

Understanding substrate recruitment by BTB-Kelch family E3 ligases



Zhuoyao Chen

Nuffield Department of Medicine & Oriel College

University of Oxford

Trinity term 2019

Thesis submitted for the degree of Doctor of Philosophy in Clinical Medicine

Understanding substrate recruitment by BTB-Kelch family E3 ligases.

Zhuoyao Chen. Oriel College.

Thesis submitted for the degree of Doctor of Philosophy in Clinical Medicine, Nuffield Department of Medicine. Trinity term 2019.

Abstract

Over 600 E3 ligases are identified that covalently modify substrates with ubiquitin to control their degradation, trafficking or activity. Many are linked to human diseases and represent an important class of drug targets. An understanding of their substrate recruitment mechanisms is necessary to elucidate their functions and to enable drug design. BTB-Kelch proteins are the substrate adaptors for a large family of Cullin3-dependent E3 ligases that remain incompletely characterised. Recurrent mutations in the Kelch domain of KBTBD4 act as an oncogenic driver of medulloblastoma. I identified specific substrates of KBTBD4 mutants using mass spectrometry-based proteomics. These data suggest a novel gain-of-function mechanism for these tumour mutations. For other targets with known substrates, I performed SPOTs peptide arrays and biophysical analyses to map their binding motifs (degrons) and was able to solve the crystal structures for the substrate complexes of the E3 ligases KLHL20, KLHL12 and KLHL3 with substrates DAPK1, DVL1 and WNK3, respectively. The mapped 'LPDLV' motif in DAPK1 and 'PGGPP' motif in DVL were observed to insert into the central pockets of the Kelch domains to form hydrophobic interactions, whereas the WNK3 interface was more polar with WNK3 Thr541 forming a potential phosphorylation site to regulate salt homeostasis.

I further used the DAPK1 degron peptide to develop an Alphascreen assay for compound screening and identified the first low micromolar inhibitors of KLHL20 with potential application in cancer, as well as Alzheimer's disease. I also established a panel of 17 human Kelch proteins for future work leading to a new structure of KLHL17 as well as a selectivity panel that revealed the high selectivity of known inhibitors of KEAP1. These results confirm the diverse structural characteristics of the Kelch domains observed from my work. Overall, these features help to explain how E3 substrate adaptors can perform their diverse roles in protein regulation.

Declaration

I declare that there are no parts of this thesis or its research herein have been reproduced or accepted for another award or degree or diploma at any other university or learning institution. This thesis contains no other person's work except where stated in text. Chapter 3 and part of Chapter 1 are based on the publication -

Chen Z, Picaud S, Filippakopoulos P, D'Angiolella V, Bullock AN. Structural Basis for Recruitment of DAPK1 to the KLHL20 E3 Ligase. *Structure*. 2019;27(9):1395–1404.e4. doi:10.1016/j.str.2019.06.005.

Acknowledgments

Throughout my DPhil project and writing of this thesis I have received a great deal of support and assistance. Firstly I would like to thank my supervisors, Dr Alex Bullock and Dr Vincenzo D'Angiolella, for their expertise, guidance and kind encouragement. I would also like to extend my gratitude to the China Scholarship Council and the Nuffield Department of Medicine for providing financial support over the last four years.

I would like to acknowledge my colleagues for their valuable advice, care and sense of humour which makes every day a pleasure to be in the lab. I would like to thank the past and present members in the GFSU group for their support and help along the way, Dr Eleanor Williams, Mrs Alice Hunt, Dr Jong Fu Wong, Ms Elizabeth Brown, Dr Daniel Pinkas, Dr Alison Howarth, Dr Caroline Sanvitale-Rudolf, Dr Roslin Adamson, Dr Sarah Dixon-Clarke, Dr Georgina Kerr, and Ms Charlotte Manning; and the SGC as a whole for the wealth of knowledge shared by all, in particular the Biotech team. I would also like to thank Dr Hongbin Yang in the department of Oncology for his enduring patience and his insight into the cellular aspects of this thesis.

I would like to thank my family for their unconditional love and support. I would like to acknowledge the Oriel College for always being a warm place to visit, for all the opportunities it provides and the amazing friends I made there. I would also like to extend my sincere thanks to Dr Laiyin Nie, Ms Agnese Baronina and Ms Qinrui Wang for joining me in the Radcliffe Camera sightseeing group (aka writing club) and the cheerful time we spent together. Lastly, thanks to my desktop pets – a cactus family of four, for their quiet company throughout my thesis writing and their faith in me.

List of Abbreviations

Å	Angstrom
Alphascreen	Amplified Luminescent Proximity Homogeneous Assay screen
BSA	Bovine serum albumin
BTB	Bric-a-brac, Tramtrack, and Broad complex
cDNA	Complementary DNA
CoREST	REST corepressor
CRL	Cullin–RING ligases
CUL	Cullin
CV	Column volume
CHX	Cycloheximide
DAPK1	Death-associated protein kinase 1
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DUB	Deubiquitinating enzyme
DVL	Dishevelled
EPAS1	Endothelial PAS domain-containing protein 1
ESI	electrospray ionization
FBS	Fetal bovine serum
FL	Full length
FP	Fluorescence Polarization
HCS	Hampton Crystal Screen
HCl	Hydrochloric acid
HDAC	Histone deacetylase
HEPEs	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEK293	Human embryonic kidney cells
HIN	Hampton screen index
HPLC	High-performance liquid chromatography
HTRF	Homogeneous Time Resolved Fluorescence
HRP	Horseradish Peroxidase
IC50	Half maximal inhibitory concentration
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
JCSG	Joint Centre for Structural Genomics
KBTBD	Kelch repeat and BTB domain-containing
K_D	Dissociation constant
kDa	Kilodaltons
KEAP1	Kelch-like ECH-associated protein 1
KLHDC	Kelch domain-containing
KLHL	Kelch-like
LB	Lysogeny broth

LC/MS	liquid chromatography/mass spectrometry
LFS	Ligand Friendly Screen
LIC	Ligation Independent Cloning
LSD1	Lysine-specific histone demethylase 1
MB	Medulloblastoma
MGC	Mammalian Gene Collection
MIDAS	Medium Alphascreen
MR	Molecular replacement
MS/MS	Tandem mass spectrometry
NC	Nitrocellulose
NCS	Non-crystallographic symmetry
NEDD-8	Neural precursor cell Expressed Developmentally Down-regulated 8
NRF2	Nuclear factor erythroid 2-related factor 2
PCR	Polymerase chain reaction
PBS	Phosphate buffered saline
PDB	Protein Data Bank
PEG	Polyethylene glycol
PEI	Polyethylenimine
PHAI1	Pseudohypoaldosteronism type II
pI	Isoelectric point
PPTID	Pineal Parenchymal Tumour of Intermediate Differentiation
PVDF	Polyvinylidene fluoride
RING	Really interesting new gene
RNA	Ribonucleic acid
RT-PCR	Reverse Transcriptase PCR
S.D.	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sf9	Spodoptera frugiperda
siRNA	Short interfering Ribonucleic acid
SPR	Surface Plasmon Resonance
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TLS	Translation/Libration/Screw
T _m	Melting temperature
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
TWEEN	Polyethylene glycol sorbitan monolaurate
Ub	Ubiquitin
UBD	Ubiquitin binding domain
ULK1	Serine/threonine-protein kinase ULK1
UPS	Ubiquitin-proteasome system
WNK	With no lysine (K)
WT	Wild type

Table of Contents

Chapter 1 Introduction.....	1
1.1. Ubiquitination.....	1
1.1.1. The ubiquitin code.....	1
1.1.2. Writing, reading and erasing the ubiquitin code.	3
1.2. Cullin RING E3 ligases	3
1.3. BTB-Kelch family	7
1.3.1. Structural basis for substrate recruitment.....	7
1.3.2. Functional landscape of the BTB-Kelch family.....	12
1.3.2.1. KLHL20 in tumorigenesis and Alzheimer’s disease	17
1.3.2.2. KLHL12 in embryonic development.....	18
1.3.2.3. KLHL3 in blood pressure control	19
1.3.2.4. KBTBD4, a novel tumour driver with unknown mechanism.	20
1.4. Current methods applied in understanding substrate recruitment by BTB- Kelch family E3 ligases.....	22
1.5. Aims of this thesis.....	25
1.6. References	26
Chapter 2 Methods and Materials	34
2.1. Construct cloning.....	34
2.1.1. Construct design	34

2.1.2. Ligation Independent Cloning (LIC)	34
2.1.3. Restriction enzyme cloning (REC)	36
2.1.4. Two-step mutagenesis [5]	37
2.1.5. Competent cells preparation	37
2.1.5.1. Buffers and Media	37
2.1.5.2. Protocols	38
2.1.6. Transformation	39
2.1.7. RNA extraction and Reverse Transcripase PCR	39
2.2. Protein expression	40
2.2.1. E.coli expression	40
2.2.2. Baculoviral expression	41
2.2.3. Mammalian expression (HEK293T)	41
2.3. Protein purification	42
2.3.1. Buffers	42
2.3.2. Cell lysis	42
2.3.3. Immobilized metal ion affinity chromatography (IMAC)	43
2.3.4. Size exclusion chromatography	43
2.3.5. Ion-exchange chromatography	43
2.3.6. Protein handling	44
2.4. SPOTs assay	44
2.4.1. Peptide array design	44
2.4.2. Peptide array interaction assay	45
2.5. Isothermal titration calorimetry	45
2.6. Surface Plasmon Resonance (SPR)	47

2.7. Alphascreen.....	48
2.8. Western blotting.....	48
2.9. Co-immunoprecipitation.....	49
2.10. Mass Spectrometry.....	50
2.10.1. Intact Mass (LC/MS).....	50
2.10.2. Phosphorylation site mapping.....	51
2.10.3. Interactome proteomics.....	51
2.11. X-ray crystallography.....	52
2.11.1. Crystallisation.....	52
2.11.2. Structure determination.....	53
2.12. Ubiquitination assay.....	54
2.13. Homology model.....	55
2.14. Nuclear extraction.....	55
2.15. Cycloheximide chase assay.....	55
2.16. References.....	56
Chapter 3 Structural basis for recruitment of DAPK1 to the KLHL20 E3 ligase.	58
3.1. Introduction.....	58
3.2. Results.....	59
3.2.1. Mapping of the DAPK1 binding motif for KLHL20 recruitment	59
3.2.2. An 'LPDLV' motif in DAPK1 is critical for KLHL20 interaction	61
3.2.3. High resolution structure of the KLHL20 Kelch domain bound to DAPK1 peptide.....	63
3.2.4. Extended interactions of the DAPK1 peptide.....	67
3.2.5. Interactions of the 'LPDLV' motif.....	68

3.2.6. Direct and water-mediated hydrogen bonding in the KLHL20 Kelch domain-DAPK1 complex.....	70
3.2.7. KLHL20-induced degradation of DAPK1 is dependent on the 'LPDLV' motif	71
3.2.8. A putative KLHL20-binding site in EPAS1 contains a 'GPDVL' motif	73
3.2.9. Small molecule inhibitor screening for KLHL20.....	74
3.3. Discussion	79
3.4. References	82
Chapter 4 Structural basis for recruitment of DVL1 to KLHL12	85
4.1. Introduction.....	85
4.2. Results.....	89
4.2.1. A 'PGGPP' motif in DVL1 is critical for KLHL12 interaction.	89
4.2.2. Crystallisation trials and structure determination for a KLHL12-DVL1 complex.....	90
4.2.3. Structure overview of KLHL12 Kelch domain bound to DVL1 peptide.	94
4.2.4. Interactions in the KLHL12 Kelch domain-DVL1 complex...	96
4.2.5. KLHL12-induced degradation of DVL1 is dependent on the 'PGGPP' motif.....	98
4.2.6. 'PGXPP' is a consensus motif in DVL1, 2 and 3 for KLHL12 interaction.	99
4.3. Discussion	100
4.4. References	105

Chapter 5 Characterisation of the KLHL3-WNK3 complex.....	107
5.1. Introduction.....	107
5.2. Results.....	109
5.2.1. The WNK3 degron peptide binds to the KLHL3 Kelch domain with a K_D in micromolar range.	109
5.2.2. Structure determination of the WNK3 acidic degron motif in complex with KLHL3	110
5.2.3. Structural comparison of WNK3 and WNK4 binding to KLHL3115	
5.2.4. Newly observed interactions in the WNK3 binding interface	117
5.2.5. Steric constraints would disfavour phosphothreonine in the WNK3 interface	118
5.2.6. Phosphorylation mapping of WNK3 under different stimuli.	121
5.3. Discussion	123
5.4. References	125
Chapter 6 Identifying the novel substrates for KBTBD4, a subtype-specific cancer driver in Medulloblastoma	126
6.1. Introduction.....	126
6.2. Results.....	127
6.2.1. Mass Spectrometry Proteomics identified CoREST-LSD1 complex as specific binders for KBTBD4 MB mutants	128
6.2.1.1. Sample preparation for mass spectrometry.....	128
6.2.1.2. Mass spectrometry hit analysis	129
6.2.2. Validation of interaction partners in HEK293T cells by co-immunoprecipitation and Western blot.....	132

6.2.2.1. CoREST1 and LSD1 bind to KBTBD4 mutants but not WT.	132
6.2.2.2. Two isoforms of CoREST3 bind to KBTBD4 mutants but not WT.....	132
6.2.3. CoREST3 is ubiquitinated and destabilised by KBTBD4 Group 3 and 4 mutants.	134
6.2.3.1. Endogenous CoREST3 are destabilised by KBTBD4 Group 3 and 4 mutants.....	134
6.2.3.2. In vitro ubiquitination assay for CoREST3 as a substrate.	135
6.3. Discussion	136
6.4. References	139
Chapter 7 A selectivity panel of Kelch proteins	142
7.1. Introduction.....	142
7.2. Results.....	143
7.2.1. Protein expression and purification	143
7.2.2. Crystallisation trials	147
7.2.3. Structure determination of KLHL17	147
7.3. Discussion	150
7.4. Reference	151
Chapter 8 Conclusions and Future Directions	152
References	157
Appendix - List of constructs and cloning primers	158

Chapter 1

Introduction

1.1. Ubiquitination

1.1.1. The ubiquitin code

Ubiquitination is a prevalent protein posttranslational modification playing critical roles in cellular processes. Ubiquitin (Ub), a 76-residue protein highly conserved among all eukaryotes, is covalently conjugated to the lysine (K) residues of substrate proteins via an isopeptide bond. A single substrate can be modified at one or multiple lysine residues with a ubiquitin molecule, termed as mono- and multi-monoubiquitination respectively. Modification of the N-terminal methionine or one of the seven lysine residues (K6, K11, K27, K29, K33, K48 and K63) of a substrate-attached ubiquitin leads to formation of polymeric chains (Figure 1.1A). Homogenous chains are formed when the same residue is modified. In contrast, heterogeneous chains contain mixed linkages within the same polymer. Moreover, modification of multiple lysine residues of a single ubiquitin initiates more complexed branched chains (Figure 1.1B). Monoubiquitination and polyubiquitination with different chain topologies, referred to as a 'ubiquitin code', stores and transmits biological information [1, 2].

The ubiquitin code determines the fates of modified substrates and controls almost every cellular process. The ubiquitin-proteasome system (UPS) is the major proteolytic system in eukaryotes. In this system, substrates are mostly modified with Lys-48 or Lys-11 linked poly-Ub chain, which is recognized by the proteasome for degradation. Inhibiting the proteasome with MG132 rapidly increases the substrate levels [3]. Apart from the UPS, other less common

chains with different linkages serve as signals for non-proteolytic regulatory mechanisms, as summarized in Figure 1.1C [4].

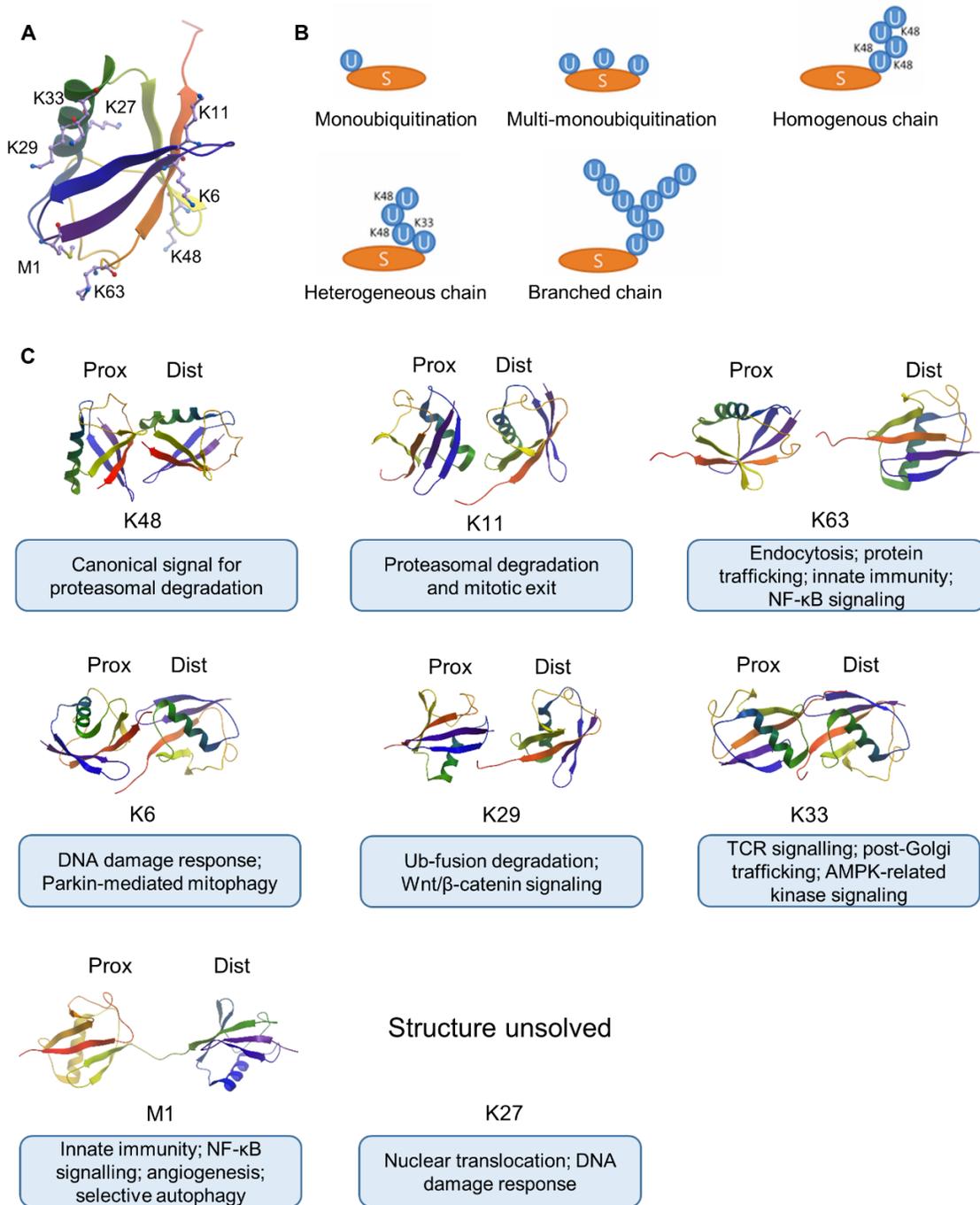


Figure 1.1 The ubiquitin code. (A) Structure of ubiquitin, showing the seven Lys residues and Met1. (B) The different topologies of ubiquitination. S, substrate. U, ubiquitin. Linkages are specified with residue numbers. (C) Structures of di-Ub with different linkages and their physiological functions. Prox, proximal to substrate. Dist, distal to substrate. PDB: K48, 2PE9; K11, 2XEW; K63, 2JF5; K6, 2XK5; K29, 4S22; K33, 5AF4; M1, 2W9N.

