B.7: Dipolar Couplings - The dipole-dipole interaction between two nuclei, P and Q, is dependent on both the distance between them, r, and the angle, θ, of the bond relative to the external magnetic field, B_0. In solids, θ is fixed and dipolar couplings, D_{PQ}, can be large (kHz). In solutions, rotational diffusion causes the dipolar coupling to average out to zero (236, 245, 304).

Several methods have been developed to induce partial molecular alignment so that residual dipolar couplings (RDCs) can be observed. These include the use of lipid bicelles, alkyl-poly(ethylene glycol)/alcohol mixtures, negatively charged filamentous bacteriophage and compressed and stretched polyacrylamide gels (236, 303).

The RDC between two nuclear spins can be measured by comparing the splitting

observed under two types of conditions. An ^{1}H, ^{15}N-HSQC is recorded without decoupling during t_1. This results in two peaks (a doublet) being observed for every $^{1}H^{N}$-$^{15}N^{H}$ resonance. The separation between each pair of peaks is a result of $^{1}J_{NH}$ scalar coupling, which is approximately -93 Hz. In an alignment medium, the observed splitting values are affected. This gives information as to how the N-H bonds are aligned relative to a molecular alignment tensor.

As this approach doubles the number of peaks observed for any given protein, it can complicate interpretation due to overcrowding. An alternative is to split the doublet subcomponents into separate spectra using the IPAP (in-phase / antiphase) technique (305). The IPAP experiment produces two spectra, one which gives the upfield doublet component and the other, which gives the downfield doublet component. Measuring this for both an isotropic sample and for a partially aligned sample (e.g. dilute liquid crystalline phase of $C_{12}E_6$/hexanol mixture (195)) gives four spectra, which can be used to measure N-H RDCs (figure B.8) (236).

The equation for observable dipolar coupling between two nuclei, D_{PQ}, can be alternatively expressed as a function of an axial and a rhombic component of the alignment tensor, D_a^{PQ} and R respectively:

$$D_{PQ}(\theta,\phi) = D_a^{PQ}\left\{(3\cos^2\theta - 1) + \frac{3}{2}R(\sin^2\theta cos2\phi)\right\}$$

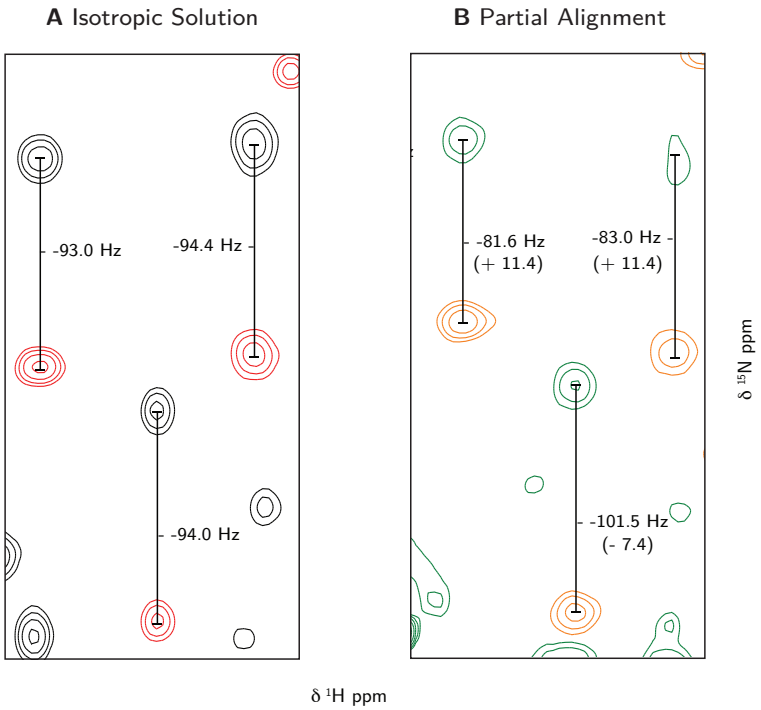


Figure B.8: Residual Dipolar Couplings - Comparison of peak separations between isotropic and anisotropic solutions for equivalent resonances. **A** Splitting in isotropic solution is uniform (no contribution from RDCs) and corresponds to the ${}^1J_{\mathrm{NH}}$ scalar coupling (magnitude approximately -93 Hz). Upfield and downfield doublet components are collected separately and shown in red and black respectively. **B** Peak separation in alignment medium (PEG/hexanol) is given by ${}^1J_{\mathrm{NH}} + D_{\mathrm{NH}}$ and may be larger or smaller than just ${}^1J_{\mathrm{NH}}$, depending on whether the RDC contribution is negative or positive. Measuring the difference between an isotropic solution and aligned medium gives the RDC contribution. Upfield and downfield doublet components shown in green and orange respectively. Experimental data from SMARCAD1 dCUE (section 5.5).

θ is the angle between the NH bond vector and the z-axis of the alignment tensor, which in addition to ϕ, the angle between the projection of the NH vector on the x-y plane of the alignment tensor and the x-axis, are the only two variables that dictate the RDC value (263). D_a^{PQ} and R can be estimated from the distribution of observed RDCs and used in structure calculations (264). Observed RDC values are binned and plotted as a histogram (figure B.9). For a given fixed distance (r_{PQ}) P-Q interaction, the values for D_a^{PQ} at either end of the histogram correspond to orientations of the P-Q vector closest to the z ($\theta = 0°$) and y axes ($\theta = 90°$ and $\phi = 90°$) of the alignment tensor. The more common dipolar coupling value is coincident with the size of the bond vector aligned along the x axis ($\theta = 90°$ and $\phi = 0°$). Based on RDC values at these axial positions, the following equations can be used to estimate values for D_a^{PQ} and R: $D_{zz}^{PQ} = 2D_a^{PQ}$; $D_{yy}^{PQ} = -D_a^{PQ}(1 + 1.5R)$; $D_{xx}^{PQ} = -D_a^{PQ}(1 - 1.5R)$ (300, 301).

B.3.2.4 Structure Calculations

3D models of protein structures can be calculated using a simulated annealing / molecular dynamics approach based on an energy minimisation algorithm. Experimentally-derived measurements such as NOE restraints, RDC values and dihedral angle estimates (from chemical shift data) provide potential energy terms for the protocol. Other covalent geometry terms, such as the improper dihedral potential, the RAMA potential (a multi-dimensional torsion angle database term, which biases angles in favour of com-

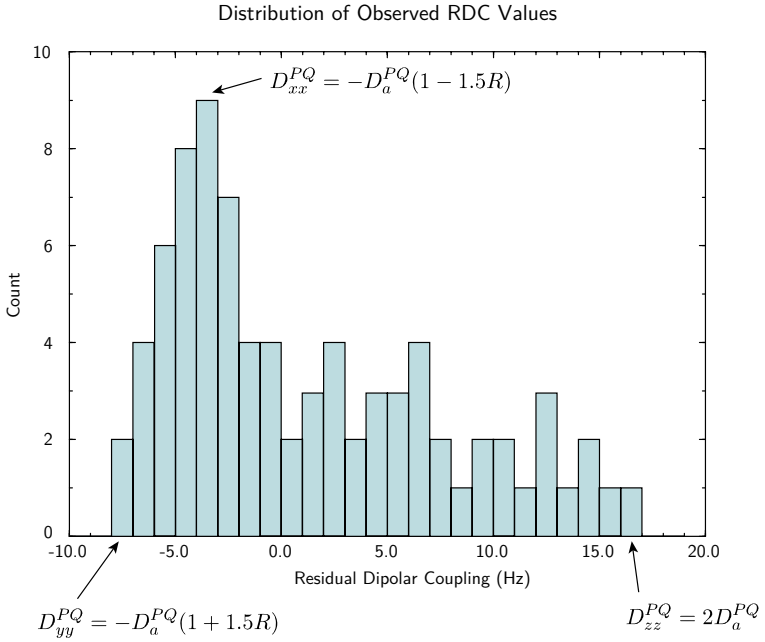


Figure B.9: Example of RDC Distribution - Binned RDC values are plotted as a histogram. The extremes of the distribution and most common value (highest count) correlate with orientations parallel to axes relative to the alignment tensor. D_a^{PQ} and R can be estimated from the equations shown (264).

mon angles from known structures) and non-bonded contacts also contribute energy terms for the calculations (figure B.10) (201, 202).

The process of simulated annealing uses a Monte Carlo method to minimise the following potential energy function:

$$E = E_{chem} + \omega_{data} \cdot E_{data}$$

This function incorporates numerous parameters, each of which have a different energy minimum, to find a global minimum. E_{chem} represents chemical information for atom radii and bond lengths, E_{data} is dependent on how well the observed experimental data fits with calculated results and ω_{data} is a weighting term (300, 301, 306, 307). These terms are summations of distinct energy parameters:

$$E_{chem} = E_{cov} + E_{vdW}$$

$$E_{cov} = E_{bond} + E_{angl} + E_{impr}$$

$$E_{data} = E_{NOE} + E_{tor} + E_{RDC}$$

The protocol utilises a gradient-based minimisation; starting molecular dynamics on a randomly disordered chain of amino acids at high temperature and reducing the temperature as the calculation proceeds to generates structures with a global energy minimum (i.e. structures that obey the measured restraints). Repeating this process