1.1.2. Writing, reading and erasing the ubiquitin code.

Ubiquitination is catalysed by a cascade of three enzymes – a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2), and a ubiquitin ligase (E3). Ub is activated by E1 in an ATP-dependent manner and transferred to E2, producing a thioester bond linked E2-Ub conjugate. E3 simultaneously interacts with E2-Ub conjugate and a specific substrate to mediate the final step of Ub transfer. The ubiquitin chain linkage specificity is likely determined by the E2 [5], with a few exceptions that the E3 or a particular substrate-E3 complex determines the chain types with a non-specific E2 [6, 7].

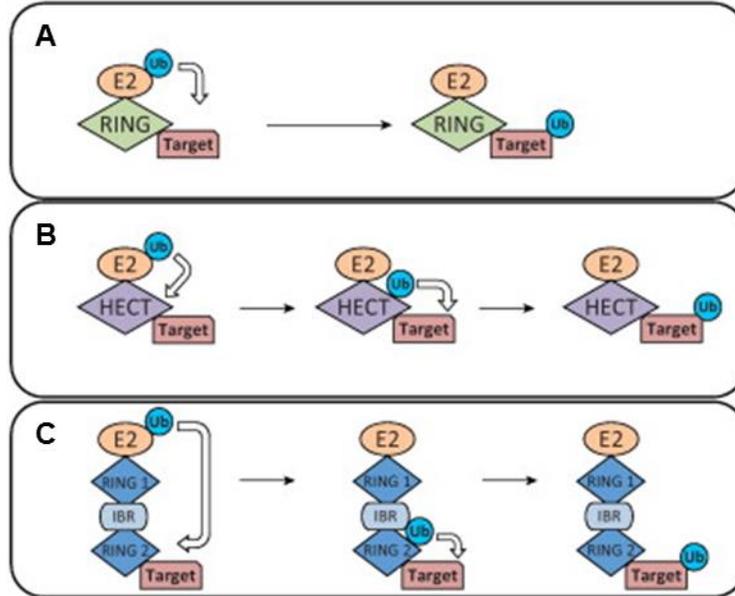
In order to decipher the ubiquitin code, cells have evolved a range of ubiquitin binding domains (UBDs) that specifically recognize and bind to distinct patches on the ubiquitin surface [8]. The UBDs couple the ubiquitination to a specific outcome; with Rpn10 and Rpn13 recognizing K48 linked chains and recruiting ubiquitinated substrates to the 26S proteasome for degradation as examples [9-12]

After being read by UBDs, the ubiquitin code is erased by deubiquitinating enzymes (DUBs) to prevent ubiquitination signal from being constitutively on [13]. Several proteasome-associated DUBs play a housekeeping role to protect ubiquitin from degradation, therefore maintaining sufficient levels of free ubiquitin for new chain assembly [1]. In other cases, sequential actions of DUBs and E3s enable ubiquitin code editing, which changes the fate of modified substrates [14]. Together, the writing, reading and erasing system of the ubiquitin code reflects the strict and sophisticated regulation of this modification.

1.2. Cullin RING E3 ligases

The human genome encodes two E1s, about 40 E2s, and more than 600 E3s. The E3s are the critical components of this cascade owing to their strict regulation of both the reaction efficiency and substrate specificity. Based on the protein folds and mechanisms of Ub transfer,

E3s are classified into two major types – the RING (really interesting new gene) type E3s and the HECT (homologous to the E6AP carboxyl terminus) type E3s (Figure 1.2A and B) [15]. The RING E3s are characterized by their zinc-bound RING catalytic domain, which recruits E2-Ub conjugates and enables direct Ub transfer from E2 to substrate [16]. A similar mechanism and fold to the RING domain is employed by U box E3 ligases, but the zinc-bound sites are replaced by a hydrophobic core to support the folding instead [17]. By contrast, HECT E3s undergo a catalytic cysteine-dependent transthiolation reaction with E2–Ub, producing a E3-Ub conjugate before transfer of the Ub to the substrate [18]. An emerging class of RING-IBR-RING (RBR) E3s has been recently discovered as a RING-HECT hybrid type [19, 20] (Figure 1.2C). The RBR protein comprises two RING fingers, RING1 and RING2, and an in-between-RINGs (IBR) domain. The RING1 domain, similar to classic RING type E3s, binds E2-Ub and brings it into close proximity to the RING2 domain, in which a conserved cysteine is capable of forming a reversible thioester intermediate with Ub.



(Figure from Rieser et al., 2013)

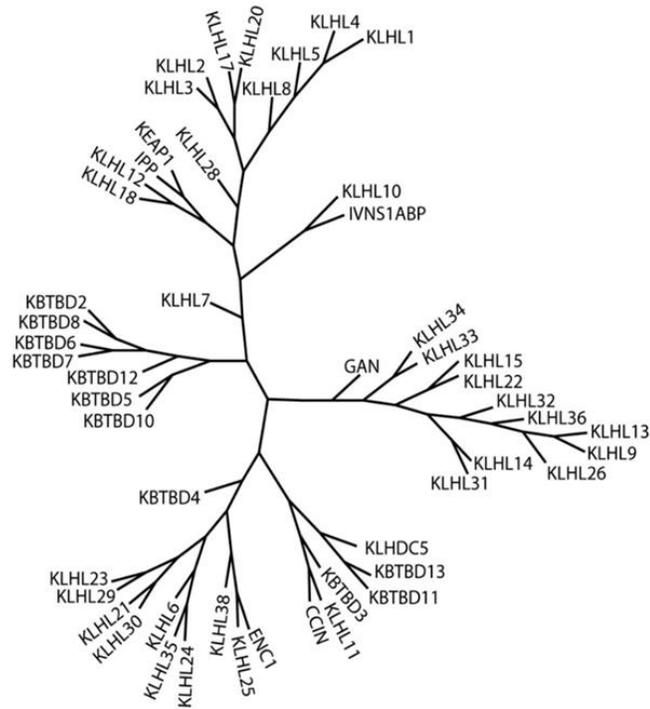
Figure 1.2 Model of ubiquitin chain formation mediated by RING, HECT, and RBR E3s. (A) RING E3 ligases mediate the ubiquitin transfer by bringing the E2-Ub into close proximity to the substrate without the formation of a covalent E3-ubiquitin conjugate. (B) HECT E3s form a thioester bond with ubiquitin before transferring the ubiquitin to the substrate. (C) RBR E3s incorporate a hybrid mechanism. The RING1 domain recruits the E2-Ub conjugate and the RING2 domain forms a thioester bond with the ubiquitin before transferring the ubiquitin to the substrates. The IBR domain connects RING1 and RING2 domains in RBR E3s.[21]

Cullin-RING ligases (CRLs) are a superfamily of multi-subunit RING-type E3 complexes, comprising the largest family of Ub ligases [15]. Over 240 members are grouped by different Cullin scaffold proteins, namely, CUL1 to CUL5, CUL7 and CUL9. The C-terminal domain (CTD) of Cullins associates with RING domain protein RBX1 or RBX2, which serves as a catalytic site for E2-Ub binding and ubiquitin transfer. The N-terminal domain (NTD) of Cullins interacts with specific sets of substrate adaptors, which recruit target proteins for ubiquitination [22].

Regardless of the identity of Cullins and specific substrate adaptor modules involved, it is well recognized that these E3 complexes share similar regulatory features [22]. CRLs are activated by a ubiquitin-like molecule Neural precursor cell Expressed Developmentally Down-

regulated 8 (NEDD-8) which is covalently linked to a conserved lysine residue in the C-terminal helical motif of Cullins [23]. This process, termed neddylation, is catalysed through an E1 NEDD8-activating enzyme (NAE)-E2 NEDD8-conjugating enzyme cascade [24]. Physiologically, neddylation is reversible by a deneddylase complex COP9 signalosome (CSN) [25]. After catalysing deneddylation, CSN remains bound to CRLs and sterically hinders substrate access, sequestering CRLs in the inactive state. Besides this, a small molecule inhibitor MLN4924 has been proven to effectively block CRL activity by targeting NAE [26, 27]. This compound is used for stabilizing CRL substrates in this thesis.

CUL3 binds BTB-containing proteins exclusively as substrate adaptors [28]. The BTB-containing adaptors make CUL3-RING ligases (CRL3) unique among all CRLs with two features. Firstly, the BTB domain is capable of dimerization, therefore recruiting two CUL3 molecules in one CRL3 complex. Secondly, unlike adaptors of other Cullins recruiting a separate module for substrate recognition, BTB-containing adaptors possess a second protein-protein interaction domain for substrates to bind. The range of substrate-binding domains include the C-terminal Kelch, PHR (PAM, Highwire, and RPM-1), zinc finger domains and N-terminal MATH (meprin and TRAF homology) domain. Among these BTB-containing adaptors, the BTB-Kelch family, which comprises 52 members, appears the largest group [29, 30] (Figure 1.3).



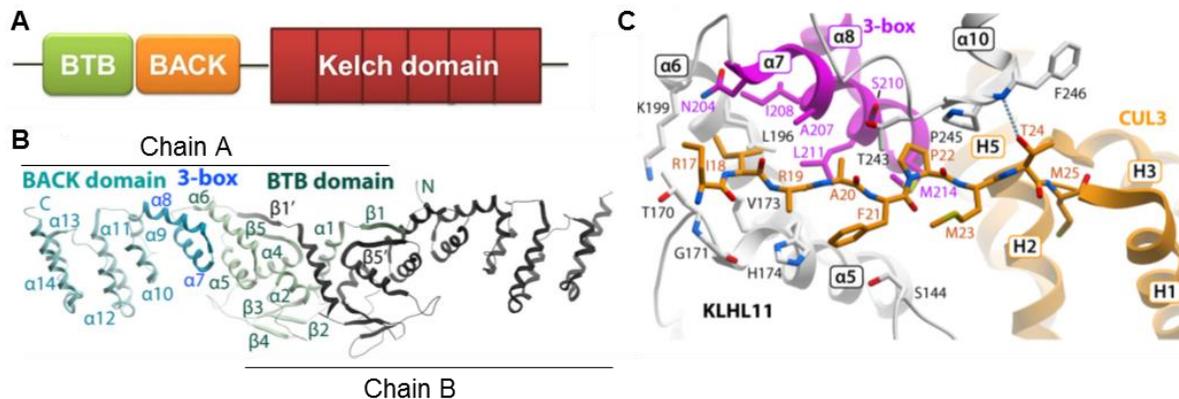
(Figure from Canning et al., 2013)

Figure 1.3 Phylogenetic tree of the BTB-Kelch family. The phylogenetic tree was generated based on multiple sequence alignment of human Kelch domains [29].

1.3. BTB-Kelch family

1.3.1. Structural basis for substrate recruitment

The BTB-Kelch proteins utilize multiple domains to facilitate their dual functions as CUL3 adaptors and substrate recognition modules (Figure 1.4A). The BTB domain (Bric-a-brac, Tramtrack, and Broad complex) shares a conserved fold with the CUL1 adaptor SKP1 and the CUL2/5 adaptor ElonginC. As mentioned above, the BTB domain mediates the homo-dimerization of BTB-Kelch proteins (Figure 1.4B). The last two helices in the BTB domain and the ‘3-box’ motif in the BACK domain form a four-helix bundle to provide a hydrophobic groove for CUL3 assembly (Figure 1.4C) [29]. The C-terminal Kelch domain, which contains six Kelch repeats, is required for substrate capture.



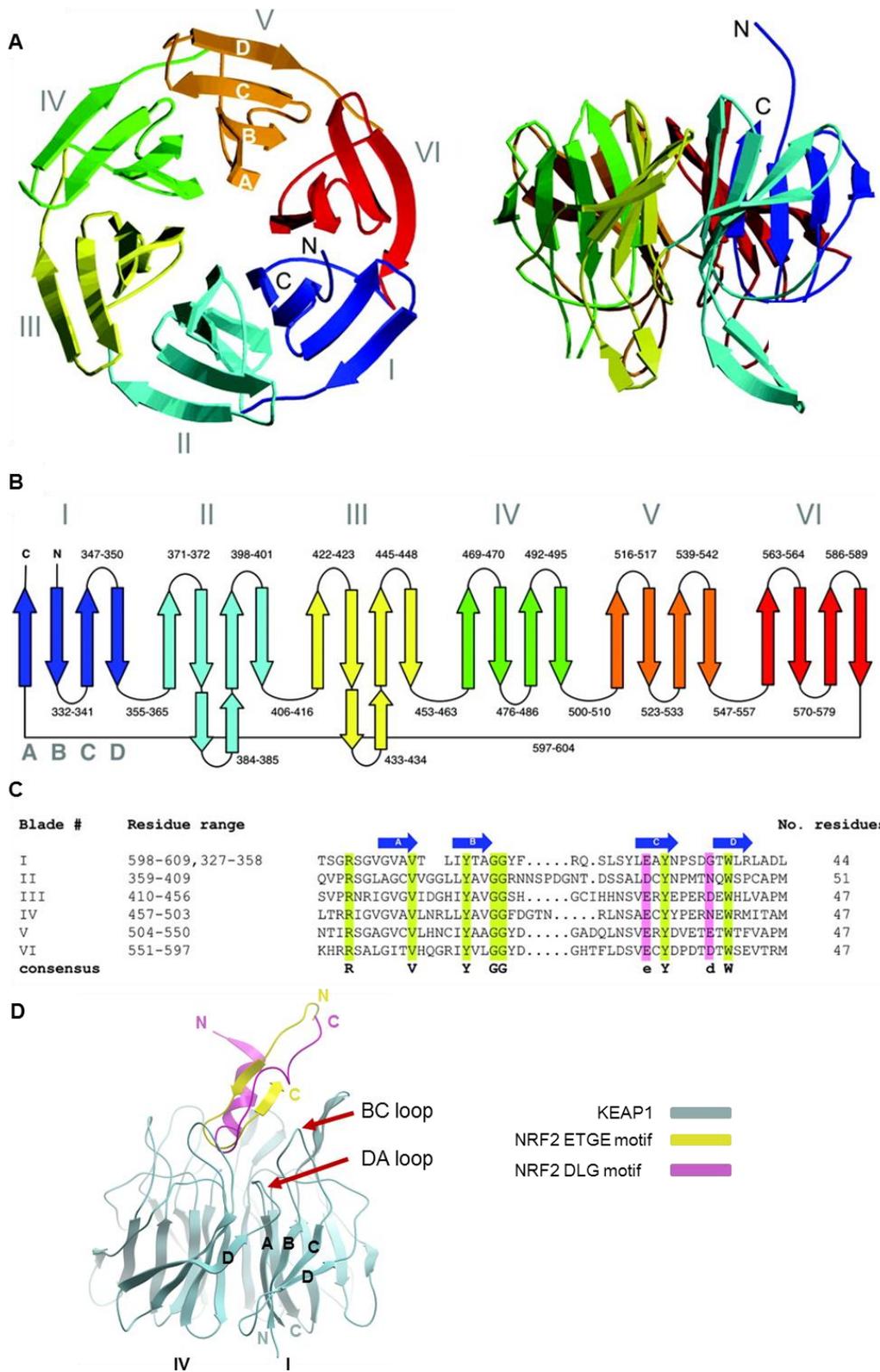
(Figure adapted from Canning et al., 2013)

Figure 1.4 A schematic diagram of a BTB-Kelch protein and the structures of BTB-BACK domains. (A) Domain organization of BTB-Kelch proteins. (B) Ribbon representation of the dimeric structure of the KLHL11 BTB-BACK domains. In Chain A, the BTB domain, 3-box motif and BACK domain are colour coded as their labels. Chain B is coloured grey. PDB: 3I3N (C) Side chain interactions of the CUL3 N-terminus in the hydrophobic groove of KLHL11. PDB: 4AP2 [29].

Several crystal structures have revealed the structural features of the Kelch domain (Table 1.1). Overall, the Kelch domain shows a β -propeller fold (Figure 1.5A) [29]. The six Kelch repeats form the six blades (I-VI) of the propeller arranged radially around a central axis. Each repeat is folded into a twisted β -sheet consisting of four antiparallel β -strands. A final C-terminal β -strand is observed to close the β -propeller and inserts into blade I as the innermost β A strand. Blade I is therefore comprised of a C-terminal β A strand and N-terminal β B, β C, and β D strands (Figure 1.5B). Packing between each blade is mediated by a number of conserved hydrophobic positions as well as several buried charged residues that recur within each kelch repeat (Figure 1.5C). The substrate binding surface is shaped by the long BC loops, which protrude outwards from the Kelch domain surface, and the largely buried DA loops which link adjacent blades and contribute to the protein core (Figure 1.5D). Due to limited sequence identity, the surface properties of the Kelch domains are strikingly diverse, which contributes to the substrate specificity of each BTB-Kelch protein [29, 31].