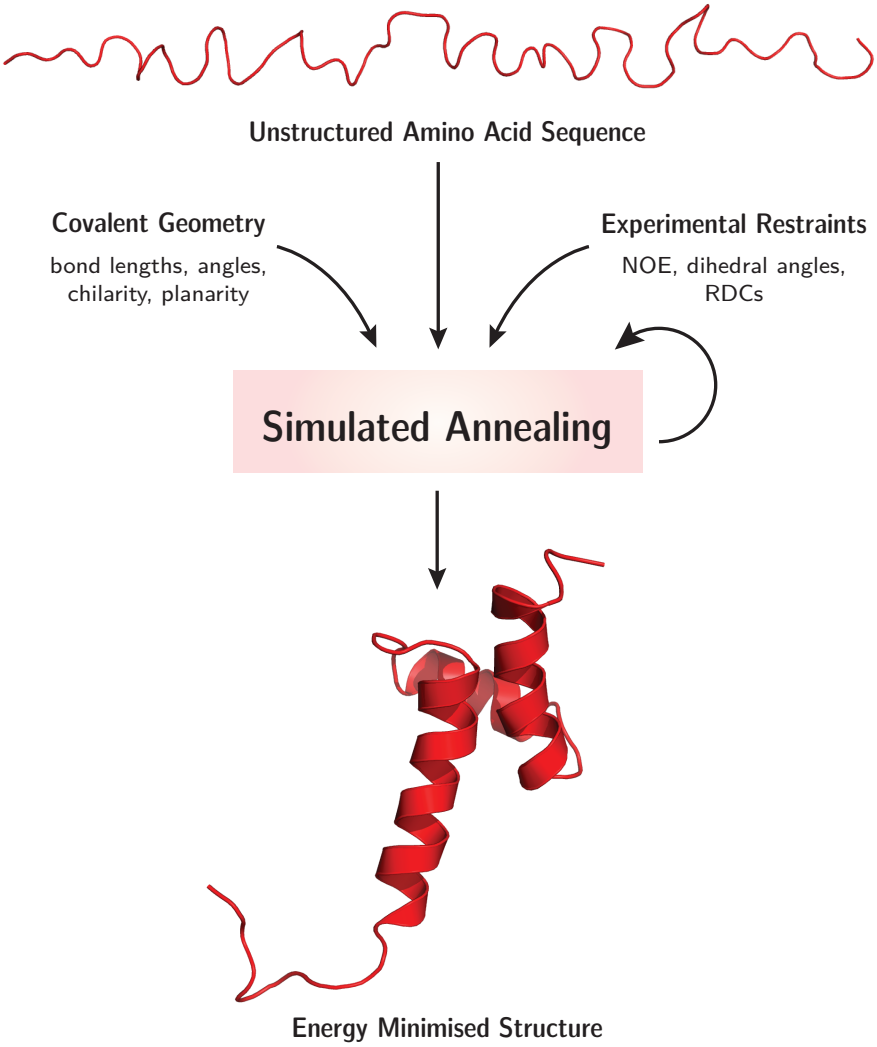


Figure B.10: Schematic of NMR Structure Calculation - The simulated annealing step uses a starting, unstructured, sequence in conjunction with covalent geometries and experimentally-derived restraints to produce energy minimised structures. This is an iterative process and repeated many times. The structures can be refined or validated. Figure adapted from Redfield, 2009 (300, 301).

numerous times with sufficient NMR restraints and refinement steps produces a family of highly similar, low potential energy models, which accurately reflect the structure of the protein in solution. Calculated structures can be validated by analysing how well the models fit with the experimental data, their stereochemical quality and any van der Waal violations (208).

References

[1] V a Spencer and J R Davie. **Role of covalent modifications of histones in regulating gene expression.** *Gene*, **240**(1):1–12, November 1999. 1

[2] Philip Ball. **Portrait of a molecule.** *Nature*, **421**(6921):421–2, January 2003. 1

[3] Diana C Hargreaves and Gerald R Crabtree. **ATP-dependent chromatin remodeling: genetics, genomics and mechanisms.** *Cell research*, **21**(3):396–420, March 2011. 1, 3, 4, 6, 21, 22, 23, 26

[4] Gary Felsenfeld and Mark Groudine. **Controlling the double helix.** *Nature*, **421**(6921):448–453, 2003. 1, 2, 3, 6

[5] Lisette Mohrmann and C Peter Verrijzer. **Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes.** *Biochimica et biophysica acta*, **1681**(2-3):59–73, January 2005. 1

[6] Kazuhiro Maeshima, Saera Hihara, and Mikhail Eltsov. **Chromatin structure: does the 30-nm fibre exist in vivo?** *Current opinion in cell biology*, **22**(3):291–7, June 2010. 2

[7] Richard I Gregory and Ramin Shiekhattar. **Chromatin modifiers and carcinogenesis.** *Trends in cell biology*, **14**(12):695–702, December 2004. 3, 4

[8] Christopher L Woodcock and Rajarshi P Ghosh. **Chromatin higher-order structure and dynamics.** *Cold Spring Harbor perspectives in biology*, **2**(5):a000596, May 2010. 3, 5

[9] Andrew J Bannister and Tony Kouzarides. **Regulation of chromatin by histone modifications.** *Cell research*, **21**(3):381–95, March 2011. 3, 4, 6, 8, 9, 10, 11

[10] Guillaume J Filion, Joke G van Bemmel, Ulrich Braunschweig, Wendy Talhout, Jop Kind, Lucas D Ward, Wim Brugman, Inês J de Castro, Ron M Kerkhoven, Harmen J Bussemaker, and Bas van Steensel. **Systematic protein location mapping reveals five principal chromatin types in Drosophila cells.** *Cell*, **143**(2):212–24, October 2010. 3

[11] Dirk Schübeler. **Chromatin in multicolor.** *Cell*, **143**(2):183–4, October 2010. 3

[12] Jingping Yang and Victor G Corces. **Chromatin insulators: a role in nuclear organization and gene expression.** *Advances in Cancer Research*, **110**:43–76, 2011. 4

[13] Jutta Vogelmann, Alessandro Valeri, Emmanuelle Guillou, Olivier Cuvier, and Marcello Nollmann. **Roles of chromatin insulator proteins in higher-order chromatin organization and transcription regulation.** *Nucleus (Austin, Tex.)*, **2**(5):358–69, September 2011. 4

[14] Chin-Tong Ong and Victor G Corces. **Insulators as mediators of intra- and inter-chromosomal interactions: a common evolutionary theme**. *Journal of Biology*, **8**(8):73, 2009. 4

[15] Thomas Schalch, Sylwia Duda, David F Sargent, and Timothy J Richmond. **X-ray structure of a tetranucleosome and its implications for the chromatin fibre.** *Nature*, **436**(7047):138–41, July 2005. 5

[16] Guohong Li and Danny Reinberg. **Chromatin higher-order structures and gene regulation.** *Current opinion in genetics & development*, **21**(2):175–86, April 2011. 5

[17] Margot P Scheffer, Mikhail Eltsov, and Achilleas S Frangakis. **Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei.** *Proceedings of the National Academy of Sciences of the United States of America*, **108**(41), October 2011. 5

[18] Philip J J Robinson, Louise Fairall, Van a T Huynh, and Daniela Rhodes. **EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure.** *Proceedings of the National Academy of Sciences of the United States of America*, **103**(17):6506–11, April 2006. 5

[19] K Luger, A W Mäder, R K Richmond, D F Sargent, and T J Richmond. **Crystal structure of the nucleosome core particle at 2.8 A resolution.** *Nature*, **389**(6648):251–260, 1997. 5, 6

[20] Chunyang Zheng and Jeffrey J Hayes. **Structures and interactions of the core histone tail domains.** *Biopolymers*, **68**(4):539–46, April 2003. 5

[21] Curt a Davey, David F Sargent, Karolin Luger, Armin W Maeder, and Timothy J Richmond. **Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution.** *Journal of molecular biology*, **319**(5):1097–113, June 2002. 5

[22] W L DeLano. **The PyMOL Molecular Graphics System**. *DeLano Scientific*, 2002. 5, 37, 204

[23] V G Allfrey, R Faulkner, and A E Mirsky. **ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS**. *Proceedings of the National Academy of Sciences of the United States of America*, **51**(5):786–794, 1964. 6

[24] Erica L Mersfelder and Mark R Parthun. **The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure.** *Nucleic acids research*, **34**(9):2653–62, January 2006. 6

[25] B D Strahl and C D Allis. **The language of covalent histone modifications.** *Nature*, **403**(6765):41–5, January 2000. 6, 13

[26] Thomas Jenuwein and C D Allis. **Translating the histone code.** *Science*, **293**(5532):1074–80, 2001. 6

[27] Abcam, Tony Kouzarides, and Andy Bannister. **Guide to Epigenetic Marks**, 2011. 7

[28] Anna Portela and Manel Esteller. **Epigenetic modifications and human disease.** *Nature biotechnology*, **28**(10):1057–68, October 2010. 7

[29] Taraswi Banerjee and Debabrata Chakravarti. **A peek into the complex realm of histone phosphorylation.** *Molecular and cellular biology*, **31**(24), October 2011. 7

[30] Alessia Peserico and Cristiano Simone. **Physical and functional HAT/HDAC interplay regulates protein acetylation balance.** *Journal of biomedicine & biotechnology*, **2011**(Table 1):371832, January 2011. 8

[31] X-J Yang and E Seto. **HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention.** *Oncogene*, **26**(37):5310–8, August 2007. 8

[32] Alison L Clayton, Catherine a Hazzalin, and Louis C Mahadevan. **Enhanced histone acetylation and transcription: a dynamic perspective.** *Molecular cell*, **23**(3):289–96, August 2006. 8

[33] Takashi Ito. **Role of histone modification in chromatin dynamics.** *Journal of biochemistry*, **141**(5):609–14, May 2007. 8

[34] Keiichi Sugiyama, Kazumitsu Sugiura, Tomohiro Hara, Kenji Sugimoto, Hiroshi Shima, Kei Honda, Koichi Furukawa, Shunichi Yamashita, and Takeshi Urano. **Aurora-B associated protein phosphatases as negative regulators of kinase activation.** *Oncogene*, **21**(20):3103, 2002. 8

[35] Hidemasa Goto, Yoshihiro Yasui, Erich a Nigg, and Masaki Inagaki. **Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation.** *Genes to cells : devoted to molecular & cellular mechanisms*, **7**(1):11–7, January 2002. 8

[36] Sukesh R Bhaumik, Edwin Smith, and Ali Shilatifard. **Covalent modifications of histones during development and disease pathogenesis.** *Nature structural & molecular biology*, **14**(11):1008–16, November 2007. 8

[37] Bingsheng Chang, Yue Chen, Yingming Zhao, and Richard K Bruick. **JMJD6 is a histone arginine demethylase.** *Science (New York, N.Y.)*, **318**(5849):444–7, October 2007. 9

[38] Yu-ichi Tsukada, Jia Fang, Hediye Erdjument-Bromage, Maria E Warren, Christoph H Borchers, Paul Tempst, and Yi Zhang. **Histone demethylation by a family of JmjC domain-containing proteins.** *Nature*, **439**(7078):811–6, February 2006. 9

[39] Bryan M Turner. **Reading signals on the nucleosome with a new nomenclature for modified histones.** *Nature structural & molecular biology*, **12**(2):110–2, February 2005. 9

[40] S S Ng, W W Yue, U Oppermann, and R J Klose. **Dynamic protein methylation in chromatin biology.** *Cellular and molecular life sciences : CMLS*, **66**(3):407–22, February 2009. 9

[41] Mark T Bedford and Steven G Clarke. **Protein arginine methylation in mammals: who, what, and why.** *Molecular cell*, **33**(1):1–13, January 2009. 9

[42] Fei Lan and Yang Shi. **Epigenetic regulation: methylation of histone and non-histone proteins.** *Science in China. Series C, Life sciences / Chinese Academy of Sciences*, **52**(4):311–22, April 2009. 9

[43] Artem Barski, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. **High-resolution profiling of histone methylations in the human genome.** *Cell*, **129**(4):823–37, May 2007. 9

[44] Jeffrey a Rosenfeld, Zhibin Wang, Dustin E Schones, Keji Zhao, Rob DeSalle, and Michael Q Zhang. **Determination of enriched histone modifications in non-genic portions of the human genome.** *BMC genomics*, **10**:143, January 2009. 9