Table 1.1 Summary of structures of human Kelch domains in protein data bank (PDB)

BTB-Kelch	Kelch structures
proteins	
KLHL2	2XN4: Apo. 4CHB: WNK4 bound.
KLHL3	4CH9: WNK4 bound. 5NKP: WNK3 bound (this thesis).
KLHL7	3II7: Apo.
KLHL12	2VPJ: Apo.
KLHL17	6HRL: Apo (this thesis).
KEAP1	1U6D: Apo. 2FLU: NRF2 ETGE bound. 3WN7: NRF2 DLG bound. Other structures
(KLHL19)	with various inhibitors
KLHL20	5YQ4: Apo. 6GY5: DAPK1 bound (this thesis).
NS-BP1	6N3H: Apo.
(KLHL39)	
KBTBD5	4ASC: Apo.
KBTBD10	2WOZ: Apo

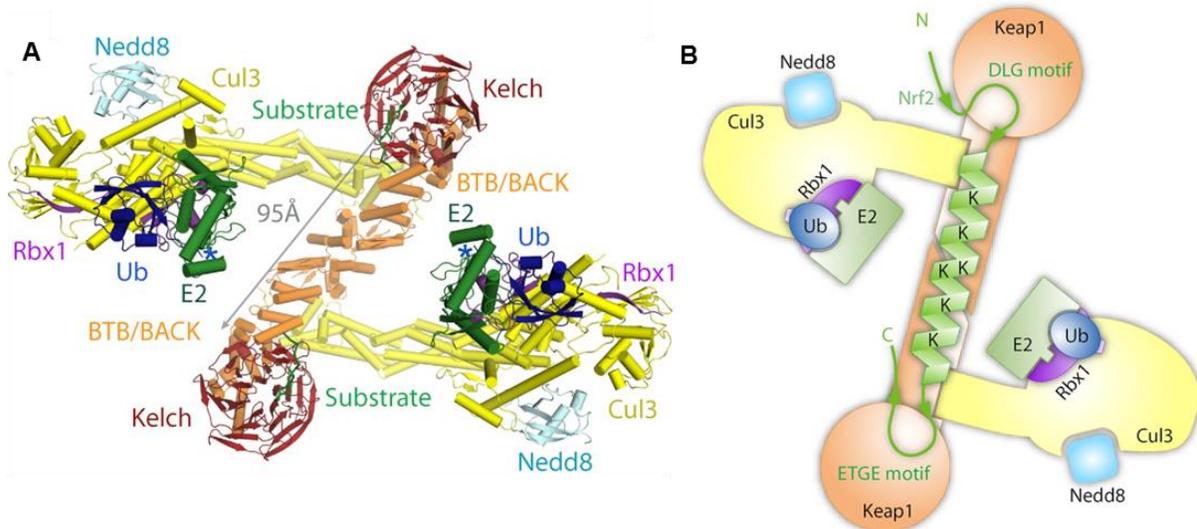


(Figure adapted from Li et al., 2004. Figure legend on next page)

Figure 1.5 The structural features of the Kelch domain. (A) Ribbon representation of the β -propeller fold of a Kelch domain. (KEAP1 as an example, PDB: 1U6D). Six blades are coloured from blue to red and labelled I-VI. The four anti-parallel β -strands in each blade are labelled A–D on blade V. The N and C termini are both in blade I and labelled with N and C, respectively. (B) Topology diagram of a Kelch domain using KEAP1 as an example, showing the four antiparallel β -strands in each blade (C) Sequence alignment of the six Kelch repeats of KEAP1. The four conserved β -strands in each blade are shown by arrows. Residues conserved in all six blades are highlighted in green; residues that play conserved structural roles are highlighted in pink. A consensus sequence for all six blades is shown at the bottom. (D) Ribbon representation of substrate-bound Kelch domain. The figure shows the superposition of the structures of KEAP1 in complex with two distinct NRF2 motifs, as labelled. The BC loop in Blade I and DA loop in Blade VI are labelled with red arrows. PDB: 2FLU, 3WN7 [32].

The best characterized BTB-Kelch protein is KEAP1 (Kelch-like ECH-associated protein 1), also known as KLHL19. KEAP1 functions as a critical sensor of oxidative stress and a master regulator of the antioxidant response by controlling the levels of transcriptional factor NRF2 (nuclear factor erythroid 2-related factor 2) [33]. Under normal conditions, KEAP1 ubiquitinates NRF2 with K48-linked chains and lead it to proteolysis [34, 35]. However, upon cellular stress, oxidation or covalent cysteine modification, the activity of KEAP1 is blocked and therefore NRF2 accumulates in the nucleus where it activates the expression of genes possessing an Antioxidant Response Element (ARE) in their promoters [36]. Two distinct degron motifs, the ETGE and DLG sites, in NRF2 were identified for KEAP1 binding. The ETGE motif binds to KEAP1 at low nanomolar concentrations ($K_D \sim 5$ nM), whereas the DLG motif binds to KEAP1 at 100-fold weaker affinity [37]. Crystal structures for these individual degrons bound to the KEAP1 Kelch domain have been solved previously (Figure 1.6D). DLG and ETGE peptides insert into the substrate binding pocket in the Kelch domain in a similar manner, despite deviation at their N-terminus owing to the helical conformation of the DLG peptide motif. Based on the crystal structures of the KEAP1-NRF2, CUL3-KLHL11 dimeric complex, the CUL5 CTD-RBX1-NEDD8 complex and the E2-ubiquitin intermediate, a molecular model has been proposed that the separate ETGE and DLG motifs allow a single NRF2 molecule to engage

two Kelch domains present in the dimeric CRL3(KEAP1) complex (Figure 1.6) [38]. The proper assembly is thought to be important to correctly position the NRF2 for ubiquitination.



(Figure adapted from Canning et al., 2015)

Figure 1.6 The two-site recognition model proposed for NRF2 recruitment by KEAP1. (A) Structural model of fully assembled CRL3 complex. In the model, the scaffolding protein CUL3, modified by Nedd8, connects the catalytic RING protein Rbx1 and a substrate adaptor BTB-Kelch. The complex dimerizes through the BTB domain. The substrate is recruited to this complex by directly binding to the BTB-Kelch protein. The Rbx1 binds to an E2-Ub conjugate. (B) Schematic representation of NRF2 recruitment by KEAP1. Two distinct motifs in NRF2 bind to two Kelch domains present in the KEAP1 dimer. The intervening α -helix contains seven substrate lysine residues of which six are predicted to fall on the same face [38].

To date, few Kelch-substrate complexes have been structurally characterized, with the major examples being the KEAP1-NRF2 [37, 39] and KLHL2/3-WNK4 systems [40] (Table 1.1). Despite the growing number of substrate proteins identified for BTB-Kelch proteins, there remains limited knowledge of their specific binding epitopes.

1.3.2. Functional landscape of the BTB-Kelch family

The BTB-Kelch family consists of 52 members, targeting a broad spectrum of substrates involved in various cell processes. The critical roles of BTB-Kelch proteins in physiological and pathophysiological states are emerging with recent functional analyses [41-44]. Existing data,

as summarized in Table 1.2, show that BTB-Kelch proteins modify substrates with both proteolytic and non-proteolytic ubiquitination signals. By ubiquitinating these substrates, BTB-Kelch proteins function in metabolism, immunization, programmed cell death and many other processes. Of note, 8 out of 27 listed BTB-Kelch proteins (marked with an asterisk) regulate the cell cycle, suggesting their fundamental roles in cells as well as their potential links to tumorigenesis. Indeed, a wide range of diseases including cancer have been found to be associated with alterations of BTB-Kelch proteins (Table 1.3).

Table 1.2 Summary of the functions of BTB-Kelch proteins

BTB-Kelch	Substrates	Ubiquitination and outcome	Biological impact	References
KLHL2	WNK1-4	Degradation	Blood pressure control. Mutations in substrate-E3 interface cause Gordon's syndrome	[40]
	NPCD	Protein complex formation	Neurodegeneration	[45]
KLHL3	WNK1-4	Degradation	Blood pressure control. Mutations in substrate-E3 interface cause Gordon's syndrome	[40]
	Claudin-8	Degradation	Blood pressure control.	[46]
	cMyBP-C	Degradation	Heart development	[47]
KLHL6*	Roquin2	Degradation	B cell survival and differentiation	[48]
	CDK2	Degradation	B cell survival and differentiation, cell cycle	[49]
KLHL7	TUT1	K48 linked poly-ub. Degradation	Nucleolar integrity	[50]
KLHL8	Rapsyn	Degradation	Essential regulation on neuromuscular junction	[51]
KLHL9*	Aurora B	Monoubiquitination. Subcellular relocalisation	mitotic progression	[52]
KLHL12	SEC31	Monoubiquitination. Protein complex formation	COPII coat size control. Collagen export.	[53]
	Dopamine D4 receptors	Unknown	Unknown	[54, 55]
	DVL1-3	Degradation	Inhibiting WNT signalling pathway	[56]

Continued on next page

KLHL13*	Aurora B	Monoubiquitination. Subcellular relocalisation	Mitotic progression	[52]
KLHL15	RBBP8	Degradation	DNA damage response	[57]
	PP2A	Degradation	Shaping dephosphorylation profiles in cells	[57]
KLHL16 (GAN)	MAP1B	Degradation	Anti-degenerative and anti- apoptosis effects in neurons	[58]
	TBCB	Degradation	Protecting from cytoskeletal abnormalities	[59]
	MAP8B	Degradation	Maintaining normal retrograde axonal transport	[59]
KLHL17 (Actinfilin)	GluR6	Degradation	Contributing to the synaptic normality among neurons	[60]
KLHL18*	Aurora A	Altered protein activity	Necessary for mitotic entry	[61]
KLHL19 (KEAP1)*	MCM3	Altered protein activity	Cell cycle control	[62]
	NRF2	Degradation	Critical for oxidative stress response; Inhibited tumorigenesis and chemo- resistance in a variety of malignancies; Triggering the onset of diabetes and excessive adipogenesis	[33]
	IKK β	Degradation	Tumour-inhibitory effects in various malignancies by suppressing NF- κ B signalling; Lowered inflammatory cytokines produced by macrophage	[63]
	PGAM5	Degradation	Activation of pro-apoptotic behaviour	[64]
	p62	Altered protein activity	Competitive inhibition of NRF2 to stimulate anti- oxidant responses; Enhanced autophagic activity	[65, 66]
	PALB2	monoubiquitination. Altered protein activity	Prohibited homologous recombination in G1 cells	[67]
	KLHL20	DAPK1	Degradation	Anti-apoptosis. Increased metastatic capability in colon cancer
	PML	Degradation	Promoting tumorigenesis by degrading tumour suppressor PML	[69]
	ULK1	Degradation	Governing autophagic termination	[70]
	PDZ-RhoGEF	Degradation	Driving neurite outgrowth	[71]
	Coronin 7	K33 linked Ub chain. Altered protein activity	Modulating post-Golgi trafficking	[72]
	EPAS1	Stabilisation	Facilitating hypoxia response	[73]

Continued on next page

KLHL21*	Aurora B	Monoubiquitination. Subcellular re-localization	Completion of mitotic event	[74]
	End binding 1 (EB1)	Monoubiquitination.	Cell migration	[75]
KLHL22*	PLK1	Monoubiquitination. Subcellular re-localization	Mitotic progression	[76]
	DEPDC5	K48-linked polyub. Degradation.	Autophagy and cell growth	[77]
KLHL24	Keratin 14	Degradation	Skin structure	[78]
KLHL25	ATP-citrate lyase (ACLY)	Degradation	Repression of the neoplastic characteristics of lung cancer cells	[79]
	4E-BP1	Degradation	Enhanced translational activity	[80]
KLHL31	Filamin-C (FlnC)		Skeletal and cardiac muscle myogenesis	[81]
KBTBD2	p85 α	Degradation	Activation of PI3K pathway and enhanced insulin sensitivity	[82]
KBTBD5	DP1	Degradation	Skeletal muscle differentiation	[83]
	LMOD and NEB	Degradation	Skeletal muscle differentiation	[84]
KBTBD6	TIAM1	Degradation	Restricted migration of breast carcinoma cells	[85]
KBTBD7	TIAM1	Degradation	Restricted migration of breast carcinoma cells	[85]
KBTBD8*	TCOF1/NOLC1	Monoubiquitination. Protein complex formation	Driving neural crest specification by regulating spindle at mitosis	[86]
KBTBD10 (KLHL41)	LASP1	Unknown	Skeletal muscle differentiation and myofibril assembly	[87]
KBTBD11	NFATc1	Degradation	negatively regulating osteoclastogenesis	[88]
KLHDC5*	p60/Katanin	Degradation	Mitosis control	[89]

*BTB-Kelch proteins that regulate the cell cycle.

Table 1.3 Summary of disease links of BTB-Kelch proteins upon gene alterations

BTB-Kelch	Disease links	Alterations	Reference
KLHL1	Spinocerebellar ataxia type 8 (SCA8)	Antisense regulation (downregulation)	[90]
KLHL2	Hypertension	Loss-of function missense mutations in BTB or Kelch domain	[91]

Continued on next page

KLHL3	Hypertension	Loss-of-function missense mutations in BTB or Kelch domain	[92]
KLHL4	X-linked cleft palate	A silent mutation	[93]
KLHL6	Lymphoma	Loss-of-function missense mutations in BTB or Kelch domain	[48]
KLHL7	Breast cancer	Loss-of-function missense mutations in BACK domain. Gene amplification.	[94]
KLHL9	Gastric cancer	Loss-of-function missense mutations in BTB or Kelch domain. Frameshift mutations and early stop codons.	[95]
KLHL10	Infertility	Loss of function missense mutations in BTB domain	[96]
KLHL13	Inherited peripheral neuropathies	A missense mutation in Kelch domain	[97]
KLHL14	Primary CNS lymphoma	Missense mutations in BTB or BACK domain	[98]
KLHL16	Carcinoma. Giant axonal neuropathy	Loss-of-function missense mutations in BTB or Kelch domain. Frameshift mutations and early stop codons.	[99]
KLHL17	Epilepsy	Deletion	[100]
KLHL19	A wide range of cancers in lung, galibladder, prostate and colorectal etc.	DNA methylation. Truncated mRNA transcripts. Loss-of-function missense mutations in BTB or Kelch domain	[101]
KLHL20	Prostate cancer and Alzheimer	Gene amplification	[69]
KLHL21	Carcinoma	Gene amplification	[102]
KLHL22	Breast cancer	Gene amplification	[77]
KLHL23	Gastric cancer	Overexpression	[103]
KLHL24	Epidermolysis bullosa	Truncation stabilising KLHL24 protein level	[104]
KLHL25	Lung cancer	Downregulation	[79]
KLHL35	Carcinoma	DNA methylation	[105, 106]
KLHL37 (ENC1)	A wide range of cancers including Melanoma, carcinomas, Hairy cell leukemia and neuroblastoma tumour	Gene amplification in Colorectal carcinomas and Hairy cell leukemia. DNA methylation in Melanoma. Gene downregulation in Neuroblastoma tumours.	[107-110]
KBTBD4	Medulloblastoma and Pineal Parenchymal Tumour of Intermediate Differentiation	Missense and indel mutations in Kelch domain.	[111, 112]
KBTBD5	Nemaline myopathy	Loss-of function missense mutations in BTB or Kelch domain. Frameshift mutations and early stop codons	[113]
KBTBD10	Nemaline myopathy	Loss-of function missense/indel mutations in BTB, BACK or Kelch domain.	[114]
KBTBD13	Nemaline myopathy	Loss of function missense mutations in Kelch domain	[115]

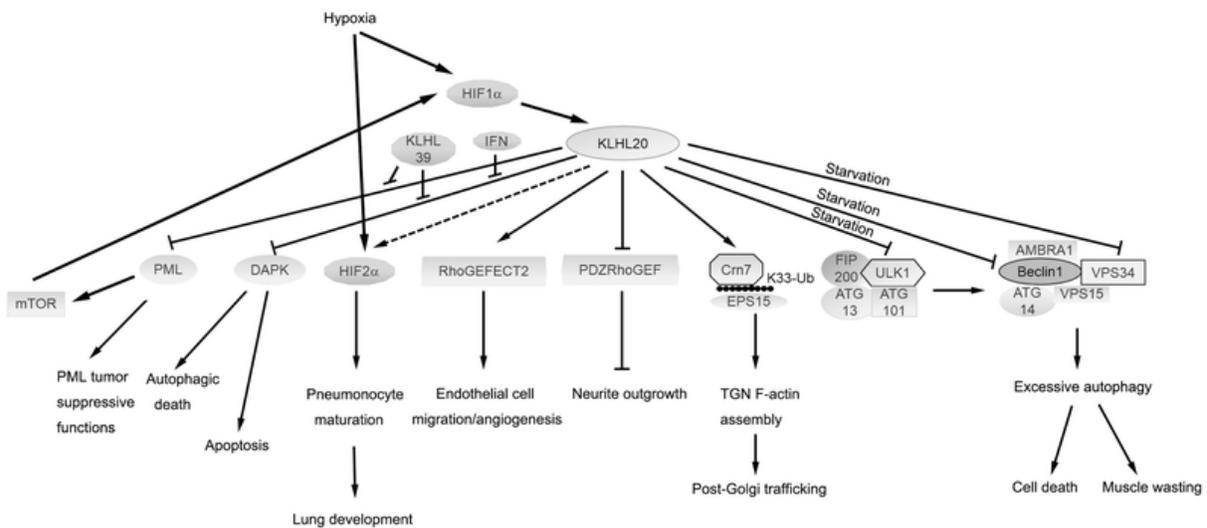
This thesis characterizes the substrate recruitment of KLHL20, KLHL12 and KLHL3, whose physiological roles are well established, as well as identifies the novel function of KBTBD4. The current understanding of these proteins are introduced below.

1.3.2.1. KLHL20 in tumorigenesis and Alzheimer's disease

KLHL20 has been shown to ubiquitinate some half a dozen protein targets that link its activities to diverse processes including autophagy, hypoxia, cancer and Alzheimer's disease [116]. To date, the majority of substrates identified for KLHL20 are targeted for proteasomal degradation, suggesting their modification by Lys48-linked polyubiquitin chains [116]. These include the substrates DAPK1 [68], PML [69], PDZ-RhoGEF [71] and ULK1 [117]. However, KLHL20 also plays an important role in protein trafficking by targeting coronin 7 to the trans-golgi network through atypical K33-linked polyubiquitination [72].

The substrates of KLHL20 reflect its function in cellular stress responses, as well as its linkage to human disease [116] (Figure 1.7). Transcription of the *KLHL20* gene is upregulated by the hypoxia-inducible factor HIF-1 α leading to its overexpression in hypoxic tumour cells. In this context, KLHL20 can promote tumorigenesis by degrading the tumour suppressor proteins DAPK1 and PML [69]. In human prostate cancer patients, higher levels of KLHL20 (and low PML) were found to correlate specifically with high grade tumours. Moreover, KLHL20 depletion in PC3 prostate cancer cells restricted the growth of tumour xenografts, suggesting KLHL20 as a potential therapeutic target [69]. KLHL20 also plays a critical role in autophagy termination by degrading the pool of activated ULK1 [117]. Thus, KLHL20 can restrict both apoptotic and autophagic cancer cell death. Importantly, interferon stimulation causes the sequestration of KLHL20 in so-called 'PML nuclear bodies', which occur as punctate membraneless substructures of the nucleus enriched with PML protein [68]. This allows DAPK1 to evade degradation and to accumulate to mediate interferon-induced cell death [68]. In addition to this inhibitory mechanism, KLHL39, another BTB-Kelch protein, competitively binds to KLHL20 and

suppresses cancer metastasis by blocking KLHL20-mediated PML and DAPK ubiquitination. Clinically, low expression of KLHL39 in human colon cancer correlates with low expression of PML and DAPK, higher tumour grade, lymph node metastasis, and distant metastasis. Notably, the stress responses of KLHL20 also appear linked to neurodegeneration with KLHL20 RNA transcript levels being among the top 20 biomarkers for Alzheimer's disease progression [118, 119]. Therefore KLHL20 has emerged as an interesting target for drug development with potential application in both oncology and Alzheimer's disease.