[45] Mark Bycroft. **Recognition of non-methyl histone marks.** *Current opinion in structural biology*, October 2011. 9, 10

[46] Vikki M Weake and Jerry L Workman. **Histone ubiquitination: triggering gene activity.** *Molecular cell*, **29**(6):653–63, March 2008. 10

[47] Lori Frappier and C Peter Verrijzer. **Gene expression control by protein deubiquitinases.** *Current opinion in genetics & development*, **21**(2):207–13, April 2011. 10

[48] Holger Richly, Luciana Rocha-Viegas, Joana Domingues Ribeiro, Santiago Demajo, Gunes Gundem, Nuria Lopez-Bigas, Tekeya Nakagawa, Sabine Rospert, Takashi Ito, and Luciano Di Croce. **Transcriptional activation of polycomb-repressed genes by ZRF1.** *Nature*, **468**(7327):1124–8, December 2010. 10

[49] Hengbin Wang, Liangjun Wang, Hediye Erdjument-Bromage, Miguel Vidal, Paul Tempst, Richard S Jones, and Yi Zhang. **Role of histone H2A ubiquitination in Polycomb silencing.** *Nature*, **431**(7010):873–878, 2004. 10

[50] Jorge A Iñiguez Lluhí. **For a healthy histone code, a little SUMO in the tail keeps the acetyl away.** *ACS Chemical Biology*, **1**(4):204–206, 2006. 10, 181

[51] Nagarama Kothapalli, Gabriela Camporeale, Alice Kueh, Yap C Chew, Anna M Oommen, Jacob B Griffin, and Janos Zempleni. **Biological functions of biotinylated histones.** *The Journal of nutritional biochemistry*, **16**(7):446–8, July 2005. 10, 12

[52] Toshinobu Kuroishi, Luisa Rios-Avila, Valerie Pestinger, Subhashinee S K Wijeratne, and Janos Zempleni. **Biotinylation is a natural, albeit rare, modification of human histones.** *Molecular genetics and metabolism*, **4**, September 2011. 10, 12

[53] GL Cuthbert, Sylvain Daujat, and AW Snowden. **Histone deimination antagonizes arginine methylation. 118**:545–553, 2004. 11

[54] Yanming Wang, Joanna Wysocka, Joyce Sayegh, Young-Ho Lee, Julie R Perlin, Lauriebeth Leonelli, Lakshmi S Sonbuchner, Charles H McDonald, Richard G Cook, Yali Dou, Robert G Roeder, Steven Clarke, Michael R Stallcup, C David Allis, and Scott a Coonrod. **Human PAD4 regulates histone arginine methylation levels via demethylimination.** *Science (New York, N.Y.)*, **306**(5694):279–83, October 2004. 11

[55] Saskia Hemmers, John R Teijaro, Sanja Arandjelovic, and Kerri a Mowen. **PAD4-mediated neutrophil extracellular trap formation is not required for immunity against influenza infection.** *PloS one*, **6**(7):e22043, January 2011. 11

[56] Volker Brinkmann, Ulrike Reichard, Christian Goosmann, Beatrix Fauler, Yvonne Uhlemann, David S Weiss, Yvette Weinrauch, and Arturo Zychlinsky. **Neutrophil extracellular traps kill bacteria.** *Science (New York, N.Y.)*, **303**(5663):1532–5, March 2004. 11

[57] Indira Neeli, Salar N Khan, and Marko Radic. **Histone deimination as a response to inflammatory stimuli in neutrophils.** *Journal of immunology (Baltimore, Md. : 1950)*, **180**(3):1895–902, February 2008. 11

[58] Michael O Hottiger. **ADP-ribosylation of histones by ARTD1: An additional module of the histone code?** *FEBS letters*, **585**(11):1595–9, June 2011. 11

[59] Simon Messner and Michael O Hottiger. **Histone ADP-ribosylation in DNA repair, replication and transcription.** *Trends in cell biology*, **21**(9):534–542, July 2011. 11

[60] Christopher J Nelson, Helena Santos-Rosa, and Tony Kouzarides. **Proline isomerization of histone H3 regulates lysine methylation and gene expression.** *Cell*, **126**(5):905–16, September 2006. 11

[61] Kaoru Sakabe, Zihao Wang, and Gerald W Hart. **Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code.** *Proceedings of the National Academy of Sciences of the United States of America*, **107**(46):19915–20, November 2010. 12

[62] Chad Slawson and Gerald W Hart. **O-GlcNAc signalling: implications for cancer cell biology.** *Nature reviews. Cancer*, **11**(9):678–84, January 2011. 12

[63] Janos Zempleni, Daniel Camara Teixeira, Toshinobu Kuroishi, Elizabeth L Cordonier, and Scott Baier. **Biotin requirements for DNA damage prevention.** *Mutation research*, pages 16–18, August 2011. 12

[64] Helena Santos-Rosa, Antonis Kirmizis, Christopher Nelson, Till Bartke, Nehme Saksouk, Jacques Cote, and Tony Kouzarides. **Histone H3 tail clipping regulates gene expression.** *Nature structural & molecular biology*, **16**(1):17–22, January 2009. 12

[65] Elizabeth M Duncan, Tara L Muratore-Schroeder, Richard G Cook, Benjamin a Garcia, Jeffrey Shabanowitz, Donald F Hunt, and C David Allis. **Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation.** *Cell*, **135**(2):284–94, October 2008. 12

[66] C D Allis, J K Bowen, G N Abraham, C V Glover, and M A Gorovsky. **Proteolytic processing of histone H3 in chromatin: a physiologically regulated event in Tetrahymena micronuclei.** *Cell*, **20**(1):55–64, 1980. 12

[67] C D Allis and J C Wiggins. **Proteolytic processing of micronuclear H3 and histone phosphorylation during conjugation in Tetrahymena thermophila.** *Experimental cell research*, **153**(2):287–98, August 1984. 12

[68] Kavitha Sarma and Danny Reinberg. **Histone variants meet their match.** *Nature reviews. Molecular cell biology*, **6**(2):139–49, February 2005. 13, 14

[69] Daniel R Foltz, Lars E T Jansen, Ben E Black, Aaron O Bailey, John R Yates, and Don W Cleveland. **The human CENP-A centromeric nucleosome-associated complex.** *Nature cell biology*, **8**(5):458–69, May 2006. 13

[70] Jolien S Verdaasdonk and Kerry Bloom. **Centromeres: unique chromatin structures that drive chromosome segregation.** *Nature reviews. Molecular cell biology*, **12**(5):320–32, May 2011. 14

[71] Sean D Taverna, Haitao Li, Alexander J Ruthenburg, C David Allis, and Dinshaw J Patel. **How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.** *Nature structural & molecular biology*, **14**(11):1025–40, November 2007. 13, 15, 16, 17, 18, 19

[72] Patricia M Thompson, Takahiro Gotoh, Marleen Kok, Peter S White, and Garrett M Brodeur. **CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system.** *Oncogene*, **22**(7):1002–11, February 2003. 13

[73] Robyn E Mansfield, Catherine a Musselman, Ann H Kwan, Samuel S Oliver, Adam L Garske, Foteini Davrazou, John M Denu, Tatiana G Kutateladze, and Joel P Mackay. **Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9.** *The Journal of biological chemistry*, **286**(13):11779–91, April 2011. 15, 17

[74] Chen Chen, Timothy J Nott, Jing Jin, and Tony Pawson. **Deciphering arginine methylation: Tudor tells the tale.** *Nature reviews. Molecular cell biology*, **12**(10):629–642, September 2011. 15

[75] Lei Zeng and Ming Ming Zhou. **Bromodomain: an acetyl-lysine binding domain.** *FEBS letters*, **513**(1):124–8, February 2002. 15

[76] D J Owen, P Ornaghi, J C Yang, N Lowe, P R Evans, P Ballario, D Neuhaus, P Filetici, and a a Travers. **The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p.** *The EMBO journal*, **19**(22):6141–9, November 2000. 15, 16

[77] Christin Tse, Elena I Georgieva, Ana B Ruiz-garc, Ramon Sendra, and Jeffrey C Hansen. **Acetylates Nucleosomes and Folded Nucleosomal Arrays in the Absence of Other Protein Subunits *.** *Biochemistry*, **273**(49):32388–32392, 1998. 16

[78] KL Yap. **Keeping It in the Family: Diverse Histone Recognition by Conserved Structural Folds**. *Critical reviews in biochemistry and*, **45**(6):488–505, 2010. 17

[79] Mariann Bienz. **The PHD finger, a nuclear protein-interaction domain.** *Trends in biochemical sciences*, **31**(1):35–40, January 2006. 17

[80] Haitao Li, Serge Ilin, Wooikoon Wang, Elizabeth M Duncan, Joanna Wysocka, C David Allis, and Dinshaw J Patel. **Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF.** *Nature*, **442**(7098):91–5, July 2006. 17, 18

[81] Joanna Wysocka, Tomek Swigut, Hua Xiao, Thomas a Milne, So Yeon Kwon, Joe Landry, Monika Kauer, Alan J Tackett, Brian T Chait, Paul Badenhorst, Carl Wu, and C David Allis. **A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.** *Nature*, **442**(7098):86–90, July 2006. 17

[82] a Arockia Jeyaprakash, Claire Basquin, Uma Jayachandran, and Elena Conti. **Structural basis for the recognition of phosphorylated histone h3 by the survivin subunit of the chromosomal passenger complex.** *Structure (London, England : 1993)*, **19**(11):1625–34, November 2011. 19

[83] Jiaxue Wu, Michael S Y Huen, Lin-Yu Lu, Lin Ye, Yali Dou, Mats Ljungman, Junjie Chen, and Xiaochun Yu. **Histone ubiquitination associates with BRCA1-dependent DNA damage response.** *Molecular and cellular biology*, **29**(3):849–60, February 2009. 19

[84] Anjanabha Saha, Jacqueline Wittmeyer, and Bradley R Cairns. **Chromatin remodelling: the industrial revolution of DNA around histones.** *Nature reviews. Molecular cell biology*, **7**(6):437–47, June 2006. 19

[85] Andrew Flaus, David M a Martin, Geoffrey J Barton, and Tom Owen-Hughes. **Identification of multiple distinct Snf2 subfamilies with conserved structural motifs.** *Nucleic acids research*, **34**(10):2887–905, January 2006. 19, 20, 29

[86] Glenn Hauk and Gregory D Bowman. **Structural insights into regulation and action of SWI2/SNF2 ATPases.** *Current opinion in structural biology*, **21**(6):719–27, December 2011. 19, 23

[87] Daniel P Ryan and Tom Owen-Hughes. **Snf2-family proteins: chromatin remodellers for any occasion.** *Current opinion in chemical biology*, **15**(5):649–56, October 2011. 19, 22, 23, 24