(Figure from Chen et al., 2016)

Figure 1.7 Overview of the regulators, substrates and biological functions of KLHL20 [116]

1.3.2.2. KLHL12 in embryonic development

KLHL12 is highly expressed in mouse embryonic stem (ES) cells, but significantly downregulated upon differentiation, suggesting critical roles of KLHL12 in certain stage of embryonic development [53]. In particular, KLHL12 negatively regulates the Wnt signalling pathway by degrading Dishevelled (DVL) proteins through the UPS. Overexpression of KLHL12 in zebrafish embryos inhibited antero-posterior axis elongation, as seen with loss of DVLS [56].

On the contrary to inhibiting cell polarity differentiation, KLHL12 promotes mouse ES cell proliferation by ubiquitinating SEC31 to drive large COPII coat assembly for collagen export [53]. Moreover, the SEC31 recruitment to Cullin3-KLHL12 complex requires PEF/ALG2 as target-specific co-adaptors which are regulated by Calcium [120].

Studies on KLHL12 have extended our understanding towards the current working model of BTB-Kelch proteins described in 1.3.1. In addition to the canonical homodimerisation, KLHL12 is also found to heterodimerise with KLHL26, which likely requires distinct motifs in its substrates for KLHL26 to bind [121]. With aforementioned Ca^{2+} responsive co-adaptors associated within the complex, the substrate recruitment to KLHL12 undergoes stringent control which regulates ubiquitination in a spatio-temporal manner [120].

KLHL12 modifies substrates with diverse ubiquitin chains. DVLs are targeted for proteasomal degradation, suggesting their modification by Lys48-linked polyubiquitination [56]. SEC31 is mono-ubiquitinated by KLHL12 for COPII complex assembly [53]. Notably, KLHL12 targets the Dopamine D4 receptor for both lysine and non-lysine ubiquitination, yet the subsequent processes remain elusive [54, 55, 122]. The Ub K29R K48R K63R triple mutant largely blocked the D4 poly-ubiquitination by KLHL12, however, none of the single mutants phenocopied the loss of ubiquitination [55], suggesting branched chain formation.

1.3.2.3. KLHL3 in blood pressure control

KLHL3 regulates mammalian blood pressure by targeting With No Lysine (WNK) kinases for ubiquitination-dependent proteolysis [123, 124]. Overexpression of WNK1 and mutations within a conserved acidic motif of WNKs located C-terminal to the kinase domain (residues 557–567 in WNK4) cause Gordon's hypertension syndrome (also known as pseudohypoaldosteronism type II, PHAII) [125, 126]. This acidic region was later found to operate as a degron motif for KLHL3 to bind, hence controlling the stability of WNKs. This conclusion is supported by the

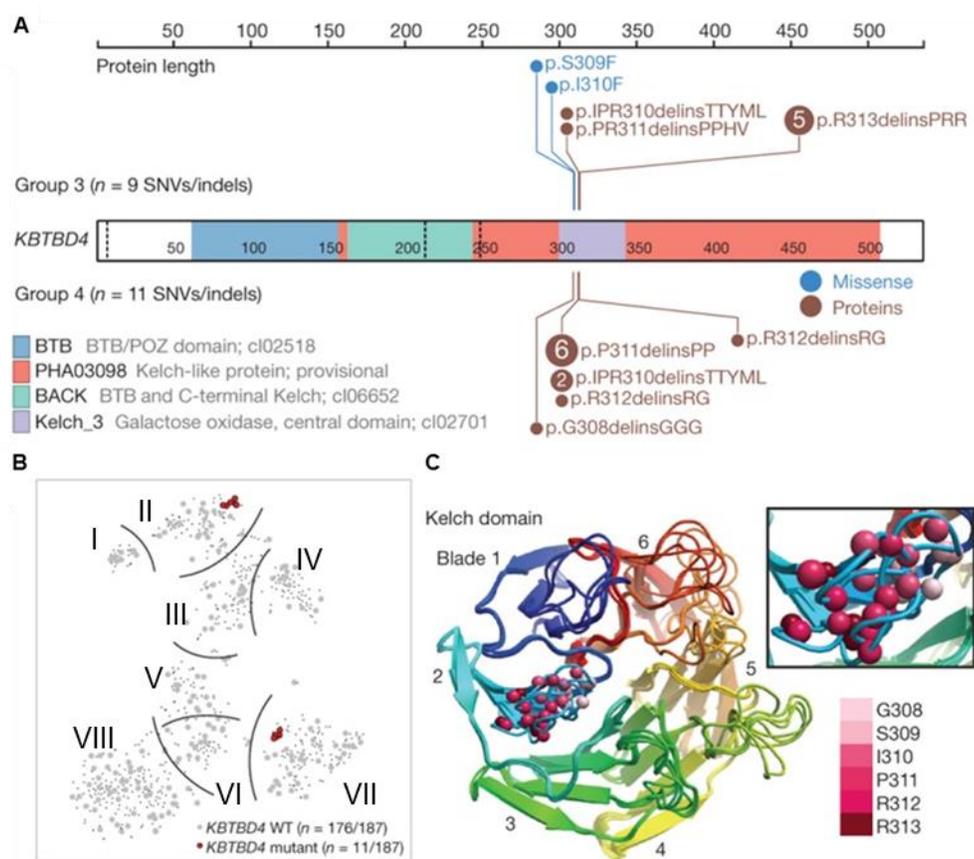
finding that mutations in KLHL3 that inhibit binding to either CUL3 or WNK4 also cause Gordon's hypertension syndrome [92, 127-129]. Consistent with this, animal model work also revealed that a WNK4 knock-in mouse incorporating a mutation within the degron motif and KLHL3 knockout mice showed similar PHAI1 phenotypes [130, 131]. These mutations inhibit the ubiquitination of WNK isoforms by the CRL3^{KLHL3} complex, and result in WNK stabilization and therefore the overactivation of the downstream kinases SPAK/OSR1 and their substrate NCC/NKCC2 ion cotransporters. Consequently, increased salt retention leads to hypertension [132].

Schumacher et al characterized the interaction between the WNK degron motif and KLHL3 as well as KLHL2, a BTB-Kelch protein sharing 86% overall sequence identity with KLHL3, using fluorescence polarization and X-ray crystallography [40]. KLHL3 and KLHL2 modify WNKs in a similar manner in distal convoluted tubules (DCTs) and medulla in kidney respectively [130, 133]. The structures of KLHL2/3-WNK4 complexes provide a molecular mechanism to explain how several of the Gordon's syndrome-associated mutations stabilize WNKs by disrupting the binding. The WNK degron motif binds to KLHL3 in a markedly different manner to how NRF2 binds to KEAP1, confirming the diversity and selectivity of the substrate binding pockets in Kelch domains.

1.3.2.4. KBTBD4, a novel tumour driver with unknown mechanism.

KBTBD4 is mainly localized to the nucleus (<https://www.proteinatlas.org/>) with unknown functions. As a member of the BTB-Kelch family, KBTBD4 is predicted to associate with CUL3 and facilitate ubiquitination on target proteins. Recent studies on medulloblastoma (MB) and Pineal Parenchymal Tumour of Intermediate Differentiation (PPTID) discovered recurrent mutations targeting *KBTBD4* [111, 112]. The majority of the somatic *KBTBD4* variants were determined to contain in-frame insertions clustered across six residues within the second blade of the Kelch domain (Figure 1.8A). In Group3 and 4 MB, *KBTBD4* mutations were tightly

clustered within subtypes II (21%) and VII (14%) (Figure 1.8B). Meanwhile, these subtypes as well as PPTID that harboured *KBTBD4* mutations, lacked other obvious oncogenic driver events, suggesting that the *KBTBD4* mutations are functional. Moreover, the *KBTBD4* mutations identified in MB and PPTID are all heterozygous mutations, which strongly suggests that these are activating, gain-of-function mutations. As seen in the homology model (Figure 1.8C), the hotspot mutations in *KBTBD4* Kelch domain reside in the substrate binding pocket [112]. Thus, *KBTBD4* mutants may function as cancer drivers by altering their substrate recruitment in MB and PPTID.



(Figure from Northcott et al., 2017)

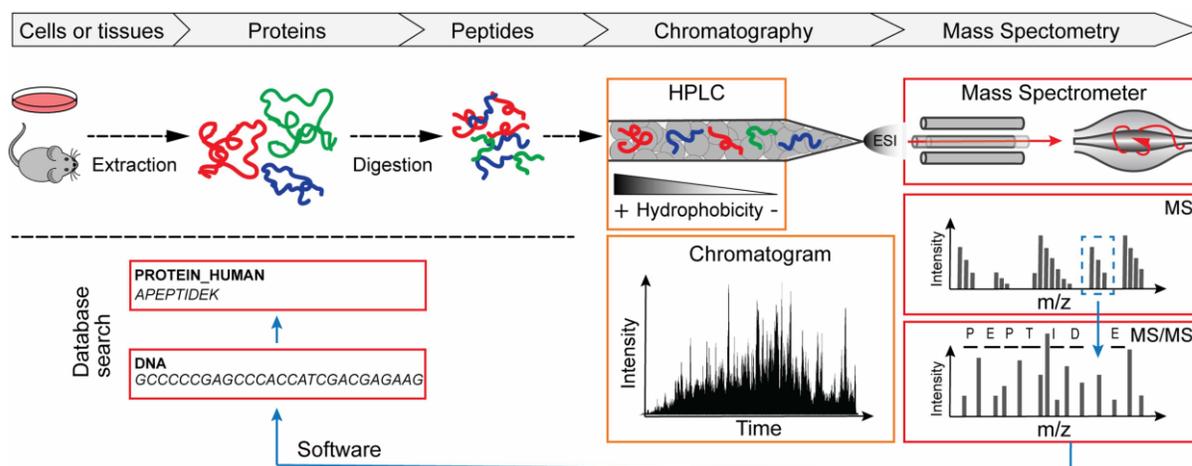
Figure 1.8 Recurrent hotspot mutations clustered within *KBTBD4*. (A) Summary of somatic mutations targeting *KBTBD4* in MB at gene-level. (B) Group 3 and Group 4 MB were classified into 8 methylation subtypes. *KBTBD4* mutants clustered in subtype II (21%) and VII (14%). (C) Homology models of the *KBTBD4* Kelch domain, highlighting the positions where hotspot insertions reside (shown as purple spheres). [112]

1.4. Current methods applied in understanding substrate recruitment by BTB-Kelch family E3 ligases

The regulatory functions of BTB-Kelch proteins are largely determined by the substrates they recognize and modify. On the other hand, the dynamic complexity of the ubiquitination process makes it a significant challenge to identify the direct substrates of individual BTB-Kelch proteins. Conventional screening methods include the yeast two-hybrid system and phage display, which use either the E3 or the substrate as a bait to fish one another from a library of proteins incorporated into the host cells/phages. However these methods can give rise to a high number of false positive/negative identifications. For example, mismatch of the physiological backgrounds may cause a mammalian protein to be misfolded or not properly modified in yeast or phage, therefore failing to capture the real binders. Moreover, the interactions between E3s and substrates are proved to dissociate rapidly; hence they are not necessarily well detected by the secondary reporter/readout systems employed in these methods. Despite successful examples such as the KLHL20-DAPK1 complex identified by the yeast two-hybrid system [68], these conventional methods are less favoured by researchers nowadays due to the transient nature of E3-substrate interaction, as well as limitations of the technique itself.

The emergence of high-sensitivity mass spectrometry technology develops a bottom-up proteomics approach to substrate identification for E3s (Figure 1.9) [134]. The progress in MS and liquid chromatography instrumentation as well as software and sample preparation techniques have substantially improved data processing speed and sensitivity, allowing the characterization of low abundant proteins. The tandem mass spectrometry (MS/MS) analysis in the bottom-up proteomics enables identification of accurate peptide sequences in the protein sample. Current applications of this technique fall into two categories – E3-substrate interaction based and ubiquitination level based [135]. E3-substrate interaction based proteomics is widely employed in substrate identification for BTB-Kelch proteins, with studies on KLHL12 and

KBTBD8 as examples [53, 86]. Substrates are co-immunoprecipitated with an E3 modified with a tag epitope, and then identified by MS/MS analysis. The hits, however, also include proteins that bind to E3s as regulators or adaptors, other than substrates. Complementary to this, the development of Tandem Ubiquitin-Binding Entities (TUBEs) provides more a direct approach to monitor the ubiquitination level change upon E3 activity perturbation [136]. With TUBE pull-down, ubiquitinated proteins are enriched, then digested and analysed by quantitative MS. The most commonly used TUBE is the ubiquitin-binding domain UBA, which binds to all ubiquitin chains. However TUBEs for specific chain types are also available, such as K29 (Trabid NZF) and K48 (MINDY-1 tUIM) [137, 138]. Furthermore, di-Gly proteomics can be used to identify ubiquitination sites on a substrate [139]. In the digestion step, trypsin cleaves proteins at the carboxyl side of Arg and Lys, which leaves a Gly-Gly motif attached to the ϵ -amine of the modified lysine. The ubiquitinated substrate peptide is then enriched by di-Gly antibody prior to MS identification. The development of these diverse methods enables unbiased E3 substrate profiling under physiological conditions, however, also results in an accumulation of ubiquitinome data that require large amounts of experimental validation and investigation.



(From Heap et al., 2017)

Figure 1.9 The general workflow of a peptide-based (bottom-up) proteomics strategy. Proteins are extracted from cells or tissues and digested by a proteolytic enzyme (e.g. trypsin). Peptides from previous digestion step are separated by High-Performance Liquid Chromatography (HPLC) and then ionized by electrospray ionization (ESI). After undergoing the 'survey scan', the peptides located at certain m/z reading window are selected for fragmentation by tandem mass spectrometry (MS/MS). The resulting spectral data are searched against protein database of certain organism for sequence identification. [134]

The substrates identified can be validated by ubiquitination assay and E3-substrate binding characterization. In the past, several studies employed co-immunoprecipitation of truncated proteins to map the interacting regions of both BTB-Kelch proteins and their substrates [56]. Recently peptide arrays have also been used to map the interaction motifs at a higher resolution [68, 69, 71, 87]. The biochemical and biophysical features of the binding event are characterized using assays such as Isothermal Titration Calorimetry (ITC) and Fluorescence Polarization (FP) [39, 40]. X-ray crystallography captures a snapshot of the substrate-E3 binding and provides the molecular basis of the substrate recruitment [39, 40]. The model predicted from the structural data then can be confirmed with cellular binding/activity assays incorporating mutagenesis. By employing these powerful methods, I'll be aiming to characterize the substrate recruitment for KLHL20, KLHL12, KLHL3 and KBTBD4.

1.5. Aims of this thesis

The aim of this thesis was to characterize the substrate recruitment of BTB-Kelch proteins, with the focus on four representative members KLHL20, KLHL12, KLHL3 and KBTBD4. So far only a few BTB-Kelch proteins have been well characterized. Given there are only a handful of conserved positions that define the Kelch repeat, their identification with bioinformatics servers remain highly challenging. In addition, the structural elements that determine the fold, the beta-strands, are especially short. This makes secondary structure predictions less reliable. Thus, it can be difficult to build homology models as well as to accurately predict the domain boundaries for construct design. For this reason, solving the structures of the Kelch domains is important to provide functional insights such as the effects of disease mutations and substrate interactions.

A generic workflow for characterising BTB-Kelch proteins has been generated in this thesis (Figure 1.10), including substrate identification and characterisation employing various techniques mentioned in section 1.4. With the results presented in this thesis, I wish to extend the understanding of the diverse substrate-binding modes in the BTB-Kelch family. The comprehensive biochemical and structural data will also provide insights for structure-based drug design, as well as prediction of putative substrates.

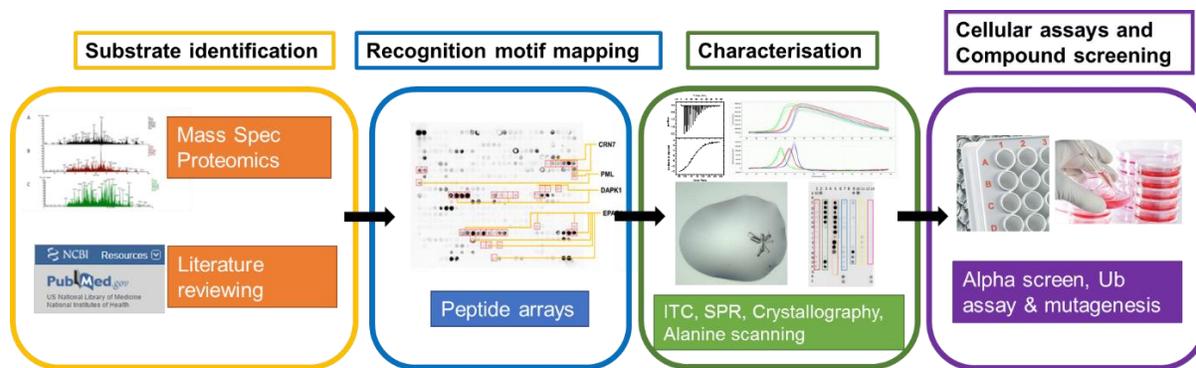


Figure 1.10 Generic workflow for characterising BTB-Kelch proteins. Firstly, substrates of a BTB-Kelch protein are identified by mass spectrometry proteomics or literature reviewing. Secondly, the recognition motif in a substrate is mapped by the SPOTs peptide arrays. Thirdly, the interaction of a substrate with a BTB-Kelch protein is characterised with biophysical and biochemical methods. Fourthly, the in vitro findings are cross-validated in cells and can generate tools for drug design.