[88] Andrew Flaus and Tom Owen-Hughes. **Mechanisms for ATP-dependent chromatin remodelling: the means to the end.** *The FEBS journal*, **278**(19):3579–95, October 2011. 20

[89] Harald Dürr, Andrew Flaus, Tom Owen-Hughes, and Karl-Peter Hopfner. **Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures.** *Nucleic acids research*, **34**(15):4160–7, January 2006. 20, 22

[90] Cedric R Clapier and Bradley R Cairns. **The biology of chromatin remodeling complexes.** *Annual review of biochemistry*, **78**:273–304, January 2009. 21

[91] Andrew Flaus, David M a Martin, Geoffrey J Barton, and Tom Owen-Hughes. **Identification of multiple distinct Snf2 subfamilies with conserved structural motifs.** *Nucleic acids research*, **34**(10):2887–905, January 2006. 22

[92] Nicolas H Thomä, Bryan K Czyzewski, Andrei a Alexeev, Alexander V Mazin, Stephen C Kowalczykowski, and Nikola P Pavletich. **Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54.** *Nature structural & molecular biology*, **12**(4):350–6, April 2005. 22

[93] Yonatan Savir and Tsvi Tlusty. **RecA-mediated homology search as a nearly optimal signal detection system.** *Molecular cell*, **40**(3):388–96, November 2010. 22

[94] Zhucheng Chen, Haijuan Yang, and Nikola P Pavletich. **Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures.** *Nature*, **453**(7194):489–4, May 2008. 22

[95] Alexandra Lusser, Debra L Urwin, and James T Kadonaga. **Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly.** *Nature structural & molecular biology*, **12**(2):160–6, February 2005. 23

[96] Robert J Sims, Chi-Fu Chen, Helena Santos-Rosa, Tony Kouzarides, Smita S Patel, and Danny Reinberg. **Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains.** *The Journal of biological chemistry*, **280**(51):41789–92, December 2005. 23

[97] John F Flanagan, Li-Zhi Mi, Maksymilian Chruszcz, Marcin Cymborowski, Katrina L Clines, Youngchang Kim, Wladek Minor, Fraydoon Rastinejad, and Sepideh Khorasanizadeh. **Double chromodomains cooperate to recognize the methylated histone H3 tail.** *Nature*, **438**(7071):1181–5, December 2005. 23

[98] Glenn Hauk, Jeffrey N McKnight, Ilana M Nodelman, and Gregory D Bowman. **The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor.** *Molecular cell*, **39**(5):711–23, September 2010. 23, 25, 26, 27

[99] Amit Sharma, Katherine R Jenkins, Annie Héroux, and Gregory D Bowman. **Crystal structure of the chromodomain helicase DNA-binding protein 1 (Chd1) DNA-binding domain in complex with DNA.** *The Journal of biological chemistry*, **286**(49):42099–104, December 2011. 25

[100] Donna M Martin. **Chromatin remodeling in development and disease: focus on CHD7.** *PLoS genetics*, **6**(7):e1001010, July 2010. 26

[101] C N Adra, J L Donato, R Badovinac, F Syed, R Kheraj, H Cai, C Moran, M T Kolker, H Turner, S Weremowicz, T Shirakawa, C C Morton, L E Schnipper, and R Drews.

SMARCAD1, a novel human helicase family-defining member associated with genetic instability: cloning, expression, and mapping to 4q22-q23, a band rich in breakpoints and deletion mutants involved in several human diseases. *Genomics*, **69**(2):162–73, October 2000. 26

[102] M Schoor, K Schuster-Gossler, and a Gossler. **The Etl-1 gene encodes a nuclear protein differentially expressed during early mouse development.** *Developmental dynamics : an official publication of the American Association of Anatomists*, **197**(3):227–37, July 1993. 26

[103] Feng Hong, Fang Fang, Xuming He, Xiaoyi Cao, Hiram Chipperfield, Dan Xie, Wing H Wong, Huck H Ng, and Sheng Zhong. **Dissecting early differentially expressed genes in a mixture of differentiating embryonic stem cells.** *PLoS computational biology*, **5**(12):e1000607, December 2009. 26

[104] Nozomi Tomimatsu, Bipasha Mukherjee, Emmanuelle Martini, Basheer Khadaroo, Haico Van Attikum, and Bertrand Llorente. **The yeast Fun30 and human SMARCAD1 chromatin remodellers promote DNA end resection**. *Nature*, pages 0–5, 2012. 26, 32, 33, 218, 219

[105] Ana Neves-Costa, W Ryan Will, Anna T Vetter, J Ross Miller, and Patrick Varga-Weisz. **The SNF2-family member Fun30 promotes gene silencing in heterochromatic loci.** *PloS one*, **4**(12):e8111, January 2009. 28, 29

[106] Hugo F Azurmendi, Sharmistha Mitra, Iriscilla Ayala, Liwu Li, Carla V Finkielstein, and Daniel G S Capelluto. **Backbone (1)H, (15)N, and (13)C resonance assignments and secondary structure of the Tollip CUE domain.** *Molecules and cells*, **30**(6):581–5, December 2010. 26

[107] William J Belden, Jennifer J Loros, and Jay C Dunlap. **Execution of the circadian negative feedback loop in Neurospora requires the ATP-dependent chromatin-remodeling enzyme CLOCKSWITCH.** *Molecular cell*, **25**(4):587–600, February 2007. 29

[108] Samuel P Rowbotham, Leila Barki, Ana Neves-Costa, Fatima Santos, Wendy Dean, Nicola Hawkes, Parul Choudhary, W Ryan Will, Judith Webster, David Oxley, Catherine M Green, Patrick Varga-Weisz, and Jacqueline E Mermoud. **Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1.** *Molecular cell*, **42**(3):285–96, May 2011. 29, 31, 32, 34, 51, 77, 78, 85, 191, 197, 200, 201, 211, 216, 222

[109] Janna Nousbeck, Bettina Burger, Dana Fuchs-Telem, Mor Pavlovsky, Shlomit Fenig, Ofer Sarig, Peter Itin, and Eli Sprecher. **A Mutation in a Skin-Specific Isoform of SMARCAD1 Causes Autosomal-Dominant Adermatoglyphia.** *American journal of human genetics*, **89**(2):302–307, August 2011. 30

[110] JE Mermoud, Samuel P. Rowbotham, and Patrick D. Varga-Weisz. **Keeping chromatin quiet: How nucleosome remodeling restores heterochromatin after replication**. *Cell cycle*, **10**(23):4017–4025, 2011. 31, 32, 33, 34, 191, 197, 216, 217

[111] Noriko Okazaki, Shun Ikeda, Reiko Ohara, Kiyo Shimada, Toshihide Yanagawa, Takahiro Nagase, Osamu Ohara, and Hisashi Koga. **The novel protein complex with SMARCAD1/KIAA1122 binds to the vicinity of TSS.** *Journal of molecular biology*, **382**(2):257–65, October 2008. 32

[112] Shinya Kamitani, Sumihito Togi, Osamu Ikeda, Misa Nakasuji, Asuka Sakauchi, Yuichi Sekine, Ryuta Muromoto, Kenji Oritani, and Tadashi Matsuda. **Krüppel-associated box-associated protein 1 negatively regulates TNF-α-induced NF-κB transcriptional activity by influencing the interactions among STAT3, p300, and NF-κB/p65.** *Journal of immunology (Baltimore, Md. : 1950)*, **187**(5):2476–83, September 2011. 32

[113] Sushma Iyengar and Peggy J Farnham. **KAP1 Protein: An Enigmatic Master Regulator of the Genome.** *The Journal of biological chemistry*, **286**(30):26267–76, July 2011. 33, 181, 197, 214

[114] Yung-Kang Lee, Stefani N Thomas, Austin J Yang, and David K Ann. **Doxorubicin down-regulates Kruppel-associated box domain-associated protein 1 sumoylation that relieves its transcription repression on p21WAF1/CIP1 in breast cancer MCF-7 cells.** *The Journal of biological chemistry*, **282**(3):1595–606, January 2007. 33

[115] Henri a Blomster, Susumu Y Imanishi, Jenny Siimes, Juha Kastu, Nick a Morrice, John E Eriksson, and Lea Sistonen. **In vivo identification of sumoylation sites by a signature tag and cysteine-targeted affinity purification.** *The Journal of biological chemistry*, **285**(25):19324–9, June 2010. 33

[116] Lei Zeng, Kyoko L Yap, Alexey V Ivanov, Xueqi Wang, Shiraz Mujtaba, Olga Plotnikova, Frank J Rauscher, and Ming-Ming Zhou. **Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing.** *Nature structural & molecular biology*, **15**(6):626–33, June 2008. 33

[117] Shuhei Matsuoka, Bryan a Ballif, Agata Smogorzewska, E Robert McDonald, Kristen E Hurov, Ji Luo, Corey E Bakalarski, Zhenming Zhao, Nicole Solimini, Yaniv Lerenthal, Yosef Shiloh, Steven P Gygi, and Stephen J Elledge. **ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.** *Science (New York, N.Y.)*, **316**(5828):1160–6, May 2007. 33, 215

[118] Xu Li, Yung-Kang Lee, Jen-Chong Jeng, Yun Yen, David C Schultz, Hsiu-Ming Shih, and David K Ann. **Role for KAP1 serine 824 phosphorylation and sumoylation/desumoylation switch in regulating KAP1-mediated transcriptional repression.** *The Journal of biological chemistry*, **282**(50):36177–89, December 2007. 33

[119] E Warbrick, D P Lane, D M Glover, and L S Cox. **A small peptide inhibitor of DNA replication defines the site of interaction between the cyclin-dependent kinase inhibitor p21WAF1 and proliferating cell nuclear antigen.** *Current biology : CB*, **5**(3):275–82, March 1995. 34

[120] Raymond a Poot, Ludmila Bozhenok, Debbie L C van den Berg, Sø ren Steffensen, Fernando Ferreira, Margaret Grimaldi, Nick Gilbert, Joao Ferreira, and Patrick D Varga-Weisz. **The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci.** *Nature cell biology*, **6**(12):1236–44, December 2004. 34

[121] Ann L Kirchmaier. **Ub-family modifications at the replication fork: Regulating PCNA-interacting components.** *FEBS letters*, August 2011. 35, 36, 191

[122] Steven Bergink and Stefan Jentsch. **Principles of ubiquitin and SUMO modifications in DNA repair.** *Nature*, **458**(7237):461–7, March 2009. 35, 181

[123] Ricardo Sánchez, Daniel Torres, Jesús Prieto, Francisco J Blanco, and Ramón Campos-Olivas. **Backbone assignment of human proliferating cell nuclear antigen.** *Biomolecular NMR assignments*, **1**(2):245–7, December 2007. 36, 191

[124] George-Lucian Moldovan, Boris Pfander, and Stefan Jentsch. **PCNA, the maestro of the replication fork.** *Cell*, **129**(4):665–79, May 2007. 36, 181, 191