1.6. References

1. Komander, D. and M. Rape, *The ubiquitin code*. *Annu Rev Biochem*, 2012. **81**: p. 203-29.
2. Yau, R. and M. Rape, *The increasing complexity of the ubiquitin code*. *Nat Cell Biol*, 2016. **18**(6): p. 579-86.
3. Rock, K.L., et al., *Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules*. *Cell*, 1994. **78**(5): p. 761-771.
4. Akutsu, M., I. Dikic, and A. Bremm, *Ubiquitin chain diversity at a glance*. *J Cell Sci*, 2016. **129**(5): p. 875-80.
5. Ye, Y. and M. Rape, *Building ubiquitin chains: E2 enzymes at work*. *Nature Reviews Molecular Cell Biology*, 2009. **10**: p. 755.
6. Bentley, M.L., et al., *Recognition of Ubch5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex*. *The EMBO journal*, 2011. **30**(16): p. 3285-3297.
7. Hibbert, R.G., et al., *E3 ligase Rad18 promotes monoubiquitination rather than ubiquitin chain formation by E2 enzyme Rad6*. *Proceedings of the National Academy of Sciences of the United States of America*, 2011. **108**(14): p. 5590-5595.
8. Husnjak, K. and I. Dikic, *Ubiquitin-Binding Proteins: Decoders of Ubiquitin-Mediated Cellular Functions*. *Annual Review of Biochemistry*, 2012. **81**(1): p. 291-322.
9. Hamazaki, J., S. Hirayama, and S. Murata, *Redundant Roles of Rpn10 and Rpn13 in Recognition of Ubiquitinated Proteins and Cellular Homeostasis*. *PLOS Genetics*, 2015. **11**(7): p. e1005401.
10. Peth, A., T. Uchiki, and A.L. Goldberg, *ATP-Dependent Steps in the Binding of Ubiquitin Conjugates to the 26S Proteasome that Commit to Degradation*. *Molecular Cell*, 2010. **40**(4): p. 671-681.
11. Husnjak, K., et al., *Proteasome subunit Rpn13 is a novel ubiquitin receptor*. *Nature*, 2008. **453**(7194): p. 481-488.
12. Schreiner, P., et al., *Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction*. *Nature*, 2008. **453**(7194): p. 548-552.

13. Clague, M.J., et al., *Deubiquitylases From Genes to Organism*. Physiological Reviews, 2013. **93**(3): p. 1289-1315.
14. Mevissen, T.E.T. and D. Komander, *Mechanisms of Deubiquitinase Specificity and Regulation*. Annual Review of Biochemistry, 2017. **86**(1): p. 159-192.
15. Zheng, N. and N. Shabek, *Ubiquitin Ligases: Structure, Function, and Regulation*. Annu Rev Biochem, 2017. **86**: p. 129-157.
16. Deshaies, R.J. and C.A.P. Joazeiro, *RING Domain E3 Ubiquitin Ligases*. Annual Review of Biochemistry, 2009. **78**(1): p. 399-434.
17. Ohi, M.D., et al., *Structural insights into the U-box, a domain associated with multi-ubiquitination*. Nature structural biology, 2003. **10**(4): p. 250-255.
18. Rotin, D. and S. Kumar, *Physiological functions of the HECT family of ubiquitin ligases*. Nature Reviews Molecular Cell Biology, 2009. **10**: p. 398.
19. Wenzel, D.M., et al., *UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids*. Nature, 2011. **474**(7349): p. 105-108.
20. Spratt, D.E., H. Walden, and G.S. Shaw, *RBR E3 ubiquitin ligases: new structures, new insights, new questions*. The Biochemical journal, 2014. **458**(3): p. 421-437.
21. Rieser, E., S.M. Cordier, and H. Walczak, *Linear ubiquitination: a newly discovered regulator of cell signalling*. Trends in Biochemical Sciences, 2013. **38**(2): p. 94-102.
22. Zimmerman, E.S., B.A. Schulman, and N. Zheng, *Structural assembly of cullin-RING ubiquitin ligase complexes*. Curr Opin Struct Biol, 2010. **20**(6): p. 714-21.
23. Duda, D.M., et al., *Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation*. Cell, 2008. **134**(6): p. 995-1006.
24. Rabut, G. and M. Peter, *Function and regulation of protein neddylation. 'Protein modifications: beyond the usual suspects' review series*. EMBO reports, 2008. **9**(10): p. 969-976.
25. Lyapina, S., et al., *Promotion of NEDD8-CUL1 Conjugate Cleavage by COP9 Signalosome*. Science, 2001. **292**(5520): p. 1382.
26. Soucy, T.A., et al., *An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer*. Nature, 2009. **458**: p. 732.
27. Brownell, J.E., et al., *Substrate-Assisted Inhibition of Ubiquitin-like Protein-Activating Enzymes: The NEDD8 E1 Inhibitor MLN4924 Forms a NEDD8-AMP Mimetic In Situ*. Molecular Cell, 2010. **37**(1): p. 102-111.
28. Pintard, L., A. Willems, and M. Peter, *Cullin-based ubiquitin ligases: Cul3-BTB complexes join the family*. The EMBO journal, 2004. **23**(8): p. 1681-1687.
29. Canning, P., et al., *Structural basis for Cul3 protein assembly with the BTB-Kelch family of E3 ubiquitin ligases*. J Biol Chem, 2013. **288**(11): p. 7803-14.
30. Genschik, P., I. Sumara, and E. Lechner, *The emerging family of CULLIN3-RING ubiquitin ligases (CRL3s): cellular functions and disease implications*. EMBO J, 2013. **32**(17): p. 2307-20.
31. Chen, Z., et al., *Structural Basis for Recruitment of DAPK1 to the KLHL20 E3 Ligase*. Structure (London, England : 1993), 2019. **27**(9): p. 1395-1404.e4.
32. Li, X., et al., *Crystal Structure of the Kelch Domain of Human Keap1*. Journal of Biological Chemistry, 2004. **279**(52): p. 54750-54758.
33. Zhang, D.D., et al., *Keap1 is a redox-regulated substrate adaptor protein for a Cul3-dependent ubiquitin ligase complex*. Molecular and cellular biology, 2004. **24**(24): p. 10941-10953.
34. Kobayashi, A., et al., *Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2*. Mol Cell Biol, 2004. **24**(16): p. 7130-9.

35. Furukawa, M. and Y. Xiong, *BTB protein Keap1 targets antioxidant transcription factor Nrf2 for ubiquitination by the Cullin 3-Roc1 ligase*. Mol Cell Biol, 2005. **25**(1): p. 162-71.
36. McMahon, M., et al., *Keap1-dependent proteasomal degradation of transcription factor Nrf2 contributes to the negative regulation of antioxidant response element-driven gene expression*. J Biol Chem, 2003. **278**(24): p. 21592-600.
37. Lo, S.-C., et al., *Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling*. The EMBO journal, 2006. **25**(15): p. 3605-3617.
38. Canning, P., F.J. Sorrell, and A.N. Bullock, *Structural basis of Keap1 interactions with Nrf2*. Free radical biology & medicine, 2015. **88**(Pt B): p. 101-107.
39. Fukutomi, T., et al., *Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1*. Molecular and cellular biology, 2014. **34**(5): p. 832-846.
40. Schumacher, F.R., et al., *Structural and biochemical characterization of the KLHL3-WNK kinase interaction important in blood pressure regulation*. Biochem J, 2014. **460**(2): p. 237-46.
41. Chen, H.Y. and R.H. Chen, *Cullin 3 Ubiquitin Ligases in Cancer Biology: Functions and Therapeutic Implications*. Front Oncol, 2016. **6**: p. 113.
42. Cheng, J., et al., *Functional analysis of Cullin 3 E3 ligases in tumorigenesis*. Biochim Biophys Acta Rev Cancer, 2018. **1869**(1): p. 11-28.
43. Dubiel, W., et al., *Cullin 3-Based Ubiquitin Ligases as Master Regulators of Mammalian Cell Differentiation*. Trends Biochem Sci, 2018. **43**(2): p. 95-107.
44. Jerabkova, K. and I. Sumara, *Cullin 3, a cellular scripter of the non-proteolytic ubiquitin code*. Semin Cell Dev Biol, 2018.
45. Tseng, L.A. and J.L. Bixby, *Interaction of an intracellular pentraxin with a BTB-Kelch protein is associated with ubiquitylation, aggregation and neuronal apoptosis*. Molecular and cellular neurosciences, 2011. **47**(4): p. 254-264.
46. Gong, Y., et al., *KLHL3 regulates paracellular chloride transport in the kidney by ubiquitination of claudin-8*. Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(14): p. 4340-4345.
47. Wang, L., et al., *cMyBP-C was decreased via KLHL3-mediated proteasomal degradation in congenital heart diseases*. Experimental Cell Research, 2017. **355**(1): p. 18-25.
48. Choi, J., et al., *Loss of KLHL6 promotes diffuse large B-cell lymphoma growth and survival by stabilizing the mRNA decay factor roquin2*. Nature cell biology, 2018. **20**(5): p. 586-596.
49. Ying, M., et al., *Ubiquitin-dependent degradation of CDK2 drives the therapeutic differentiation of AML by targeting PRDX2*. Blood, 2018. **131**(24): p. 2698.
50. Kim, J., et al., *KLHL7 promotes TUT1 ubiquitination associated with nucleolar integrity: Implications for retinitis pigmentosa*. Biochemical and Biophysical Research Communications, 2017. **494**(1): p. 220-226.
51. Nam, S., et al., *Control of rapsyn stability by the CUL-3-containing E3 ligase complex*. The Journal of biological chemistry, 2009. **284**(12): p. 8195-8206.
52. Sumara, I., et al., *A Cul3-Based E3 Ligase Removes Aurora B from Mitotic Chromosomes, Regulating Mitotic Progression and Completion of Cytokinesis in Human Cells*. Developmental Cell, 2007. **12**(6): p. 887-900.
53. Jin, L., et al., *Ubiquitin-dependent regulation of COPII coat size and function*. Nature, 2012. **482**(7386): p. 495-500.
54. Rondou, P., et al., *BTB Protein KLHL12 targets the dopamine D4 receptor for ubiquitination by a Cul3-based E3 ligase*. J Biol Chem, 2008. **283**(17): p. 11083-96.

55. Rondou, P., et al., *KLHL12-mediated ubiquitination of the dopamine D4 receptor does not target the receptor for degradation*. Cell Signal, 2010. **22**(6): p. 900-13.
56. Angers, S., et al., *The KLHL12-Cullin-3 ubiquitin ligase negatively regulates the Wnt-beta-catenin pathway by targeting Dishevelled for degradation*. Nat Cell Biol, 2006. **8**(4): p. 348-57.
57. Ferretti, L.P., et al., *Cullin3-KLHL15 ubiquitin ligase mediates CtIP protein turnover to fine-tune DNA-end resection*. Nat Commun, 2016. **7**: p. 12628.
58. Allen, E., et al., *Gigaxonin-controlled degradation of MAP1B light chain is critical to neuronal survival*. Nature, 2005. **438**(7065): p. 224-228.
59. Wang, W., et al., *Gigaxonin Interacts with Tubulin Folding Cofactor B and Controls Its Degradation through the Ubiquitin-Proteasome Pathway*. Current Biology, 2005. **15**(22): p. 2050-2055.
60. Salinas, G.D., et al., *Actinfilin is a Cul3 substrate adaptor, linking GluR6 kainate receptor subunits to the ubiquitin-proteasome pathway*. J Biol Chem, 2006. **281**(52): p. 40164-73.
61. Moghe, S., et al., *The CUL3-KLHL18 ligase regulates mitotic entry and ubiquitylates Aurora-A*. Biology open, 2012. **1**(2): p. 82-91.
62. Mulvaney, K.M., et al., *Identification and Characterization of MCM3 as a Kelch-like ECH-associated Protein 1 (KEAP1) Substrate*. The Journal of biological chemistry, 2016. **291**(45): p. 23719-23733.
63. Lee, D.-F., et al., *KEAP1 E3 ligase-mediated downregulation of NF-kappaB signaling by targeting IKKbeta*. Molecular cell, 2009. **36**(1): p. 131-140.
64. Lo, S.-C. and M. Hannink, *PGAM5, a Bcl-XL-interacting Protein, Is a Novel Substrate for the Redox-regulated Keap1-dependent Ubiquitin Ligase Complex*. Journal of Biological Chemistry, 2006. **281**(49): p. 37893-37903.
65. Lee, Y., et al., *Keap1/Cullin3 Modulates p62/SQSTM1 Activity via UBA Domain Ubiquitination*. Cell reports, 2017. **19**(1): p. 188-202.
66. Komatsu, M., et al., *The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1*. Nature Cell Biology, 2010. **12**: p. 213.
67. Orthwein, A., et al., *A mechanism for the suppression of homologous recombination in G1 cells*. Nature, 2015. **528**(7582): p. 422-426.
68. Lee, Y.R., et al., *The Cullin 3 substrate adaptor KLHL20 mediates DAPK ubiquitination to control interferon responses*. EMBO J, 2010. **29**(10): p. 1748-61.
69. Yuan, W.-C., et al., *A Cullin3-KLHL20 Ubiquitin Ligase-Dependent Pathway Targets PML to Potentiate HIF-1 Signaling and Prostate Cancer Progression*. Cancer Cell, 2011. **20**(2): p. 214-228.
70. Liu, C.C. and R.H. Chen, *KLHL20 links the ubiquitin-proteasome system to autophagy termination*. Autophagy, 2016. **12**(5): p. 890-1.
71. Lin, M.Y., et al., *PDZ-RhoGEF ubiquitination by Cullin3-KLHL20 controls neurotrophin-induced neurite outgrowth*. J Cell Biol, 2011. **193**(6): p. 985-94.
72. Yuan, W.C., et al., *K33-Linked Polyubiquitination of Coronin 7 by Cul3-KLHL20 Ubiquitin E3 Ligase Regulates Protein Trafficking*. Mol Cell, 2014. **54**(4): p. 586-600.
73. Higashimura, Y., et al., *Kelch-like 20 up-regulates the expression of hypoxia-inducible factor-2alpha through hypoxia- and von Hippel-Lindau tumor suppressor protein-independent regulatory mechanisms*. Biochem Biophys Res Commun, 2011. **413**(2): p. 201-5.
74. Maerki, S., et al., *The Cul3-KLHL21 E3 ubiquitin ligase targets aurora B to midzone microtubules in anaphase and is required for cytokinesis*. J Cell Biol, 2009. **187**(6): p. 791-800.

75. Courtheoux, T., et al., *Cortical dynamics during cell motility are regulated by CRL3(KLHL21) E3 ubiquitin ligase*. Nature communications, 2016. **7**: p. 12810-12810.
76. Beck, J., et al., *Ubiquitylation-dependent localization of PLK1 in mitosis*. Nature Cell Biology, 2013. **15**: p. 430.
77. Chen, J., et al., *KLHL22 activates amino-acid-dependent mTORC1 signalling to promote tumorigenesis and ageing*. Nature, 2018. **557**(7706): p. 585-589.
78. Lin, Z., et al., *Stabilizing mutations of KLHL24 ubiquitin ligase cause loss of keratin 14 and human skin fragility*. Nature Genetics, 2016. **48**: p. 1508.
79. Zhang, C., et al., *Cullin3-KLHL25 ubiquitin ligase targets ACLY for degradation to inhibit lipid synthesis and tumor progression*. Genes & development, 2016. **30**(17): p. 1956-1970.
80. Miller, W.P., et al., *The Translational Repressor 4E-BP1 Contributes to Diabetes-Induced Visual Dysfunction*. Investigative ophthalmology & visual science, 2016. **57**(3): p. 1327-1337.
81. Papizan, J.B., et al., *Deficiency in Kelch protein Khl31 causes congenital myopathy in mice*. The Journal of clinical investigation, 2017. **127**(10): p. 3730-3740.
82. Zhang, Z., et al., *Insulin resistance and diabetes caused by genetic or diet-induced KBTBD2 deficiency in mice*. Proceedings of the National Academy of Sciences, 2016. **113**(42): p. E6418.
83. Gong, W., et al., *Kbtbd5 regulates skeletal muscle myogenesis through the E2F1-DP1 complex*. Journal of Biological Chemistry, 2015.
84. Garg, A., et al., *KLHL40 deficiency destabilizes thin filament proteins and promotes nemaline myopathy*. The Journal of Clinical Investigation, 2014. **124**(8): p. 3529-3539.
85. Genau, H.M., et al., *CUL3-KBTBD6/KBTBD7 ubiquitin ligase cooperates with GABARAP proteins to spatially restrict TIAM1-RAC1 signaling*. Mol Cell, 2015. **57**(6): p. 995-1010.
86. Werner, A., et al., *Cell-fate determination by ubiquitin-dependent regulation of translation*. Nature, 2015. **525**(7570): p. 523-7.
87. Gray, C.H., et al., *Novel beta-propeller of the BTB-Kelch protein Krp1 provides a binding site for Lasp-1 that is necessary for pseudopodial extension*. J Biol Chem, 2009. **284**(44): p. 30498-507.
88. Narahara, S., et al., *KBTBD11, a novel BTB-Kelch protein, is a negative regulator of osteoclastogenesis through controlling Cullin3-mediated ubiquitination of NFATc1*. Scientific reports, 2019. **9**(1): p. 3523-3523.
89. Cummings, C.M., et al., *The Cul3/Klhdc5 E3 ligase regulates p60/katanin and is required for normal mitosis in mammalian cells*. J Biol Chem, 2009. **284**(17): p. 11663-75.
90. Chen, W.-L., et al., *SCA8 mRNA expression suggests an antisense regulation of KLHL1 and correlates to SCA8 pathology*. Brain Research, 2008. **1233**: p. 176-184.
91. Zhang, C., et al., *Degradation by Cullin 3 and effect on WNK kinases suggest a role of KLHL2 in the pathogenesis of Familial Hyperkalemic Hypertension*. Biochemical and biophysical research communications, 2016. **469**(1): p. 44-48.
92. Boyden, L.M., et al., *Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities*. Nature, 2012. **482**(7383): p. 98-102.
93. Braybrook, C., et al., *Identification and Characterization of KLHL4, a Novel Human Homologue of the Drosophila Kelch Gene That Maps within the X-Linked Cleft Palate and Ankyloglossia (CPX) Critical Region*. Genomics, 2001. **72**(2): p. 128-136.
94. Kurozumi, S., et al., *Clinical and biological roles of Kelch-like family member 7 in breast cancer: a marker of poor prognosis*. Breast Cancer Research and Treatment, 2018. **170**(3): p. 525-533.

95. Lee, B., et al., *Homozygous deletions at 3p22, 5p14, 6q15, and 9p21 result in aberrant expression of tumor suppressor genes in gastric cancer*. *Genes, Chromosomes and Cancer*, 2015. **54**(3): p. 142-155.
96. Yatsenko, A.N., et al., *Non-invasive genetic diagnosis of male infertility using spermatozoal RNA: KLHL10 mutations in oligozoospermic patients impair homodimerization*. *Human Molecular Genetics*, 2006. **15**(23): p. 3411-3419.
97. Schabhüttl, M., et al., *Whole-exome sequencing in patients with inherited neuropathies: outcome and challenges*. *Journal of Neurology*, 2014. **261**(5): p. 970-982.
98. Vater, I., et al., *The mutational pattern of primary lymphoma of the central nervous system determined by whole-exome sequencing*. *Leukemia*, 2014. **29**: p. 677.
99. Kang, J.J., et al., *A review of gigaxonin mutations in giant axonal neuropathy (GAN) and cancer*. *Human Genetics*, 2016. **135**(7): p. 675-684.
100. Paciorkowski, A.R., et al., *Copy number variants and infantile spasms: evidence for abnormalities in ventral forebrain development and pathways of synaptic function*. *European journal of human genetics : EJHG*, 2011. **19**(12): p. 1238-1245.
101. Yoo, N.J., et al., *Somatic mutations of the KEAP1 gene in common solid cancers*. *Histopathology*, 2012. **60**(6): p. 943-952.
102. Shi, L., et al., *KLHL21, a novel gene that contributes to the progression of hepatocellular carcinoma*. *BMC cancer*, 2016. **16**(1): p. 815-815.
103. Choi, E.-S., et al., *Overexpression of KLHL23 protein from read-through transcription of PHOSPHO2-KLHL23 in gastric cancer increases cell proliferation*. *FEBS open bio*, 2016. **6**(11): p. 1155-1164.
104. Lee, J.Y.W., et al., *Mutations in KLHL24 Add to the Molecular Heterogeneity of Epidermolysis Bullosa Simplex*. *Journal of Investigative Dermatology*, 2017. **137**(6): p. 1378-1380.
105. Shitani, M., et al., *Genome-wide analysis of DNA methylation identifies novel cancer-related genes in hepatocellular carcinoma*. *Tumor Biology*, 2012. **33**(5): p. 1307-1317.
106. Morris, M.R., et al., *Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma*. *Oncogene*, 2010. **30**: p. 1390.
107. Bonazzi, V.F., D. Irwin, and N.K. Hayward, *Identification of candidate tumor suppressor genes inactivated by promoter methylation in melanoma*. *Genes, Chromosomes and Cancer*, 2009. **48**(1): p. 10-21.
108. Fujita, M., et al., *Up-Regulation of the Ectodermal-Neural Cortex 1 (ENC1) Gene, a Downstream Target of the β -Catenin/T-Cell Factor Complex, in Colorectal Carcinomas*. *Cancer Research*, 2001. **61**(21): p. 7722.
109. Hammarsund, M., et al., *Disruption of a novel ectodermal neural cortex 1 antisense gene, ENC-1AS and identification of ENC-1 overexpression in hairy cell leukemia*. *Human Molecular Genetics*, 2004. **13**(23): p. 2925-2936.
110. Hernandez, M.-C., et al., *Cloning of Human ENC-1 and Evaluation of Its Expression and Regulation in Nervous System Tumors*. *Experimental Cell Research*, 1998. **242**(2): p. 470-477.
111. Lee, J.C., et al., *Recurrent KBTBD4 small in-frame insertions and absence of DROSHA deletion or DICER1 mutation differentiate pineal parenchymal tumor of intermediate differentiation (PPTID) from pineoblastoma*. *Acta Neuropathologica*, 2019. **137**(5): p. 851-854.
112. Northcott, P.A., et al., *The whole-genome landscape of medulloblastoma subtypes*. *Nature*, 2017. **547**(7663): p. 311-317.

113. Ravenscroft, G., et al., *Mutations in KLHL40 are a frequent cause of severe autosomal-recessive nemaline myopathy*. American journal of human genetics, 2013. **93**(1): p. 6-18.
114. Gupta, V.A., et al., *Identification of KLHL41 Mutations Implicates BTB-Kelch-Mediated Ubiquitination as an Alternate Pathway to Myofibrillar Disruption in Nemaline Myopathy*. American journal of human genetics, 2013. **93**(6): p. 1108-1117.
115. Sambuughin, N., et al., *Dominant mutations in KBTBD13, a member of the BTB/Kelch family, cause nemaline myopathy with cores*. American journal of human genetics, 2010. **87**(6): p. 842-847.
116. Chen, H.Y., C.C. Liu, and R.H. Chen, *Cul3-KLHL20 ubiquitin ligase: physiological functions, stress responses, and disease implications*. Cell Div, 2016. **11**: p. 5.
117. Liu, C.C., et al., *Cul3-KLHL20 Ubiquitin Ligase Governs the Turnover of ULK1 and VPS34 Complexes to Control Autophagy Termination*. Mol Cell, 2016. **61**(1): p. 84-97.
118. Gomez Ravetti, M., et al., *Uncovering molecular biomarkers that correlate cognitive decline with the changes of hippocampus' gene expression profiles in Alzheimer's disease*. PLoS One, 2010. **5**(4): p. e10153.
119. Arefin, A.S., et al., *Unveiling clusters of RNA transcript pairs associated with markers of Alzheimer's disease progression*. PLoS One, 2012. **7**(9): p. e45535.
120. McGourty, C.A., et al., *Regulation of the CUL3 Ubiquitin Ligase by a Calcium-Dependent Co-adaptor*. Cell, 2016. **167**(2): p. 525-538 e14.
121. Mena, E.L., et al., *Dimerization quality control ensures neuronal development and survival*. Science, 2018. **362**(6411): p. eaap8236.
122. Skieterska, K., et al., *KLHL12 Promotes Non-Lysine Ubiquitination of the Dopamine Receptors D4.2 and D4.4, but Not of the ADHD-Associated D4.7 Variant*. PLoS One, 2015. **10**(12): p. e0145654.
123. Shibata, S., et al., *Kelch-like 3 and Cullin 3 regulate electrolyte homeostasis via ubiquitination and degradation of WNK4*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(19): p. 7838-7843.
124. Richardson, C. and D.R. Alessi, *The regulation of salt transport and blood pressure by the WNK-SPAK/OSR1 signalling pathway*. Journal of Cell Science, 2008. **121**(20): p. 3293.
125. Wilson, F.H., et al., *Human Hypertension Caused by Mutations in WNK Kinases*. Science, 2001. **293**(5532): p. 1107.
126. Wu, G. and J.-B. Peng, *Disease-causing mutations in KLHL3 impair its effect on WNK4 degradation*. FEBS letters, 2013. **587**(12): p. 1717-1722.
127. Mori, Y., et al., *Decrease of WNK4 ubiquitination by disease-causing mutations of KLHL3 through different molecular mechanisms*. Biochemical and Biophysical Research Communications, 2013. **439**(1): p. 30-34.
128. Wakabayashi, M., et al., *Impaired KLHL3-Mediated Ubiquitination of WNK4 Causes Human Hypertension*. Cell Reports, 2013. **3**(3): p. 858-868.
129. Ohta, A., et al., *The CUL3-KLHL3 E3 ligase complex mutated in Gordon's hypertension syndrome interacts with and ubiquitylates WNK isoforms: disease-causing mutations in KLHL3 and WNK4 disrupt interaction*. The Biochemical journal, 2013. **451**(1): p. 111-122.
130. Sasaki, E., et al., *KLHL3 Knockout Mice Reveal the Physiological Role of KLHL3 and the Pathophysiology of Pseudohypoaldosteronism Type II Caused by Mutant KLHL3*. Molecular and Cellular Biology, 2017. **37**(7): p. e00508-16.
131. Susa, K., et al., *Impaired degradation of WNK1 and WNK4 kinases causes PHAI1 in mutant KLHL3 knock-in mice*. Human Molecular Genetics, 2014. **23**(19): p. 5052-5060.

132. Alessi, D.R., et al., *The WNK-SPAK/OSR1 pathway: Master regulator of cation-chloride cotransporters*. *Science Signaling*, 2014. **7**(334): p. re3.
133. Kasagi, Y., et al., *Impaired degradation of medullary WNK4 in the kidneys of KLHL2 knockout mice*. *Biochemical and Biophysical Research Communications*, 2017. **487**(2): p. 368-374.
134. Heap, R.E., et al., *Mass spectrometry techniques for studying the ubiquitin system*. *Biochem Soc Trans*, 2017. **45**(5): p. 1137-1148.
135. Beaudette, P., O. Popp, and G. Dittmar, *Proteomic techniques to probe the ubiquitin landscape*. *Proteomics*, 2016. **16**(2): p. 273-87.
136. Hjerpe, R., et al., *Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities*. *EMBO reports*, 2009. **10**(11): p. 1250-1258.
137. Kristariyanto, Y.A., et al., *K29-selective ubiquitin binding domain reveals structural basis of specificity and heterotypic nature of k29 polyubiquitin*. *Molecular cell*, 2015. **58**(1): p. 83-94.
138. Kulathu, Y., et al., *Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain*. *Nature Structural & Molecular Biology*, 2009. **16**: p. 1328.
139. Marotti, L.A., Jr., et al., *Direct identification of a G protein ubiquitination site by mass spectrometry*. *Biochemistry*, 2002(0006-2960 (Print)).

Chapter 2

Methods and Materials

2.1. Construct cloning

2.1.1. Construct design

Constructs for KLHL3, KLHL12, KLHL17 and KLHL20 used in crystallisation were designed before my DPhil programme started. New constructs (see Appendix) were designed for follow-up functional study and new findings for protein of interests. The sources of cDNA include Mammalian Genomics Collection (MGC), Addgene, DNASU plasmid repository and reverse-transcription from HEK293T mRNA template. Truncation series were designed based on secondary structure prediction using bioinformatics tools such as Globplot (<http://globplot.embl.de/>) [1] and Psipred (<http://bioinf.cs.ucl.ac.uk/psipred/>) [2]. The constructs were made using two methods – ligation independent cloning and restriction enzyme cloning.

2.1.2. Ligation Independent Cloning (LIC)

The LIC method is based on the exonuclease activity of T4 DNA polymerase. A 12-16 base pair overhang was added to polymerase chain reaction (PCR) primers for the T4 DNA polymerase to cut to create sticky ends for ligase-independent annealing [3].

Before conducting cloning, DNA template was diluted to 2.5 ng/μL, while forward and reverse primers were mixed and diluted to 5 μM each. PCR was performed with Herculase II Fusion DNA polymerase (Agilent, 600675) employing the touch-down PCR condition [4].

95°C, 10 min
 (95°C, 30 sec; 68°C, 30 sec; 68°C, 1-3 min*) x5 cycles
 (95°C, 30 sec; 60°C, 30 sec; 68°C, 1-3 min*) x5 cycles
 (95°C, 30 sec; 55°C, 30 sec; 68°C, 1-3 min*) x5 cycles
 (95°C, 30 sec; 50°C, 30 sec; 68°C, 1-3 min*) x10 cycles
 68°C, 10 min
 15°C hold.

The PCR product was analysed on an Invitrogen pre-cast ethidium bromide gel. For PCR optimisation, additives such as DMSO or glycerol was added to facilitate primer-template annealing. In the case where the template construct and cloning vector shared the same antibiotics, the PCR product was treated with DpnI to remove the template DNA. Then all PCR product was cleaned up using the mini-columns PCR purification kit (Qiagen, 28104). The vectors were digested with BsaI as indicated in the vector maps (show below) for 2hours in 37°C. The cut vectors were purified using the PCR purification kit.

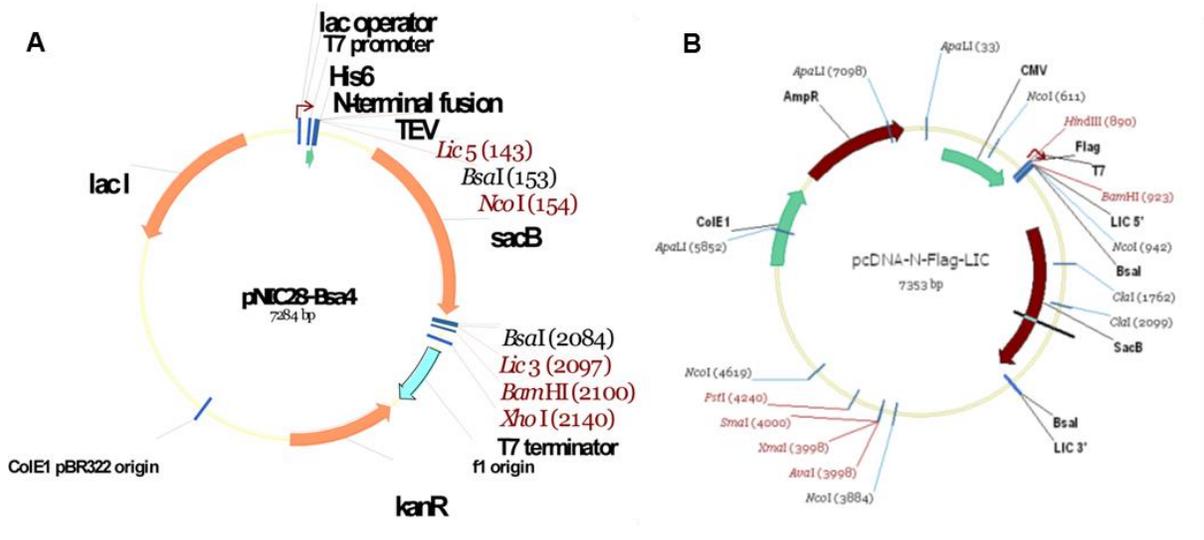


Figure 2.1 LIC-modified vectors (figures taken from <http://www.thesgc.org/reagents/vectors>. Credit to Dr P. Savitsky, Dr O.Gileadi and Dr G. Kochan)

To create annealing sticky ends, T4 polymerase was added to vectors in presence of 25 mM dGTP. Similarly, T4 polymerase was added to PCR product in presence of 25 mM dCTP. This reaction was incubated for 30 minutes at 22°C, then inactivated at 75°C for 20 minutes in a thermocycler. For annealing, 1 µL vector was mixed with 2 µL PCR insertion and incubated at 22°C for at least 30min, followed by transformation into MACH I cells with 45 sec heatshock at 42°C. Transformed cells were plated out onto LB Agar with appropriate antibiotics containing 5% sucrose for SacB selection.

2.1.3. Restriction enzyme cloning (REC)

Restriction enzymes (restriction endonucleases) cut DNA at specific recognition sites and create sticky ends. They cut both strands of the target DNA at different spots creating 3'- or 5'-overhangs of 1 to 4 nucleotides. To be able to clone a DNA insert into a cloning vector, both insert and vector have to be treated with two restriction enzymes to create compatible ends for ligation. The ligation step requires the T4 ligase.

Prior to cloning, restriction enzymes were decided based on the cloning vector as well as the target sequences in NEBcutter (<http://nc2.neb.com/NEBcutter2/>) to make sure there was no in frame cutting sites. The cutting condition and capability of double digest can be assessed with NEB cloner tool <https://nebcloner.neb.com/#!/redigest>. Both forward and reverse primers were then designed with the appropriate restriction site and a 6 mer random sequence at 5' end to ensure efficient DNA cleavage by the restriction enzymes. PCR was performed in the same way as described in 2.1.2. PCR product was cleaned up with PCR purification kit, then treated with appropriate restriction enzymes following the instructions from NEBcloner. Vectors were digested with same restriction enzymes, followed by dephosphorylation with FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific, EF0654) to prevent self-ligation. Both treated insert and vectors were then purified by Gel extraction (Qiagen, 28704). The

ligation was then performed with 50 ng vector, three time molar ratio of insert and 1 μ L T4 ligase in 10 μ L compatible buffer for at least 30 min at 22°C. Ligated constructs were transformed into DH5 α cells (Bioline, BIO-85025).

2.1.4. Two-step mutagenesis [5]

Two-step mutagenesis generates mutations with primers that incorporate 20 mer overhangs at each end and mutated sites in the middle. Primer design graph is shown below.

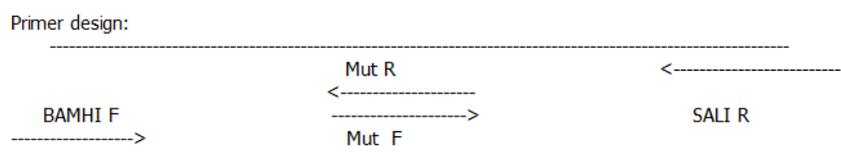


Figure 2.2 Two-step mutagenesis primer design

The first round PCR employed (1) BAMHI F and Mut R and (2) Mut F and SALI R to generate mutated 5' end and 3' end fractions respectively. The PCR product was then treated with DpnI to remove the wild type template. The second round PCR used the first round product as template, and performed the reaction with BAMHI F and SALI R primers. The 20 mer overhangs in Mut R and Mut F facilitated annealing of 5' end and 3' end mutated fractions during PCR. The PCR product of second round was then processed with standard cloning steps.

2.1.5. Competent cells preparation

2.1.5.1. Buffers and Media

SOB medium:

1 litre: 20 g bacto tryptone (or Peptone with casein), 5 g bacto yeast extract, 0.5 g NaCl, 2.5 ml of 1 M KCl (MW 74.55)

Autoclave, cool down, then add 10 ml of filter sterilised 2 M MgCl₂ hexahydrate (MW 203.31)

SOC medium:

To 100 ml of SOB, add 2 ml of 1 M (18%) filtered glucose stock (anhydrous glucose MW 180.16)

TFB1:

	Molarity	MW	g/L	g/500mL
Potassium acetate	30 mM	98.15	2.945	1.472
RbCl	100 mM	120.92	12.092	6.046
CaCl ₂ •2H ₂ O	10 mM	147.02	1.470	0.735
**MnCl ₂ •4H ₂ O	50 mM	197.91	9.896	4.948
Glycerol	15% (v/v)		150 mL	75 mL
Water			850 mL	425 mL

**Bring pH to 5.85 with 1 M acetic acid before adding MnCl₂

Filter sterilise.

TFB2:

	Molarity	MW	g/L	g/200mL
MOPS free acid	10 mM	209.30	2.093	0.419
CaCl ₂ •2H ₂ O	75 mM	147.02	11.027	2.205
RbCl	10 mM	120.92	1.209	0.242
Glycerol	15% (v/v)		150 mL	30 mL
Water			850 mL	170 mL

pH to 6.5 with KOH, and then filter sterilise.

Agar plates

25 g LB broth and 15 g Agar were dissolved in 1 L ddH₂O. After autoclaved, agar was cooled down to 40°C, added antibiotics and poured to Petri dishes.

2.1.5.2. Protocols

Competent cells are ready-to-use bacterial cells that possess more easily altered cell walls by which foreign DNA can be passed through easily. With CaCl_2 method, the competency can be obtained by creating pores in bacterial cells by suspending them in a solution containing high concentration of calcium [6].

BL21(DE3)-pRARE Rosetta and MACH I cells were streaked out onto agar plates containing appropriate antibiotics (Chloramphenicol for Rosetta, N/A for MACH I). A single colony for each strain was picked from agar plates and grown in 10 mL SOB medium at 37°C, 250 RPM overnight. On the next day, 1 L of SOB was inoculated with 10 mL overnight culture and grown at 37°C, 180 RPM until OD_{600} reached 0.45-0.55. Cells were harvested by centrifugation at 4000 RPM for 10 min in 4°C. Cell pellets were then washed thoroughly but gently with 320 mL/L icecold TFB1 with 10 mM Ca^{2+} . Once the TFB1 buffer was removed, cells were then resuspended thoroughly but gently in 50 mL/L ice-cold TFB2 containing 75mM Ca^{2+} to create cell wall pores in bacteria. Suspended competent cells were then aliquoted to prechilled Eppendorf tubes and frozen down in liquid nitrogen.