[125] James H Hurley, Sangho Lee, and Gali Prag. **Ubiquitin-binding domains.** *The Biochemical journal*, **399**(3):361–72, November 2006. 36, 42, 47, 80

[126] a Ciechanover, H Heller, S Elias, a L Haas, and a Hershko. **ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation.** *Proceedings of the National Academy of Sciences of the United States of America*, **77**(3):1365–8, March 1980. 36

[127] I L Goldknopf and H Busch. **Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24.** *Proceedings of the National Academy of Sciences of the United States of America*, **74**(3):864–8, March 1977. 36

[128] L. Gregori, M.S. Poosch, G. Cousins, and V. Chau. **A uniform isopeptide-linked multiubiquitin chain is sufficient to target substrate for degradation in ubiquitin-mediated proteolysis.** *Journal of Biological Chemistry*, **265**:8354–8357, 1990. 36

[129] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P Walter. *Moelcular Biology of the Cell*. Garland Science, New York, fourth edition, 2002. 36, 38

[130] Avram Hershko, Aaron Ciechanover, Hannah Heller, Arthur L Haas, and Irwin A Rose. **Proposed role of ATP in protein breakdown : Conjugation of proteins with multiple chains of the polypeptide of ATP-dependent proteolysis**. *Biochemistry*, **77**(4):1783–1786, 1980. 36

[131] C M Pickart. **Mechanisms underlying ubiquitination.** *Annual review of biochemistry*, **70**:503–33, January 2001. 37

[132] Takayoshi Kirisako, Kiyoko Kamei, Shigeo Murata, Michiko Kato, Hiromi Fukumoto, Masato Kanie, Soichi Sano, Fuminori Tokunaga, Keiji Tanaka, and Kazuhiro Iwai. **A ubiquitin ligase complex assembles linear polyubiquitin chains.** *The EMBO journal*, **25**(20):4877–87, October 2006. 37

[133] Fumiyo Ikeda and Ivan Dikic. **Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series.** *EMBO reports*, **9**(6):536–42, June 2008. 37, 38, 211, 212

[134] S Vijay-Kumar, C E Bugg, and W J Cook. **Structure of ubiquitin refined at 1.8 A resolution.** *Journal of molecular biology*, **194**(3):531–44, April 1987. 37

[135] Avram Hershkos, Hannah Heller, Sarah Elias, and Aaron Ciechanover. **Components of Ubiquitin-Protein Ligase System**. **258**(13):8206–8214, 1983. 38

[136] Joshua D Schnell and Linda Hicke. **Non-traditional functions of ubiquitin and ubiquitin-binding proteins.** *The Journal of biological chemistry*, **278**(38):35857–60, September 2003. 38

[137] Boyko S Atanassov, Evangelia Koutelou, and Sharon Y Dent. **The role of deubiquitinating enzymes in chromatin regulation.** *FEBS letters*, **585**(13):2016–23, July 2011. 38

[138] Francisca E Reyes-Turcu, Karen H Ventii, and Keith D Wilkinson. **Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes.** *Annual review of biochemistry*, **78**:363–97, January 2009. 38

[139] K D Wilkinson, M K Urban, and a L Haas. **Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes.** *The Journal of biological chemistry*, **255**(16):7529–32, August 1980. 38

[140] Koraljka Husnjak, Suzanne Elsasser, Naixia Zhang, Xiang Chen, Leah Randles, Yuan Shi, Kay Hofmann, Kylie J Walters, Daniel Finley, and Ivan Dikic. **Proteasome subunit Rpn13 is a novel ubiquitin receptor.** *Nature*, **453**(7194):481–8, May 2008. 38

[141] Q Deveraux, C Jensen, and M Rechsteiner. **Molecular cloning and expression of a 26 S protease subunit enriched in dileucine repeats.** *The Journal of biological chemistry*, **270**(40):23726–9, October 1995. 38

[142] Christiane Riedinger, Jonas Boehringer, Jean-Francois Trempe, Edward D Lowe, Nicholas R Brown, Kalle Gehring, Martin E M Noble, Colin Gordon, and Jane a Endicott. **Structure of Rpn10 and its interactions with polyubiquitin chains and the proteasome subunit Rpn12.** *The Journal of biological chemistry*, **285**(44):33992–4003, October 2010. 38

[143] Anna von Mikecz. **The nuclear ubiquitin-proteasome system.** *Journal of cell science*, **119**(Pt 10):1977–84, May 2006. 38

[144] J S Thrower, L Hoffman, M Rechsteiner, and C M Pickart. **Recognition of the polyubiquitin proteolytic signal.** *The EMBO journal*, **19**(1):94–102, January 2000. 38

[145] V Chau, J W Tobias, A Bachmair, D Marriott, D J Ecker, D K Gonda, and A Varshavsky. **A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein**. *Science*, **243**(4898):1576–1583, 1989. 38

[146] L Hicke. **Protein regulation by monoubiquitin.** *Nature reviews. Molecular cell biology*, **2**(3):195–201, March 2001. 38

[147] W J Cook, L C Jeffrey, M Carson, Z Chen, and C M Pickart. **Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2).** *The Journal of biological chemistry*, **267**(23):16467–71, August 1992. 39

[148] David Komander, Francisca Reyes-Turcu, Julien D F Licchesi, Peter Odenwaelder, Keith D Wilkinson, and David Barford. **Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains.** *EMBO reports*, **10**(5):466–73, May 2009. 39, 40

[149] Ivan Dikic, Soichi Wakatsuki, and Kylie J Walters. **Ubiquitin-binding domains - from structures to functions.** *Nature reviews. Molecular cell biology*, **10**(10):659–71, October 2009. 38, 40, 42, 45, 47

[150] Ian Hammond-Martel, Helen Yu, and El Bachir Affar. **Roles of ubiquitin signaling in transcription regulation.** *Cellular signalling*, **24**(2):410–421, October 2011. 38

[151] Ivan Martinez-Forero, Ana Rouzaut, Asis Palazon, Juan Dubrot, and Ignacio Melero. **Lysine 63 polyubiquitination in immunotherapy and in cancer-promoting inflammation.** *Clinical cancer research : an official journal of the American Association for Cancer Research*, **15**(22):6751–7, November 2009. 38

[152] W-L Yang, X Zhang, and H-K Lin. **Emerging role of Lys-63 ubiquitination in protein kinase and phosphatase activation and cancer development.** *Oncogene*, **29**(32):4493–503, August 2010. 38

[153] Edward W Harhaj and Vishva M Dixit. **Deubiquitinases in the regulation of NF-κB signaling.** *Cell research*, **21**(1):22–39, January 2011. 40

[154] Kazuhiro Iwai and Fuminori Tokunaga. **Linear polyubiquitination: a new regulator of NF-kappaB activation.** *EMBO reports*, **10**(7):706–13, July 2009. 40

[155] Alexis Verger, José Perdomo, and Merlin Crossley. **Modification with SUMO. A role in transcriptional regulation.** *EMBO reports*, **4**(2):137–42, February 2003. 40

[156] Ronald T Hay. **SUMO: a history of modification.** *Molecular cell*, **18**(1):1–12, April 2005. 40

[157] Alfred C O Vertegaal. **Uncovering ubiquitin and ubiquitin-like signaling networks.** *Chemical reviews*, **111**(12):7923–40, December 2011. 40

[158] S Vijay-Kumar, C E Bugg, and W J Cook. **Structure of ubiquitin refined at 1.8 A resolution.** *Journal of Molecular Biology*, **194**(3):531–544, 1987. 41

[159] P Bayer, a Arndt, S Metzger, R Mahajan, F Melchior, R Jaenicke, and J Becker. **Structure determination of the small ubiquitin-related modifier SUMO-1.** *Journal of molecular biology*, **280**(2):275–86, July 1998. 41

[160] Yun-Seok Choi, Young Ho Jeon, Kyoung-Seok Ryu, and Chaejoon Cheong. **60th residues of ubiquitin and Nedd8 are located out of E2-binding surfaces, but are important for K48 ubiquitin-linkage.** *FEBS letters*, **583**(20):3323–8, October 2009. 41

[161] TheUniprotConsortium. **Reorganizing the protein space at the Universal Protein Resource (UniProt).** *Nucleic acids research*, **40**(Database issue):D71–5, January 2012. 41, 206

[162] Mickael Goujon, Hamish McWilliam, Weizhong Li, Franck Valentin, Silvano Squizzato, Juri Paern, and Rodrigo Lopez. **A new bioinformatics analysis tools framework at EMBL-EBI.** *Nucleic acids research*, **38**(Web Server issue):W695–9, July 2010. 41

[163] M a Larkin, G Blackshields, N P Brown, R Chenna, P a McGettigan, H McWilliam, F Valentin, I M Wallace, a Wilm, R Lopez, J D Thompson, T J Gibson, and D G Higgins. **Clustal W and Clustal X version 2.0.** *Bioinformatics (Oxford, England)*, **23**(21):2947–8, November 2007. 41, 206

[164] Andrew M Waterhouse, James B Procter, David M a Martin, Michèle Clamp, and Geoffrey J Barton. **Jalview Version 2–a multiple sequence alignment editor and analysis workbench.** *Bioinformatics (Oxford, England)*, **25**(9):1189–91, May 2009. 41

[165] CP Ponting. **Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction**. *Biochemical Journal*, **535**:527–535, 2000. 42

[166] Christopher D Lima. **CUE'd up for Monoubiquitin.** *Cell*, **113**(5):554–556, 2003. 42, 43

[167] Richard S Kang, Cynthia M Daniels, Smitha A Francis, Susan C Shih, William J Salerno, Linda Hicke, and Ishwar Radhakrishnan. **Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding.** *Cell*, **113**(5):621–630, 2003. 42, 43, 82, 86, 115, 149, 182, 183, 213

[168] Linda Hicke, Heidi L Schubert, and Christopher P Hill. **Ubiquitin-binding domains.** *Nature reviews. Molecular cell biology*, **6**(8):610–21, August 2005. 42, 47, 181

[169] Gali Prag, Saurav Misra, Eudora A Jones, Rodolfo Ghirlando, Brian A Davies, Bruce F Horazdovsky, and James H Hurley. **Mechanism of ubiquitin recognition by the CUE domain of Vps9p.** *Cell*, **113**(5):609–620, 2003. 42, 43, 114, 182

[170] Susan C Shih, Gali Prag, Smitha A Francis, Myra A Sutanto, James H Hurley, and Linda Hicke. **A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain.** *EMBO Journal*, **22**(6):1273–1281, 2003. 42, 43

[171] Alexey V Ivanov, Hongzhuang Peng, Vyacheslav Yurchenko, Kyoko L Yap, Dmitri G Negorev, David C Schultz, Elyse Psulkowski, William J Fredericks, David E White, Gerd G Maul, Moshe J Sadofsky, Ming-Ming Zhou, and Frank J Rauscher. **PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing.** *Molecular Cell*, **28**(5):823–837, 2007. 42