2.1.6. Transformation

An aliquot of competent cells were thawed on ice and dispensed 50 μL into each autoclaved Eppendorf tube. 1 μL plasmid or 3~5 μL ligation product was incubated with 50 μL cells on ice for at least 30 min. Then the mixture was heat-shocked for 45 sec then incubated on ice for 1-2 min. 400 μL plain LB was added to this and the cells were incubated in 37°C with shaking for 1~2 hours. After incubation, cells were plated out on LB agar with appropriate antibiotics. The plates were then incubated at 37°C in a stationary incubator overnight.

2.1.7. RNA extraction and Reverse Transcripase PCR

RNA was extracted from a 10 cm dish of 95% confluent HEK293T cells using RNeasy Mini Kit (Qiagen, 74104). 0.5 µg RNA was converted to cDNA using Superscript III reverse transcriptase (Invitrogen, 18080085) using oligo (dT) primers and following the manufacturer's protocol. The cDNA was used as template DNA in the PCR.

2.2. Protein expression

2.2.1. E.coli expression

Kelch domains that were used in crystallisation and other biophysical studies were expressed from E.coli. The target constructs cloned into pNIC28-Bsa4 vector were transformed into expressing strain BL21(DE3)-R3-pRARE2. To make glycerol stocks for future inoculation, small scale LB culture (up to 5 mL) was inoculated with a single colony with 50 µg/mL Kanamycin and Chloramphenicol and grown for overnight at 37°C, 250 RPM. Glycerol stocks were made by mixing 50% autoclaved glycerol with 50% overnight culture and frozen in -80°C freezer. For 1 L expression, LB culture containing 50µg/mL Kanamycin and Chloramphenicol was inoculated with 5 mL overnight culture, and grown at 37°C with shaking at 180 RPM until OD₆₀₀ reached 0.6. Then the cells were induced with 0.4mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated at 18°C with shaking overnight. In pNIC28-Bsa4 vector, target genes are inserted to 3' end of the lac operator. IPTG interacts with the lac repressor and allosterically releases it from the lac operator, therefore allowing the target genes to transcribe [7]. For targets with low expression, TB culture was used to aim for higher cell density. Expression was induced with 1mM IPTG when OD₆₀₀ reached 1~1.2. Cells were harvested by centrifugation at 4000 RPM in 4°C for 20min. Pellets were resuspended in binding buffer with 1/1000 SET III protease inhibitor cocktail (Millipore, 539134). (See 2.3.1. for binding buffer recipe)

2.2.2. Baculoviral expression

The SGC employs Invitrogen Bac-to-Bac® system which generates recombinant baculovirus by site-specific transposition in DH10Bac™ E. coli strain (Thermo fisher, 10361012). The DH10Bac™ incorporates a parent bacmid with lacZ-mini-attTn7 and a helper plasmid encoding transposition enzymes. When the pFastBac™ plasmid (Thermo fisher, 10584-027) that contains the gene of interest is transformed into DH10Bac™, transposition occurs and the gene of interest is integrated into the bacmid. The successful transposition can be selected through blue/white screening as the recombinant cassette disrupts the lacZ gene expression. The bacmid DNA is isolated and transfected into Sf9 cells using JetPrime® (Polyplus, 114-15) to produce P0 virus under manufacturer's instruction.

Bacmid and P0-P1 viruses were generated prior to my DPhil started. I amplified viruses from P1 to P2 in a 50 mL scale and infected Sf9 insect cells at 2×10^6 /mL with P2 viruses. Cells were grown at 27°C with 100 RPM shaking for 48 to 72 hours. Cells were harvested by centrifugation at 4000 RPM in 4°C for 20min. Pellets were resuspended in binding buffer with 1/1000 SET III protease inhibitor cocktail.

2.2.3. Mammalian expression (HEK293T)

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco/Invitrogen) supplemented with 10% fetal bovine serum (FBS, Sigma Aldrich), 100 U/mL penicillin sodium and 100 Ig/mL streptomycin sulfate (Sigma Aldrich) in a humidified incubator at 37°C with 5% CO₂.

DNA was maxiprepmed for transient transfection. Per 10cm dish, HEK293T cells were transfected with 5 µg DNA and 15-35 µg Polyethylenimine (PEI) at 60% cell confluency, amount of PEI adjusted based on transfection efficiency in test expression. After 40 hours incubation in 37°C stationary incubator with 5% CO₂, cells were washed with ice cold PBS and then

scraped down in 1 mL ice cold PBS in a 1.5 mL Eppendorf tube. Cells were harvested by 'pop spin' (5-10 sec at 10000 RPM) at 4°C. Supernatant was aspirated and the pelleted was either kept in -80°C freezer until lysis, or was lysed right away.

2.3. Protein purification

2.3.1. Buffers

Binding Buffer: 500 mM NaCl, 5% glycerol, 50 mM HEPES pH7.5, 5 mM imidazole

Wash Buffer: 500 mM NaCl, 5% glycerol, 50 mM HEPES pH7.5, 30 mM imidazole

Elution Buffers: 500 mM NaCl, 50 mM HEPES pH7.5, 5% glycerol, 50 ~ 250 mM imidazole

Size exclusion chromatography buffer: 300 mM NaCl, 50 mM HEPES pH7.5, 0.5 mM TCEP.

Mammalian cell lysis buffer (NP40 based): 50 mM Tris pH7.5 Ultrapure, 150 mM NaCl, 0.1% NP40, 10% Glycerol, 1 mM EDTA, 5 mM MgCl₂ (Protease inhibitors and phosphatase inhibitors were added on day of experiment).

2.3.2. Cell lysis

E.coli and insect cell pellets were resuspended in binding buffer with SET III protease inhibitor cocktail. Buffer volume was adjusted with pellet size (usually 20 mL per 1 L culture). Cells were lysed by sonication for 15 min at 35% amplitude (pulsing 5 sec on and 10 sec off to prevent overheating samples) on ice. After sonication, PEI or Benzonase was added in the total lysate to precipitate/remove DNA fibre. Cell debris was then pelleted down by centrifugation at 21500 RPM for 30min in 4°C.

Proteins were extracted from mammalian cells in a NP40 based lysis buffer. This buffer is also suitable for native co-immunoprecipitation (see 2.9.) as the mild ionic force of NP40

solubilises cytoplasmic proteins without interfering with the protein-protein interaction. Cell pellet was resuspended in 3 volume of complete lysis buffer and incubated on ice for 30 min. Cell debris was then pelleted down by centrifugation at maximum speed for 15min in 4°C.

2.3.3. Immobilized metal ion affinity chromatography (IMAC)

Histidine can chelate with metal ions such as Nickel. Thus, Nickel columns are used for IMAC for the purification of recombinant proteins with a polyhistidine tag.

The supernatant from section 2.3.2. was filtered and then put through 5 mL 50% pre-equilibrated Ni slurry column under gravity flow. The column was washed with 10 column volume (CV) binding buffer, followed by 10 CV wash buffer. Bound protein was then eluted with 2 CV elution buffers containing gradient concentrations of imidazole ranging from 50~250 mM. All fractions were collected and analysed on SDS-PAGE.

2.3.4. Size exclusion chromatography

Size exclusion can be used to separate proteins according to their sizes. Protein samples were concentrated using an Amicon Ultra centrifugal filter with a 10 kDa cut-off to up to 5 mL, and loaded onto a pre-equilibrated size-exclusion chromatography column (sephadex 200 HiLoad 16/60 column or sephadex 75 HiLoad 16/60 column) and ran at 1 mL/min flowrate on the AKTA express at 4°C. Fractions were collected and protein was detected by UV absorbance peak. Protein in each fraction was analysis by SDS-PAGE.

2.3.5. Ion-exchange chromatography

Ion exchange separates proteins based on reversible absorption of charged molecules in solution to the opposite charged immobilised group. Depending on the pI of the protein and protein complex, a buffer was chosen with a pH 1-1.5 difference to the pI. After applying protein

onto an ion-exchange column, the protein is bound to the opposite charged column and eluted with an increase of ionic strength in the running buffer. Anion exchange, using a HiTrap Q HP column (GE Healthcare) and cation exchange, using a HiTrap SP column (GE Healthcare) were performed. The column was pre-washed with high salt buffer (1 M NaCl) and then equilibrated with the no salt buffer. The sample was loaded onto a pre-equilibrated column (flow rate 3 mL/min) and a linear elution gradient was run from 0 to 1 M NaCl.

2.3.6. Protein handling

Between each purification stage, 1 mM DTT or 1 mM TCEP was added to prevent protein oxidation. Proteins were either stored at 4°C or on ice during purification.

Purified protein was concentrated to 5~10 mg/mL and flash-frozen in Liquid Nitrogen and then stored in -80°C for future use. Intact Mass Spectrometry was performed for quality control before assays or crystallisation trials.

All LIC vectors used in this thesis produce a TEV cleavage site ENLYFQ*(G/S) between the tag and the target. If desired, TEV protease could be added to cleave the tag by overnight incubation at 4°C or 2 hours at room temperature, at a protease to target protein ratio of 1:100 (w/w). For proteins expressed in insect cells, proteins were dephosphorylated with lambda phosphatase in presence of 1 mM Mn²⁺ at room temperature for 2 hours, at an enzyme to target protein ratio of 1:100 (w/w).

2.4. SPOTs assay

2.4.1. Peptide array design

Cellulose-bound peptide arrays were prepared employing standard Fmoc solid phase peptide synthesis using a MultiPep-RSi-Spotter (INTAVIS, Köln, Germany).

For binding motif screening, the screening region of interest was taken from literature. A peptide of 15 mer in length was printed on each dot in the peptide array. Consecutive dots shared 12 mer overlapping region, with 3 mer forward shift to screen throughout the sequence of the interesting region.

For Alanine scanning, to find the key residues in the peptide-protein interaction, a triple Alanine mutation was shift throughout the sequence of the peptide from N-terminus to C-terminus.

For truncation series, to find the minimal residues in the peptide that were sufficient for protein binding, a series peptides were printed onto array membrane. Residues were chopped from N-terminus or from C-terminus one by one in the series.

2.4.2. Peptide array interaction assay

Peptide array membranes can be stored in -20°C for a few months, and can be rehydrated and activated with 100% ethanol before conducting interaction assay. Membranes were then equilibrated with PBST before being blocked with 5% BSA in PBST for 6 hours. 1 uM his-tagged recombinant proteins were diluted in PBS and probed with membranes for 16 hours (overnight) at 4°C with rocking. Excessive unbound proteins were removed with PBST. HPR-conjugated anti-His antibody (1:3000 dilution) was prepared in 2.5% BSA PBST and incubated with membrane for 1 hour at room temperature. Unbound antibody was washed away with PBST thoroughly. The membrane was developed using an ECL kit containing Peroxide solution and the Luminol Enhancer solution and was imaged using Las-4000 imager (chemiluminescence settings).

2.5. Isothermal titration calorimetry

Isothermal titration calorimetry is a biophysical method that measures the thermodynamics of binding reactions by titrating ligands to macromolecules. The initial measurements directly determine the binding affinity (K_A or K_D), enthalpy (ΔH) and binding stoichiometry (n). From those initial measurements, Gibbs free energy changes (ΔG) can be calculated based on the equation $\Delta G = -RT \ln K_A$, and entropy changes (ΔS) can be calculated using $\Delta G = \Delta H - T\Delta S$. In a word, ITC provides a label free, in solution and informative characterisation for protein-ligand binding reaction [8].

In ITC instrument, the sample and reference cells are surrounded by an adiabatic jacket. The system is able to detect the temperature difference (ΔT) between the sample and reference cells, and to maintain $\Delta T = 0$ by supplying power to both cells via two heaters. The heat change when the interactions occur is measured.

In this thesis, multiple injection method was used. Calorimetric measurements were performed at 15°C using a VP-ITC calorimeter (MicroCal). Samples were dialysed against 50 mM HEPES pH 7.5, 300 mM NaCl and 1 mM TCEP buffer overnight at 4°C in D-tube dialyzer midi tubes, with a 3.5 kDa cut-off (Millipore) and degassed before use. Each titration experiment consisted of an initial (2 μ l) injection followed by a further 19 sequential injections of 15 μ l. Intervals between each injection allowed equilibrium to complete. Heats of dilution were measured in blank titrations by injecting the ligand into a cell containing dialysis buffer and were subtracted from the binding heats. Data were fit to a single binding model using the Origin software package provided with the instrument (MicroCal). Previous experiments indicated $\Delta H \sim -10^4 \text{ cal/mol}$ and $K_A \sim 10^6 \text{ M}^{-1}$. To ensure ideal experimental K window, $c = K_A M_{tot} n$ is required to be within the range 5~50 (M_{tot} - total concentration of macromolecule in sample cell). First experiment started with 30 μ M KLHL3 protein, 250 μ M WNK3 peptide. Then the peptide concentration was raised to 500 μ M to achieve better data quality.

2.6. Surface Plasmon Resonance (SPR)

Surface Plasmon Resonance (SPR) is a biophysical technique to measure biomolecular interactions in a real-time manner. Under total internal reflection, plane-polarized light hits a thin metal (usually gold) film and excites the surface plasmon polaritons (plasmon oscillation). Incident light at a certain angle excites the surface plasmon in a resonance manner, which results in maximum energy absorption and leaves a gap in reflective light intensity (Figure 2.3). This angle, termed 'SPR angle', is dependent on local index of refraction of the metal surface. The Biacore system detects the change of the SPR angle and converts the signal to response unit (RU) which is proportional to the mass of macromolecules attached to the metal surface (e.g. 1RU = 1pg protein). [9]

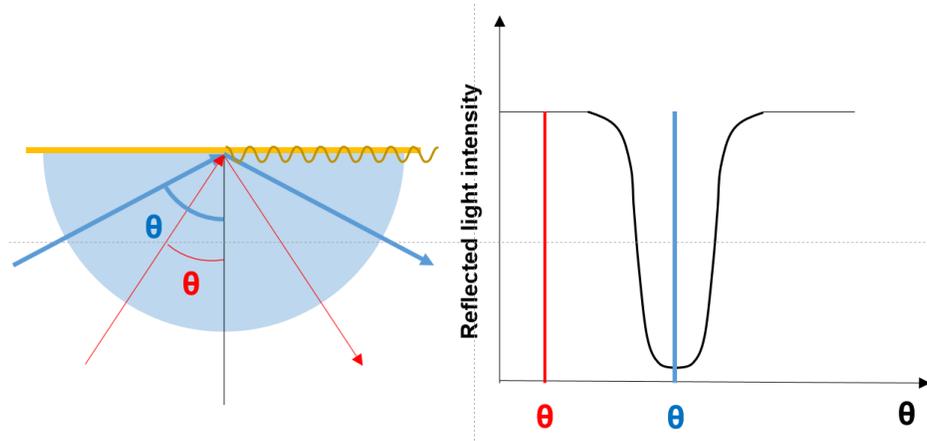


Figure 2.3 Resonance of surface plasmons and photons results in energy absorption and a gap in reflective light intensity. The blue arrows refer to incident light at SPR angle (blue θ) and its reflective light. The red arrows refer to incident light that doesn't resonate with surface plasmons. The yellow wave demonstrates the surface plasmon polariton.

Assays were performed at 25°C using a BIACORE S200 (GE Healthcare) surface plasmon resonance (SPR) instrument. The Kelch domains of KLHL20 and KLHL3 were immobilized on sensor chip CM5 (GE Healthcare) using amine coupling. Reference flow cells had no immobilized protein. Binding was monitored using a flow rate of 30 $\mu\text{L}/\text{min}$. The peptide analytes were prepared in HBS-P buffer (GE Healthcare). Data reported were after reference flow cell signal subtraction. Data were analysed by one-site steady-state affinity analysis using

the Biacore S200 Evaluation software and the fitting equation $R_{eq} = \frac{CR_{max}}{K_D+C} + RI$. (RI, bulk refractive index contribution; R_{max} , maximum response; K_D , dissociation constant; C, analyte concentration). Peptides were obtained from Severn Biotech Ltd.

2.7. Alphascreen

Alphascreen is a bead-based chemical method to study biomacromolecule interactions in plate-format. ALPHA stands for Amplified Luminescent Proximity Homogeneous Assay. Donor beads and acceptor beads are coated with different types of chemical layers that capture biomolecules in solution. Binding of molecules captured on beads leads to proximity of the donor and acceptor beads, which results in energy transfer from one to another and ultimately produces a luminescent/fluorescent signal. The assays were performed with hexahistidine tagged KLHL20 Kelch domain (303-605a.a.) and biotinylated DAPK1 peptide (Biotin-LLAMNGLPDLVAKYNTSNGA). A library of 10K Compounds (MIDAS library) were dispensed in duplicate into 384-well reaction plates using Echo to aim for 33 μ M in final 12 μ L reactions. 60 nM of protein and 60 nM of peptide were added into every well in the reaction plates, and incubated with compounds for 30 min at room temperature. Alphascreen Biotinylated-HIS donor and acceptor beads were mixed and prepared at 1:200 dilution in dark. 4 μ L of beads mixture was added to each well in the reaction plates using combi liquid handler (Thermo fisher). Reaction was incubated O/N at room temperature in dark. Signals were read using Pherastar-fsx plate reader using 'Alphascreen' settings (excitation at 680 nm, reading emission at 520-620 nm). An unlabelled DAPK1 (LGLPDLVAKYN) peptide was added to the first column of each plate at 33 μ M as an inhibition control.

Assay buffer: 20 mM HEPES pH7.5, 150 mM NaCl, 0.1% BSA, 0.05% Tween

2.8. Western blotting

Western blotting is an analytical technique for detection of specific proteins in a sample of tissue homogenate or extract.

Samples were denatured in sample buffer (Invitrogen NP0008) with boiling for 5 minutes, followed by gel electrophoresis (SDS-PAGE in all cases mentioned in this thesis). To enable antibody detection, proteins were transferred from gel onto a nitrocellulose (NC) or polyvinylidene difluoride (PVDF) membrane. Negative charged proteins within the gel were pulled towards the positive charged anode by electric current and onto the NC or PVDF membrane. Other available protein binding sites on the membrane were blocked by 3~5% BSA or Skim milk in PBST after transfer. The membrane was then probed with specific antibodies conjugated with Horseradish Peroxidase (HRP). HRP catalyses the oxidation of luminol to 3-aminophthalate via several intermediates, which results in emission of low-intensity light at 428 nm. After the unbound antibodies were washed away, the membrane was incubated with to Enhanced chemiluminescence substrates (Thermo fisher, Pierce™, 32209) briefly, and imaged by LAS4000 camera. The bands visualised on the Western Blot were analysed and quantified by densitometry using Image Studio Lite Ver 5.2.