[172] Thomas P Garner, Joanna Strachan, Elizabeth C Shedden, Jed E Long, James R Cavey, Barry Shaw, Robert Layfield, and Mark S Searle. **Independent interactions of ubiquitin-binding domains in a ubiquitin-mediated ternary complex.** *Biochemistry*, **50**(42):9076–87, October 2011. 44, 45

[173] Simin Rahighi and Ivan Dikic. **Selectivity of the ubiquitin-binding modules.** *FEBS letters*, **586**(17):2705–2710, May 2012. 44, 47

[174] Michael French, Kurt Swanson, Susan C Shih, Ishwar Radhakrishnan, and Linda Hicke. **Identification and characterization of modular domains that bind ubiquitin.** *Methods in enzymology*, **399**(05):135–57, January 2005. 44, 172

[175] Gali Prag, Sangho Lee, Rafael Mattera, Cecilia N Arighi, Bridgette M Beach, Juan S Bonifacino, and James H Hurley. **Structural mechanism for ubiquitinated-cargo recognition by the Golgi-localized, gamma-ear-containing, ADP-ribosylation-factor-binding proteins.** *Proceedings of the National Academy of Sciences of the United States of America*, **102**(7):2334–9, March 2005. 44, 45, 183

[176] J Wade Harper and Brenda a Schulman. **Structural complexity in ubiquitin recognition.** *Cell*, **124**(6):1133–6, March 2006. 45

[177] Daoning Zhang, Shahri Raasi, and David Fushman. **Affinity makes the difference: nonselective interaction of the UBA domain of Ubiquilin-1 with monomeric ubiquitin and polyubiquitin chains.** *Journal of molecular biology*, **377**(1):162–80, March 2008. 45

[178] Satoshi Hirano, Nobuhiro Suzuki, Thomas Slagsvold, Masato Kawasaki, Daniel Trambaiolo, Ryuichi Kato, Harald Stenmark, and Soichi Wakatsuki. **Structural basis of ubiquitin recognition by mammalian Eap45 GLUE domain.** *Nature structural & molecular biology*, **13**(11):1031–2, November 2006. 45

[179] Peter S Brzovic, Alexei Lissounov, Devin E Christensen, David W Hoyt, and Rachel E Klevit. **A UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination.** *Molecular cell*, **21**(6):873–80, March 2006. 45

[180] Francisca E Reyes-Turcu, John R Horton, James E Mullally, Annie Heroux, Xiaodong Cheng, and Keith D Wilkinson. **The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin.** *Cell*, **124**(6):1197–208, March 2006. 46

[181] Sangho Lee, Yien Che Tsai, Rafael Mattera, William J Smith, Michael S Kostelansky, Allan M Weissman, Juan S Bonifacino, and James H Hurley. **Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5.** *Nature structural & molecular biology*, **13**(3):264–71, March 2006. 46, 183

[182] Yusuke Sato, Azusa Yoshikawa, Masami Yamashita, Atsushi Yamagata, and Shuya Fukai. **Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by NZF domains of TAB2 and TAB3.** *The EMBO journal*, **28**(24):3903–9, December 2009. 47

[183] Yogesh Kulathu, Masato Akutsu, Anja Bremm, Kay Hofmann, and David Komander. **Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain.** *Nature structural & molecular biology*, **16**(12):1328–30, December 2009. 47

[184] Yusuke Sato, Azusa Yoshikawa, Hisatoshi Mimura, Masami Yamashita, Atsushi Yamagata, and Shuya Fukai. **Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80.** *The EMBO journal*, **28**(16):2461–8, August 2009. 47, 183

[185] J Sambrook and D W Russell. *Molecular cloning: a laboratory manual. 3rd ed.* Cold Spring Harbor Laboratory Press, 2001. 50

[186] R. Novy, D. Drott, K. Yaeger, and R. Mierendorf. **Overcoming the codon bias of E. coli for enhanced protein expression**. *inNovations*, **12**:1–3, 2001. 52

[187] F Sanger and a R Coulson. **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.** *Journal of molecular biology*, **94**(3):441–8, May 1975. 53

[188] G. Lu and E.N. Moriyama. **Vector NTI, a balanced all-in-one sequence analysis suite**. *Briefings in bioinformatics*, **5**(4):378–388, 2004. 53

[189] Doryen Bubeck, Martin a M Reijns, Stephen C Graham, Katy R Astell, E Yvonne Jones, and Andrew P Jackson. **PCNA directs type 2 RNase H activity on DNA replication and repair substrates.** *Nucleic acids research*, **39**(9):3652–66, May 2011. 64

[190] Thomas S Walter, Jonathan M Diprose, Chris J Mayo, Christian Siebold, Mike G Pickford, Lester Carter, Geoff C Sutton, Nick S Berrow, James Brown, Ian M Berry, Guillaume B E Stewart-Jones, Jonathan M Grimes, David K Stammers, Robert M Esnouf, E Yvonne Jones, Ray J Owens, David I Stuart, and Karl Harlos. **A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization.** *Acta crystallographica. Section D, Biological crystallography*, **61**(Pt 6):651–7, June 2005. 66

[191] Ian M Berry, O Dym, R M Esnouf, K Harlos, R Meged, a Perrakis, J L Sussman, T S Walter, J Wilson, and Albrecht Messerschmidt. **SPINE high-throughput crystallization, crystal imaging and recognition techniques: current state, performance analysis, new technologies and future aspects.** *Acta crystallographica. Section D, Biological crystallography*, **62**(Pt 10):1137–49, October 2006. 66

[192] Thomas S Walter, Christoph Meier, Rene Assenberg, Kin-Fai Au, Jingshan Ren, Anil Verma, Joanne E Nettleship, Raymond J Owens, David I Stuart, and Jonathan M Grimes. **Lysine methylation as a routine rescue strategy for protein crystallization.** *Structure (London, England : 1993)*, **14**(11):1617–22, November 2006. 66, 90

[193] F Delaglio, S Grzesiek, G W Vuister, G Zhu, J Pfeifer, and a Bax. **NMRPipe: a multidimensional spectral processing system based on UNIX pipes.** *Journal of biomolecular NMR*, **6**(3):277–93, November 1995. 68

[194] Wim F Vranken, Wayne Boucher, Tim J Stevens, Rasmus H Fogh, Anne Pajon, Miguel Llinas, Eldon L Ulrich, John L Markley, John Ionides, and Ernest D Laue. **The CCPN data model for NMR spectroscopy: development of a software pipeline.** *Proteins*, **59**(4):687–96, June 2005. 68, 72

[195] Markus Rückert and Gottfried Otting. **Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments**. *Journal of the American Chemical Society*, (14):7793–7797, 2000. 68, 250

[196] A. Wang, S. Grzesiek, and A. Bax. **15N, 13C HSQC ubiquitin spectrum**. 72, 168, 169

[197] Grace Development Team. **Grace (Software)**, 2008. 72, 190

[198] Sjoerd J. de Vries, Aalt D. J. van Dijk, Mickael Krzeminski, Mark van Dijk, Aurelien Thureau, Victor Hsu, Tsjerk Wassenaar, and Alexandre M. J. J. Bonvin. **HADDOCK versus HADDOCK: new features and performance of HADDOCK2. 0 on the CAPRI targets**. *Proteins: Structure, ...*, pages 726–733, 2007. 72

[199] Sjoerd J de Vries, Marc van Dijk, and Alexandre M J J Bonvin. **The HADDOCK web server for data-driven biomolecular docking.** *Nature protocols*, **5**(5):883–97, May 2010. 72, 179, 180

[200] Yinshan Yang, Daniel Auguin, Stéphane Delbecq, Emilie Dumas, Gérard Molle, Virginie Molle, Christian Roumestand, and Nathalie Saint. **Structure of the Mycobacterium tuberculosis OmpATb protein: a model of an oligomeric channel in the mycobacterial cell wall.** *Proteins*, **79**(2):645–61, February 2011. 72

[201] C Schwieters, J Kuszewski, and G Mariusclore. **Using XplorNIH for NMR molecular structure determination**. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **48**(1):47–62, March 2006. 74, 254

[202] Charles D Schwieters, John J Kuszewski, Nico Tjandra, and G Marius Clore. **The Xplor-NIH NMR molecular structure determination package.** *Journal of magnetic resonance (San Diego, Calif. : 1997)*, **160**(1):65–73, January 2003. 74, 254

[203] J Kuszewski, a M Gronenborn, and G M Clore. **Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids.** *Journal of magnetic resonance (San Diego, Calif. : 1997)*, **125**(1):171–7, March 1997. 74

[204] Yang Shen, Frank Delaglio, Gabriel Cornilescu, and Ad Bax. **TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.** *Journal of biomolecular NMR*, **44**(4):213–23, August 2009. 74, 125

[205] Mark V Berjanskii and David S Wishart. **A simple method to predict protein flexibility using secondary chemical shifts.** *Journal of the American Chemical Society*, **127**(43):14970–1, November 2005. 74, 126

[206] Douglas L Theobald and Deborah S Wuttke. **Accurate structural correlations from maximum likelihood superpositions.** *PLoS computational biology*, **4**(2):e43, February 2008. 76

[207] Douglas L Theobald and Phillip a Steindel. **Optimal simultaneous superpositioning of multiple structures with missing data.** *Bioinformatics (Oxford, England)*, **28**(15):1972–9, August 2012. 76

[208] Aneerban Bhattacharya, Roberto Tejero, and Gaetano T Montelione. **Evaluating Protein Structures Determined by Structural Genomics Consortia Tools for Structure Quality Evaluation**. *PROTEINS: Structure, Function, and Bioinformatics*, **66**:778–795, 2007. 76, 144, 157, 161, 256

[209] R. a. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. **PROCHECK: a program to check the stereochemical quality of protein structures**. *Journal of Applied Crystallography*, **26**(2):283–291, April 1993. 76

[210] SC Lovell and IW Davis. **Structure Validation by CA Geometry: Phi, Psi and CB Deviation**. *Proteins: Structure, ...*, **450**(August 2002):437–450, 2003. 76, 144, 157, 161

[211] R Liithy, JU Bowie, and D Eisenberg. **Assessment of Protein Models with Three-dimensional Profiles**. *Nature*, 1992. 76

[212] M J Sippl. **Recognition of errors in three-dimensional structures of proteins.** *Proteins*, **17**(4):355–62, December 1993. 76

[213] Erica Cavellán, Patrik Asp, Piergiorgio Percipalle, and Ann-Kristin Ostlund Farrants. **The WSTF-SNF2h chromatin remodeling complex interacts with several nuclear proteins in transcription.** *The Journal of biological chemistry*, **281**(24):16264–71, July 2006. 78

[214] S Mujtaba, L Zeng, and M-M Zhou. **Structure and acetyl-lysine recognition of the bromodomain.** *Oncogene*, **26**(37):5521–7, August 2007. 79

[215] Joel C Eissenberg. **Structural biology of the chromodomain: Form and function.** *Gene*, pages 1–11, January 2012. 79

[216] Weiwei Dang and Blaine Bartholomew. **Domain architecture of the catalytic subunit in the ISW2-nucleosome complex.** *Molecular and cellular biology*, **27**(23):8306–17, December 2007. 79

[217] Fabian Erdel, Jana Krug, Gernot Längst, and Karsten Rippe. **Targeting chromatin remodelers: Signals and search mechanisms.** *Biochimica et biophysica acta*, **1809**(9):497–508, June 2011. 80

[218] Jianghong Man and Xuemin Zhang. **CUEDC2: an emerging key player in inflammation and tumorigenesis.** *Protein & cell*, **2**(9):699–703, September 2011. 80

[219] Marco Punta, Penny C Coggill, Ruth Y Eberhardt, Jaina Mistry, John Tate, Chris Boursnell, Ningze Pang, Kristoffer Forslund, Goran Ceric, Jody Clements, Andreas Heger, Liisa Holm, Erik L L Sonnhammer, Sean R Eddy, Alex Bateman, and Robert D Finn. **The Pfam protein families database.** *Nucleic acids research*, **40**(November 2011):290–301, November 2011. 81

[220] A Gattiker, E Gasteiger, and A Bairoch. **ScanProsite: a reference implementation of a PROSITE scanning tool**. *Applied Bioinformatics*, **1**(2):107–108, 2002. 81

[221] Christian J a Sigrist, Lorenzo Cerutti, Edouard de Castro, Petra S Langendijk-Genevaux, Virginie Bulliard, Amos Bairoch, and Nicolas Hulo. **PROSITE, a protein domain database for functional characterization and annotation.** *Nucleic acids research*, **38**(Database issue):D161–6, January 2010. 81, 206, 212

[222] Evgeni M Zdobnov and Rolf Apweiler. **InterProScan an integration platform for the signature-recognition methods in InterProrecognition methods in InterPro**. *Bioinformatics*, **17**(9):847–848, 2001. 81

[223] G.E. Crooks, Gary Hon, J.M. Chandonia, and S.E. Brenner. **WebLogo: a sequence logo generator**. *Genome research*, **14**(6):1188–1190, 2004. 81

[224] Marco Pagni, Vassilios Ioannidis, Lorenzo Cerutti, Monique Zahn-Zabal, C Victor Jongeneel, Jörg Hau, Olivier Martin, Dmitri Kuznetsov, and Laurent Falquet. **MyHits: improvements to an interactive resource for analyzing protein sequences.** *Nucleic acids research*, **35**(Web Server issue):W433–7, July 2007. 82

[225] D W a Buchan, S M Ward, a E Lobley, T C O Nugent, K Bryson, and D T Jones. **Protein annotation and modelling servers at University College London.** *Nucleic acids research*, **38**(Web Server issue):W563–8, July 2010. 83

[226] D T Jones. **Protein secondary structure prediction based on position-specific scoring matrices.** *Journal of molecular biology*, **292**(2):195–202, September 1999. 83

[227] Christian Cole, Jonathan D Barber, and Geoffrey J Barton. **The Jpred 3 secondary structure prediction server.** *Nucleic acids research*, **36**(Web Server issue):W197–201, July 2008. 83

[228] Christopher J Oldfield, Eldon L Ulrich, Yugong Cheng, a Keith Dunker, and John L Markley. **Addressing the intrinsic disorder bottleneck in structural proteomics.** *Proteins*, **59**(3):444–53, May 2005. 84

[229] Zheng Rong Yang, Rebecca Thomson, Philip McNeil, and Robert M Esnouf. **RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins.** *Bioinformatics (Oxford, England)*, **21**(16):3369–76, August 2005. 84

[230] J J Ward, J S Sodhi, L J McGuffin, B F Buxton, and D T Jones. **Prediction and functional analysis of native disorder in proteins from the three kingdoms of life.** *Journal of molecular biology*, **337**(3):635–45, March 2004. 84

[231] Takashi Ishida and Kengo Kinoshita. **PrDOS: prediction of disordered protein regions from amino acid sequence.** *Nucleic acids research*, **35**(Web Server issue):W460–4, July 2007. 84

[232] Kurt Wüthrich. **NMR studies of structure and function of biological macromolecules (Nobel Lecture).** *Journal of biomolecular NMR*, **27**(1):13–39, September 2003. 95, 243

[233] H Jane Dyson and Peter E Wright. **Unfolded proteins and protein folding studied by NMR.** *Chemical reviews*, **104**(8):3607–22, August 2004. 95

[234] Manuel Rico. **Protein structure , dynamics and function by NMR**. *Chemical Communications*, **48**(21):1–10, 2010. 96

[235] a E Ferentz and G Wagner. **NMR spectroscopy: a multifaceted approach to macromolecular structure.** *Quarterly reviews of biophysics*, **33**(1):29–65, February 2000. 97, 246

[236] John Cavanagh, Wayne J Fairbrother, Arthur G. Palmer III, Nicholas J. Skelton, and Mark Rance. *Protein NMR Spectroscopy: Principles and Practice*, **2nd**. Academic Press, 2nd edition, 2007. 99, 104, 106, 109, 111, 132, 194, 236, 237, 242, 243, 244, 245, 249, 250

[237] G Bodenhausen and D. Ruben. **Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy**. *Chemical Physics Letters*, **69**(1), 1980. 99, 244, 245

[238] K Wüthrich. *NMR of Proteins and Nucleic Acids*, **32**. Wiley, 1986. 102

[239] Lewis E Kay, Mitsuhiko Ikura, Rolf Tschudin, and Ad Bax. **Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins.** *Journal of magnetic resonance*, **89**:496–514, December 1990. 102, 104, 106

[240] Stephan Gfczesiek, Mitsuhiko Ikura, and G Marius Clore. **A 3D Triple-Resonance NMR Technique for Qualitative Measurement of Carbonyl-H / 3 J Couplings in Isotopically Enriched Proteins**. *Journal of Magnetic Resonance*, **1**:215–221, 1992. 104, 106, 109

[241] Stephan Grzesiek. **Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein**. *Journal of magnetic resonance*, **96**(2):432–440, February 1992. 104, 106, 117, 118

[242] R T Clubb, V Thanabal, and G Wagner. **A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme To Correlate Intraresidue H-1(N), N-15, And C-13(') Chemical-Shifts In N-15-C-13-Labeled Proteins**. *Journal of Magnetic Resonance*, **97**(1):213–217, 1992. 109

[243] Robert T Clubb. **A Constant-Time Three-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue 1HN, 15N, and 13C Chemical Shifts in 15N- 13C-Labeled Proteins**. *Biological Chemistry*, pages 213–217, 1992. 109

[244] Jian Yao, H.Jane Dyson, and Peter E Wright. **Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins**. *FEBS Letters*, **419**(2-3):285–289, December 1997. 113

[245] Iain Campbell. *Biophysical Techniques*. Oxford Univ Press, Oxford, 2012. 115, 130, 132, 133, 237, 238, 240, 241, 242, 244, 245, 246, 247, 248, 249

[246] Haiyan Zhang, Stephen Neal, and David S Wishart. **RefDB: a database of uniformly referenced protein chemical shifts.** *Journal of biomolecular NMR*, **25**(3):173–95, March 2003. 117

[247] Victoria A. Higman. **Protein NMR - A Practical Guide**. *http://www.protein-nmr.org.uk (accessed May 2012)*. 118, 121

[248] D Marion, P C Driscoll, L E Kay, P T Wingfield, a Bax, a M Gronenborn, and G M Clore. **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** . *Biochemistry*, **28**(15):6150–6, July 1989. 118

[249] A Bax, G M Clore, and A M Gronenborn. **1H- 1H Correlation via Isotropic Mixing of 13C a New Three-Dimensional Approach for Assigning 1H and 13C Spectra of 13C-Enriched Proteins**. *Journal of Magnetic Resonance*, **88**(2):425–431, 1990. 121

[250] Edward T Olejniczak, Robert X Xu, and Stephen W Fesik. **A 4D HCCH-TOCSY experiment for assigning the side chain1H and13C resonances of proteins**. *Journal of Biomolecular NMR*, **2**:655–659, 1992. 121

[251] James Keeler. *Understanding NMR Spectroscopy*, **1**. Wiley, 2005. 123, 237, 238, 239, 240, 248

[252] Tobias S Ulmer, Benjamin E Ramirez, Frank Delaglio, and Ad Bax. **Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy**. *Journal of the American Chemical Society*, **125**(30):9179–91, July 2003. 124

[253] J S Richardson. *The anatomy and taxonomy of protein structure.*, **34**. January 1981. 125, 128

[254] C C McDonald and W D Phillips. **Manifestations of the tertiary structures of proteins in high-frequency nuclear magnetic resonance.** *Journal of the American Chemical Society*, **89**(24):6332–41, November 1967. 125

[255] Ming-Sin Cheung, Mahon L Maguire, Tim J Stevens, and R William Broadhurst. **DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure.** *Journal of magnetic resonance (San Diego, Calif. : 1997)*, **202**(2):223–33, February 2010. 125, 127, 128

[256] Yang Shen and Ad Bax. **Identification of helix capping and β-turn motifs from NMR chemical shifts.** *Journal of biomolecular NMR*, pages 211–232, February 2012. 127, 129

[257] D. Eisenberg, R. M. Weiss, and T. C. Terwilliger. **The helical hydrophobic moment: a measure of the amphiphilicity of a helix**. *Nature*, **299**:371–374, 1982. 131

[258] Don Armstrong and Raphael Zidovetzki. **Helical Wheel Script**. *http://rzlab.ucr.edu/scripts/wheel/wheel.cgi*. 131

[259] Lewis E Kay, Dennis A Torchia, and Ad Bax. **Backbone Dynamics of Proteins As Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy : Application to Staphylococcal Nuclease ?** *Analysis*, pages 8972–8979, 1989. 132

[260] Quincy Teng. *Structural Biology: Practical NMR Applications*, **1**. Springer Science, 2005. 132

[261] Rieko Ishima and Dennis A Torchia. **Protein dynamics from NMR**. *Nature structural biology*, **7**(9):7–10, 2000. 132, 248

[262] Lishan Yao, Jinfa Ying, and Ad Bax. **Improved accuracy of 15N-1H scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins.** *Journal of biomolecular NMR*, **43**(3):161–70, March 2009. 132

[263] Rebecca S Lipsitz and Nico Tjandra. **Residual dipolar couplings in NMR structure analysis.** *Annual review of biophysics and biomolecular structure*, **33**:387–413, January 2004. 136, 252

[264] G M Clore, a M Gronenborn, and a Bax. **A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information.** *Journal of magnetic resonance (San Diego, Calif. : 1997)*, **133**(1):216–21, July 1998. 137, 252, 253

[265] Christina Redfield. **Homology-based Determination of Da and R (Personal Communication)**. 2012. 138

[266] MP Williamson. **Applications of the NOE in Molecular Biology**. *Annual Reports on NMR Spectroscopy*, **65**(08):77–109, 2009. 138

[267] C. Zhao, T. Kigawa, M. Sato, S. Koshiba, M. Inoue, and S. Yokoyama. **Solution Structure of the CUE Domain in the Human CUE Domain Containing Protein 1 (CUEDC1).** *Unpublished; PDB: 2DHY*, 2006. 149

[268] S P Mielke and V V Krishnan. **An evaluation of chemical shift index-based secondary structure determination in proteins: influence of random coil chemical shifts.** *Journal of biomolecular NMR*, **30**(2):143–53, October 2004. 153

[269] L Fielding. **NMR methods for the determination of proteinligand dissociation constants**. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **51**(4):219–242, November 2007. 166, 178

[270] Filipp Frank, Nahum Sonenberg, and Bhushan Nagar. **Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2.** *Nature*, **465**(7299):818–22, June 2010. 171

[271] Gabriel Merino, Thomas Heine, and Gotthard Seifert. **The induced magnetic field in cyclic molecules.** *Chemistry (Weinheim an der Bergstrasse, Germany)*, **10**(17):4367–71, September 2004. 173, 179

[272] Nicholas C. Price, Raymond A. Dwek, R. George Ratcliffe, and Mark R. Wormald. *Principles and Problems in Physical Chemistry.* Oxford Univ Press, 2001. 175

[273] Julia Vaynberg and Jun Qin. **Weak protein-protein interactions as probed by NMR spectroscopy.** *Trends in biotechnology*, **24**(1):22–7, January 2006. 178

[274] Ayako Ohno, JunGoo Jee, Kenichiro Fujiwara, Takeshi Tenno, Natsuko Goda, Hidehito Tochio, Hideki Kobayashi, Hidekazu Hiroaki, and Masahiro Shirakawa. **Structure of the UBA domain of Dsk2p in complex with ubiquitin molecular determinants for ubiquitin recognition.** *Structure (London, England : 1993)*, **13**(4):521–32, April 2005. 183

[275] Kurt a Swanson, Linda Hicke, and Ishwar Radhakrishnan. **Structural basis for monoubiquitin recognition by the Ede1 UBA domain.** *Journal of molecular biology*, **358**(3):713–24, May 2006. 183

[276] Kurt a Swanson, Richard S Kang, Svetoslava D Stamenova, Linda Hicke, and Ishwar Radhakrishnan. **Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation.** *The EMBO journal*, **22**(18):4597–606, September 2003. 183

[277] Satoshi Hirano, Masato Kawasaki, Hideaki Ura, Ryuichi Kato, Camilla Raiborg, Harald Stenmark, and Soichi Wakatsuki. **Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting.** *Nature structural & molecular biology*, **13**(3):272–7, March 2006. 183

[278] Grace Gill. **SUMO and ubiquitin in the nucleus: different functions, similar mechanisms?** *Genes & development*, **18**(17):2046–59, September 2004. 181

[279] Jing Song, Linda K Durrin, Thomas a Wilkinson, Theodore G Krontiris, and Yuan Chen. **Identification of a SUMO-binding motif that recognizes SUMO-modified proteins.** *Proceedings of the National Academy of Sciences of the United States of America*, **101**(40):14373–8, October 2004. 188

[280] Jing Song, Ziming Zhang, Weidong Hu, and Yuan Chen. **Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation.** *The Journal of biological chemistry*, **280**(48):40122–9, December 2005. 188, 212

[281] Candace S. Seu and Yuan Chen. **Identification of SUMO Binding Motifs by NMR**. *Methods Mol Biol.*, **497**(6):1–17, 2009. 194, 212

[282] X Li, HH Lin, H Chen, and X Xu. **SUMOylation of the Transcriptional Corepressor KAP1 is Regulated by the Serine and Threonine Phosphatase PP1**. *Science Signaling*, **3**(119):1–22, 2010. 197

[283] Qiming Liang, Hongying Deng, Xiaojuan Li, Xianfang Wu, Qiyi Tang, Tsung-Hsien Chang, Hongzhuang Peng, Frank J Rauscher, Keiko Ozato, and Fanxiu Zhu. **Tripartite motif-containing protein 28 is a small ubiquitin-related modifier E3 ligase and negative regulator of IFN regulatory factor 7.** *Journal of immunology*, **187**(9):4754–63, December 2011. 197

[284] Woong Kim, Eric J Bennett, Edward L Huttlin, Ailan Guo, Jing Li, Anthony Possemato, Mathew E Sowa, Ramin Rad, John Rush, Michael J Comb, J Wade Harper, and Steven P Gygi. **Systematic and quantitative assessment of the ubiquitin-modified proteome.** *Molecular cell*, **44**(2):325–40, October 2011. 201

[285] Patrice Gouet, Emmanuel Courcelle, and DI Stuart. **ESPript : analysis of multiple sequence alignments in PostScript**. *Bioinformatics*, **15**(4):305–308, 1999. 206, 208

[286] P. Gouet. **ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins**. *Nucleic Acids Research*, **31**(13):3320–3323, July 2003. 206, 208

[287] Yuji Chikashige and Yasushi Hiraoka. **[Genome sequence of Schizosaccharomyces pombe].** *Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme*, **47**(9):1215–20, July 2002. 210

[288] Jennifer M Doyle, Jinlan Gao, Jiawei Wang, Maojun Yang, and Patrick Ryan Potts. **MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases.** *Molecular cell*, **39**(6):963–74, September 2010. 214

[289] Kendra L Cann and Graham Dellaire. **Heterochromatin and the DNA damage response: the need to relax.** *Biochemistry and cell biology = Biochimie et biologie cellulaire*, **89**(1):45–60, February 2011. 214

[290] Yael Ziv, Dana Bielopolski, Yaron Galanty, Claudia Lukas, Yoichi Taya, David C Schultz, Jiri Lukas, Simon Bekker-Jensen, Jiri Bartek, and Yosef Shiloh. **Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway.** *Nature cell biology*, **8**(8):870–6, August 2006. 214

[291] Aaron a Goodarzi, Angela T Noon, Dorothee Deckbar, Yael Ziv, Yosef Shiloh, Markus Löbrich, and Penny a Jeggo. **ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin.** *Molecular cell*, **31**(2):167–77, July 2008. 214

[292] Sean a Beausoleil, Mark Jedrychowski, Daniel Schwartz, Joshua E Elias, Judit Villén, Jiaxu Li, Martin a Cohn, Lewis C Cantley, and Steven P Gygi. **Large-scale characterization of HeLa cell nuclear phosphoproteins.** *Proceedings of the National Academy of Sciences of the United States of America*, **101**(33):12130–5, August 2004. 215

[293] Jesper V Olsen, Blagoy Blagoev, Florian Gnad, Boris Macek, Chanchal Kumar, Peter Mortensen, and Matthias Mann. **Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.** *Cell*, **127**(3):635–48, November 2006. 215

[294] Makoto Tachibana, Yasuko Matsumura, Mikiko Fukuda, Hiroshi Kimura, and Yoichi Shinkai. **G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription.** *The EMBO journal*, **27**(20):2681–90, October 2008. 216

[295] Chen Hu, Shengping Zhang, Xuan Gao, Xiaojing Gao, Xiaohong Xu, Ya Lv, Yan Zhang, Zhenhong Zhu, Changqing Zhang, Qiao Li, Jiemin Wong, Yongping Cui, Wen Zhang, Lin Ma, and Chuangui Wang. **Roles of Kruppel-associated Box (KRAB)-associated Co-repressor KAP1 Ser-473 Phosphorylation in DNA Damage Response.** *The Journal of biological chemistry*, **287**(23):18937–52, June 2012. 219

[296] David White, Ilona U Rafalska-Metcalf, Alexey V Ivanov, Andrea Corsinotti, Hongzhuang Peng, Sheng-Chung Lee, Didier Trono, Susan M Janicki, and Frank J Rauscher. **The ATM substrate KAP1 controls DNA repair in heterochromatin: regulation by HP1 proteins and serine 473/824 phosphorylation.** *Molecular cancer research : MCR*, **10**(3):401–14, March 2012. 219

[297] Zhu Zhu, Woo-Hyun Chung, Eun Yong Shim, Sang Eun Lee, and Grzegorz Ira. **Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends.** *Cell*, **134**(6):981–94, September 2008. 219

[298] EL Hahn. **Spin Echoes**. *Physical Review*, **297**(1946), 1950. 240

[299] D a MacKenzie, J a Spencer, M F Le Gal-Coëffet, and D B Archer. **Efficient production from Aspergillus niger of a heterologous protein and an individual protein domain, heavy isotope-labelled, for structure-function analysis.** *Journal of biotechnology*, **46**(2):85–93, April 1996. 243

[300] Christina Redfield. **NMR Primer (Course Handout)**. *Department of Biochemistry, University of Oxford*, 2009. 246, 252, 254, 255

[301] Christina Redfield. **Advanced NMR (Course Handout)**. *Department of Biochemistry, University of Oxford*, 2009. 246, 252, 254, 255

[302] W. a. Anderson and R. Freeman. **Influence of a Second Radiofrequency Field on High-Resolution Nuclear Magnetic Resonance Spectra**. *The Journal of Chemical Physics*, **37**(1):85, 1962. 246

[303] J H Prestegard, C M Bougault, and a I Kishore. **Residual dipolar couplings in structure determination of biomolecules.** *Chemical reviews*, **104**(8):3519–40, August 2004. 249

[304] Maximillian H Bailor, Catherine Musselman, Alexandar L Hansen, Kush Gulati, Dinshaw J Patel, and Hashim M Al-Hashimi. **Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings.** *Nature protocols*, **2**(6):1536–46, January 2007. 249

[305] J J Chou, S Gaemers, B Howder, J M Louis, and a Bax. **A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles.** *Journal of biomolecular NMR*, **21**(4):377–82, December 2001. 250

[306] a T Brünger, P D Adams, and L M Rice. **New applications of simulated annealing in X-ray crystallography and solution NMR.** *Structure (London, England : 1993)*, **5**(3):325–36, March 1997. 254

[307] M Nilges. **Structure calculation from NMR data.** *Current opinion in structural biology*, **6**(5):617–23, October 1996. 254

Statement of Authorship

This thesis is approximately 35,000 words long; exclusive of references, figures and tables.

The research presented here and the thesis itself are the work of me, Philip West, with the following exceptions:

Some of the work to clone the dCUE-long and dCUE-short constructs was carried out in conjunction with undergraduate project student Christopher Anderton.

Multiangle light scattering (MALS) analysis was carried out with help from Geoff Sutton.

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



Philip West

October 2012