Transfer buffer (10x) per 4L - 121.1 g Tris base, 576 g glycine. Bring up the volume to 4 L with ddH₂O

Transfer buffer (1x) for 1 L - 800 mL ddH₂O, 100 mL 10x Transfer buffer, 100 mL methanol

Antibodies – Flag antibody – Sigma-Aldrich, F1804; HA antibody – Biolegend, 901501; Myc antibody – Cell signalling technology, 2040S; GAPDH antibody – Thermo Fisher, MA5-15738; LaminA antibody – Cell signalling technology, 4C11; RCOR1 – Atlas, HPA054241; RCOR3 – Bethyl laboratory, A301-273A; KDM1 – Abcam, ab17721.

2.9. Co-immunoprecipitation

Immunoprecipitation is a biochemical technique of precipitating a protein out of extract sample using corresponding antibody that is coupled to a solid basis (agarose beads for

example). Protein complex immunoprecipitation (Co-IP) is commonly used for protein-protein interaction characterisation. By pulling out a component of the complex with a specific antibody, the entire complex of proteins are co-precipitated and can be analysed with Western Blot or Mass Spectrometry.

Samples were always kept on ice or in cold room during Flag co-immunoprecipitation. Prior to immunoprecipitation, Flag M2 affinity gel (Sigma Aldrich, A2220) were washed three times with lysis buffer. Per 1 mg total lysate protein, 1 μ L of Flag beads were prepared. To remove non-specific binding, plain agarose beads were added to the lysate and incubated for 30 min with rotation at 4°C. The proteins sticky to agarose beads were pelleted down by centrifugation at 2000 RPM for 2 min at 4°C. The supernatant of each sample in the same experiment was transferred to a clean tube containing equal amount of Flag beads. The samples were incubated at 4°C with rotation for 3 hours or overnight to allow epitope-antibody binding. After washing steps, bound proteins were eluted with either LDS sample buffer or with 0.4 mg/mL 3xFlag peptides. The latter increases the elution specificity, therefore was employed in mass spectrometry sample preparation.

2.10. Mass Spectrometry

2.10.1. Intact Mass (LC/MS)

Protein masses were determined using an Agilent LC/MS TOF system combining the liquid chromatography with the mass analyser. Proteins were desalted prior to mass spectrometry by rapid elution off a C3 column with a gradient of 5-95% isopropanol in water with 0.1% formic acid. In the liquid chromatography-mass analyser interface, the proteins were ionized by an electrospray and accelerated by an electric field of known strength. The m/z ratio was determined by the time-of-flight in the mass analyser. Spectra were analysed and deconvoluted using the MassHunter software (Agilent).

2.10.2. Phosphorylation site mapping

Full length human WNK3 (isoform 2) was cloned into a pCMV5-FLAG vector providing an N-terminal Flag affinity tag (MRC-PPU plasmid DU4949). Cells were split into 150 mm dishes and transfected with 14 µg WNK3-Flag construct using polyethylenimine (Polysciences). Some 36 hours post-transfection, culture medium was removed and the cells were treated with either isotonic buffer or hypotonic buffer to allow for phosphorylation under different stimuli. Following 30 mins incubation, cells were washed in PBS buffer and harvested for analysis. WNK3 was immunoprecipitated by anti-Flag M2 affinity resin and eluted with 3xFlag peptide. Protein samples were digested in elastase after DTT reduction, iodoacetamide alkylation and methanol-chloroform precipitation. The samples were then desalted using a SOLA HRP SPE Cartridge (Thermo Fisher, 60109-001) following the manufacturer's instruction. Samples were then submitted to Dr Roman Fisher in Target Discovery Institute, Oxford for LC-MS/MS analysis in search of post translational modification. LC-MS/MS data were analysed with PEAKS 7.0 against Uniprot database using an Elastase digest pattern and peptide level false discovery rate was adjusted to 1%

Isotonic high potassium buffer contained 20 mM HEPES pH 7.4, 95 mM NaCl, 50 mM KCl, 1 mM CaCl₂, 1 mM MgCl₂, 1 mM Na₂HPO₄, 1 mM Na₂SO₄. Hypotonic high potassium buffer contained 20 mM HEPES pH 7.4, 80 mM KCl, 1 mM CaCl₂, 1 mM MgCl₂, 1 mM Na₂HPO₄, 1 mM Na₂SO₄.

2.10.3. Interactome proteomics

The idea of interactome proteomics in identifying E3 substrates has been introduced in 1.4. Full length human KBTBD4 (isoform 1) wild type (WT) and two MB mutants were cloned into pcDNA3-3xFlag vector providing an N-terminal 3xFlag affinity tag Cells were split into 150 mm dishes and transfected with 14 µg KBTBD4 constructs using polyethylenimine (Polysciences).

An empty vector control was also set up for comparison. 36 hours post-transfection, MLN4924 was added to culture medium at 100 nM. Following 4 hours incubation, cells were washed in PBS buffer and harvested for analysis. KBTBD4 WT and mutants were immunoprecipitated by anti-Flag M2 affinity resin and eluted with 3xFlag peptide. Protein samples were digested with trypsin after DTT reduction, iodoacetamide alkylation and methanol-chloroform precipitation. The samples were then desalted using a SOLA HRP SPE Cartridge (Thermo Fisher, 60109-001) following the manufacturer's instruction. Samples were then submitted to Dr Iolanda Vendrell in Target Discovery Institute, Oxford for LC-MS/MS analysis. Data was searched with Mascot against Swissprot database at 1% false discovery rate (FDR) cut-off. Non-specific interactors were subtracted either by comparing with the empty vector hit list or with the Contaminant Repository for Affinity Purification (CRAPome) list [10]

2.11. X-ray crystallography

2.11.1. Crystallisation

Protein crystals grow from an oversaturated solution. Being polyions, protein solubility is dependent on other charged components in the solution. Salt, such as ammonium sulphate, and polymers such as polyethylene glycol (PEG) which sequesters water, are the most common precipitants present in crystallisation screen conditions. In addition, variable parameters, including protein concentration, pH and crystal growing temperature, are also tested in crystallisation trials [11, 12].

Proteins were concentrated to 9~12 mg/mL using a 10 kDa centrifugal concentrator (Millipore). For co-crystallisation, peptides were added to the protein solutions aiming for a final concentration at 2 mM. Crystallisation was performed using sitting drop vapour diffusion. Four sparse matrix coarse screens were employed, namely the Hampton Research Crystal Screen

(HCS), the Hampton Research Index (HIN) screen, the Joint Centre for Structural Genomics (JCSG) screen and the Ligand Friendly Screen (LFS). Among them, HCS and JCSG are both biased and derived from known crystallization conditions for over 1000 macromolecules, offering maximum coverage of different crystallisation parameters. On the other hand, HIN and LFS samples a series of reagent zones for pH and salt testing, which efficiently identifies the pH and the reagent classes that favour crystal growth or at least reduces protein solubility [13].

Sometimes, fine screening were performed to improve the crystal quality. Fine screens were designed based on the conditions of the initial crystals/crystalline precipitate obtained from the coarse screens. Initial crystals obtained from coarse screens were used to make seed stocks to help initiate nucleation [14]. To make seed stocks, crystals were resuspended by 1 μ L of the reservoir solution and transferred to a 1.5mL Eppendorf tube containing 50 μ L reservoir solution and a seed bead. The tube was vortexed for 30 sec and the seed stock can be diluted to 10, 50, 200 fold when in use. 20 nL of the seed stock was added to each subwell in the fine screening. The working dilution of the seedstock for microseeding was tested before large scale screening. Using too concentrated seed stock may result in crystal pathology such as twinning [15].

Prior to vitrification in liquid nitrogen, crystals were cryoprotected by direct addition of reservoir solution supplemented with 25 % ethylene glycol. Diffraction data were collected on beamline I03, I04 and I24 at the Diamond Light Source, Didcot, U.K.

2.11.2. Structure determination

The structure of a protein is determined by measuring the intensities and directions of the spots in X-ray diffraction images, and reconstituting the 3 dimensional molecular model using Fourier transformation based on the 2D images along with a good estimate of phase information.

Diffraction data were indexed and integrated using MOSFLM, scaled and merged using AIMLESS [16, 17]. The occurrence of multimerization state in an asymmetric unit was determined by resolution dependent Matthews's probabilities (<http://www.ruppweb.org/mattprob/default.html>). Molecular replacement was performed with Phaser MR in Phenix using solved Kelch structures as models [18]. COOT was used for manual model building and refinement [19, 20], whereas PHENIX.REFINE was used for automated refinement [20]. Real space and geometry parameters were restrained at the first few rounds of refinement. Finding non-crystallographic symmetry (NCS) and Translation/Libration/Screw (TLS) parameters were included at later stages of refinement.

The refined models were validated by MolProbity [21] and JCSG quality control check (<https://smb.slac.stanford.edu/jcsg/QC/>). Structure figures were made with PyMOL [22]. For protein-peptide co-structures, the interaction interfaces were analysed using the ePISA server [23](http://www.ebi.ac.uk/pdbe/prot_int/pistart.html).

2.12. Ubiquitination assay

Ubiquitination assay is introduced to determine the level of cellular ubiquitination in a protein of interest by preserving and enriching the ubiquitinated species by immunoprecipitation under denatured lysis condition.

To establish the assay, ubiquitin, E3s and substrates were overexpressed in HEK293T cells. Ubiquitin, E3 and substrate were tagged with Myc, Flag and HA respectively. 40 hours post transfection, cells were treated with proteasome inhibitor MG132 at 5 μ M for 4 hours. Cells were lysed in lysis buffer containing 2% SDS with heating at 90°C for 10min. Cell lysate was sonicated and diluted in buffer containing 1% Triton to make it compatible for HA immunoprecipitation. Anti-HA–Agarose antibody (Sigma Aldrich, A2095) were washed with buffer containing 1 M NaCl and bound proteins were eluted with LDS sample buffer.

Cell lysis buffer – 2% SDS, 150 mM NaCl, 10 mM Tris-HCl pH8.0. Dilution Buffer – 10 mM Tris-HCl pH8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton. Washing buffer – 10 mM Tris-HCl pH8.0, 1 M NaCl, 1 mM EDTA, 1% NP40.

2.13. Homology model

Homology model of DAPK1 Death domain were built in Molsoft ICM-Pro software using MyD88 (PDB 3MOP chain A, 25% sequence identity) as the structural template. The initial model was refined by energy minimization and side chain optimization in ICM-Pro (Molsoft) [24].

2.14. Nuclear extraction

A rapid, efficient, and practical (REAP) method was employed to separate nucleus from cells [25]. A mild detergent NP40 was used to disrupt plasma membrane without disturbing the nuclear membrane.

Cells were harvested until 90% confluent (10^6 cells/dish) after rinsing in ice-cold PBS. Plasma membrane was disrupted by resuspending cell pellet in PBS containing 0.1% NP40 and protease inhibitors. Intact nuclei were pelleted down by a brief centrifugation at 10000 RPM for 5 sec. Nuclei were then washed with PBS containing 0.1% NP40 with inhibitors and lysed with Laemmli sample buffer. Total cell lysate and nuclear fraction samples were sonicated before being analysed on Western Blot together with cytosolic fraction. LaminA/C and GAPDH were blotted as nuclear and cytosolic controls respectively.

2.15. Cycloheximide chase assay

Cycloheximide is a chemical compound that blocks eukaryotic protein translation by interfering with protein translocation step in synthesis. In this thesis, cycloheximide chase assay is used for half-life of substrate variants in presence of corresponding E3s.

HEK293T cells were cultured as described above until 60% confluency. Full length BTB-Kelch and substrate constructs were transfected with polyethylenimine. 24 hours after transfection, 100 µg/mL cycloheximide was added to inhibit protein synthesis. Cells were harvested at different time points – e.g. 0, 0.5 h, 1 h, 2 h, 4 h and 6 h. Results were analyzed using Western blotting with corresponding antibodies. GAPDH level in each sample was also detected for control (anti-GAPDH antibody; Thermo Fisher, MA5-15738). Western blot band intensities were quantified using Image Studio Lite Ver 5.2 and normalized for the GAPDH control.

2.16. References

1. Linding, R., et al., *GlobPlot: Exploring protein sequences for globularity and disorder*. Nucleic acids research, 2003. **31**(13): p. 3701-3708.
2. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices* Edited by G. Von Heijne. Journal of Molecular Biology, 1999. **292**(2): p. 195-202.
3. Savitsky, P., et al., *High-throughput production of human proteins for crystallization: The SGC experience*. Journal of Structural Biology, 2010. **172**(1): p. 3-13.
4. Korbie, D.J. and J.S. Mattick, *Touchdown PCR for increased specificity and sensitivity in PCR amplification*. Nature Protocols, 2008. **3**: p. 1452.
5. Li, S. and M.F. Wilkinson, *Site-Directed Mutagenesis: A Two-Step Method Using PCR and Dpn I*. BioTechniques, 1997. **23**(4): p. 588-590.
6. Dagert, M. and S.D. Ehrlich, *Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells*. Gene, 1979. **6**(1): p. 23-28.
7. Marbach, A. and K. Bettenbrock, *lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA*. Journal of Biotechnology, 2012. **157**(1): p. 82-88.
8. Pierce, M.M., C.S. Raman, and B.T. Nall, *Isothermal Titration Calorimetry of Protein-Protein Interactions*. Methods, 1999. **19**(2): p. 213-221.
9. Quinn, J.G., et al., *Development and Application of Surface Plasmon Resonance-Based Biosensors for the Detection of Cell-Ligand Interactions*. Analytical Biochemistry, 2000. **281**(2): p. 135-143.
10. Mellacheruvu, D., et al., *The CRAPome: a contaminant repository for affinity purification-mass spectrometry data*. Nature methods, 2013. **10**(8): p. 730-736.
11. Rupp, B., *Biomolecular crystallography: principles, practice, and application to structural biology*. 2009: Garland Science.

12. Chayen, N.E. and E. Saridakis, *Protein crystallization: from purified protein to diffraction-quality crystal*. Nature methods, 2008. **5**(2): p. 147.
13. Newman, J., et al., *Towards rationalization of crystallization screening for small-to medium-sized academic laboratories: the PACT/JCSG+ strategy*. Acta Crystallographica Section D: Biological Crystallography, 2005. **61**(10): p. 1426-1431.
14. D'Arcy, A., et al., *Microseed matrix screening for optimization in protein crystallization: what have we learned?* Acta crystallographica. Section F, Structural biology communications, 2014. **70**(Pt 9): p. 1117-1126.
15. Oswald, C., et al., *Microseeding - a powerful tool for crystallizing proteins complexed with hydrolyzable substrates*. International journal of molecular sciences, 2008. **9**(7): p. 1131-1141.
16. Leslie A.G.W., P.H.R., *Processing diffraction data with mosflm*, in *Evolving Methods for Macromolecular Crystallography.*, S.J.L. Read R.J., Editor. 2007, Springer: Dordrecht.
17. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta crystallographica. Section D, Biological crystallography, 2013. **69**(Pt 7): p. 1204-1214.
18. McCoy, A.J., et al., *Phaser crystallographic software*. J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
19. Emsley, P., et al., *Features and development of Coot*. Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 4): p. 486-501.
20. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 2): p. 213-221.
21. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography*. Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 1): p. 12-21.
22. LLC., S., *The PyMOL Molecular Graphics System*.
23. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state*. J Mol Biol, 2007. **372**(3): p. 774-97.
24. Abagyan, R., et al., *Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search*. Proteins: Structure, Function, and Bioinformatics, 1997. **29**(S1): p. 29-37.
25. Nabbi, A. and K. Riabowol, *Rapid isolation of nuclei from cells in vitro*. Cold Spring Harbor Protocols, 2015. **2015**(8): p. pdb. prot083733.

Chapter 3

Structural basis for recruitment of DAPK1 to the KLHL20 E3 ligase

This chapter is based on publication:

Chen Z, Picaud S, Filippakopoulos P, D'Angiolella V, Bullock AN. Structural Basis for Recruitment of DAPK1 to the KLHL20 E3 Ligase. *Structure*. 2019;27(9):1395–1404.e4. doi:10.1016/j.str.2019.06.005

3.1. Introduction

KLHL20 is a CUL3-dependent ubiquitin ligase linked to autophagy, cancer and Alzheimer's disease [1-4]. High expression of KLHL20 correlates specifically with high grade tumours [5-7]. Transcription of the *KLHL20* gene is upregulated by the hypoxia-inducible factor HIF-1 α leading to its overexpression in hypoxic tumour cells [2]. In this context, KLHL20 restricts both apoptosis and autophagy-induced cell death by promoting the ubiquitination and degradation of substrates including DAPK1, PML and ULK1 [1, 2, 8]. Furthermore, KLHL20 enhances the hypoxia signalling by stabilizing HIF-2 α (also known as EPAS) [9, 10]. Thus, KLHL20 can promote tumour growth and metastasis.

DAPK1 is the first reported substrate of KLHL20 [1]. Yeast two-hybrid studies previously mapped the interaction to the death domain of DAPK1 and the Kelch domain of KLHL20. It was further shown that the death domain was required for DAPK1 ubiquitination and degradation by KLHL20 [1]. DAPK1 suppresses cancer metastasis and facilitates cell apoptosis when induced by a variety of stimuli, including interferon- γ , TNF- α and TGF- β [11]. Importantly, interferon stimulation causes the sequestration of KLHL20 in so-called 'PML nuclear bodies', which occur as punctate membraneless substructures of the nucleus enriched with PML protein. This inhibitory mechanism allows DAPK1 to evade degradation and to accumulate to

mediate interferon-induced cell death [1]. Therefore, inhibiting KLHL20 to increase DAPK1 level is considered an alternative or complementary cancer therapy strategy to IFN treatment.

In this chapter, through a peptide scanning approach I identified a 'LPDLV'-containing recruitment site within this DAPK1 region that bound to KLHL20 with low micromolar affinity. I also determined the crystal structure of their complex at 1.1 Å resolution revealing a distinct peptide binding mode compared to the previously determined structural complexes of KEAP1 and KLHL3 [12-14].

3.2. Results

3.2.1. Mapping of the DAPK1 binding motif for KLHL20 recruitment

The recombinant death domain of DAPK1 (Figure 3.1A) has been shown to display intrinsic disorder and a high propensity for aggregation making it unsuitable for structural studies [15]. Given the lack of structural order, I set out to map the DAPK1 binding epitope using the SPOT peptide technology. I synthesized a peptide array to span the length of the DAPK1 death domain using 15-mer peptides and a three amino acid frameshift at each position. Probing of the array with recombinant 6xHis-KLHL20 Kelch domain and anti-His-antibody for detection revealed protein capture at two sites encompassing DAPK1 residues 1327-1350 and 1378-1395, respectively (Figure 3.1B). A control experiment indicated that the binding epitope was likely to reside within the N-terminal region since peptides from the second site also bound to the anti-His antibody alone marking them as likely false positives (Figure 3.1B).

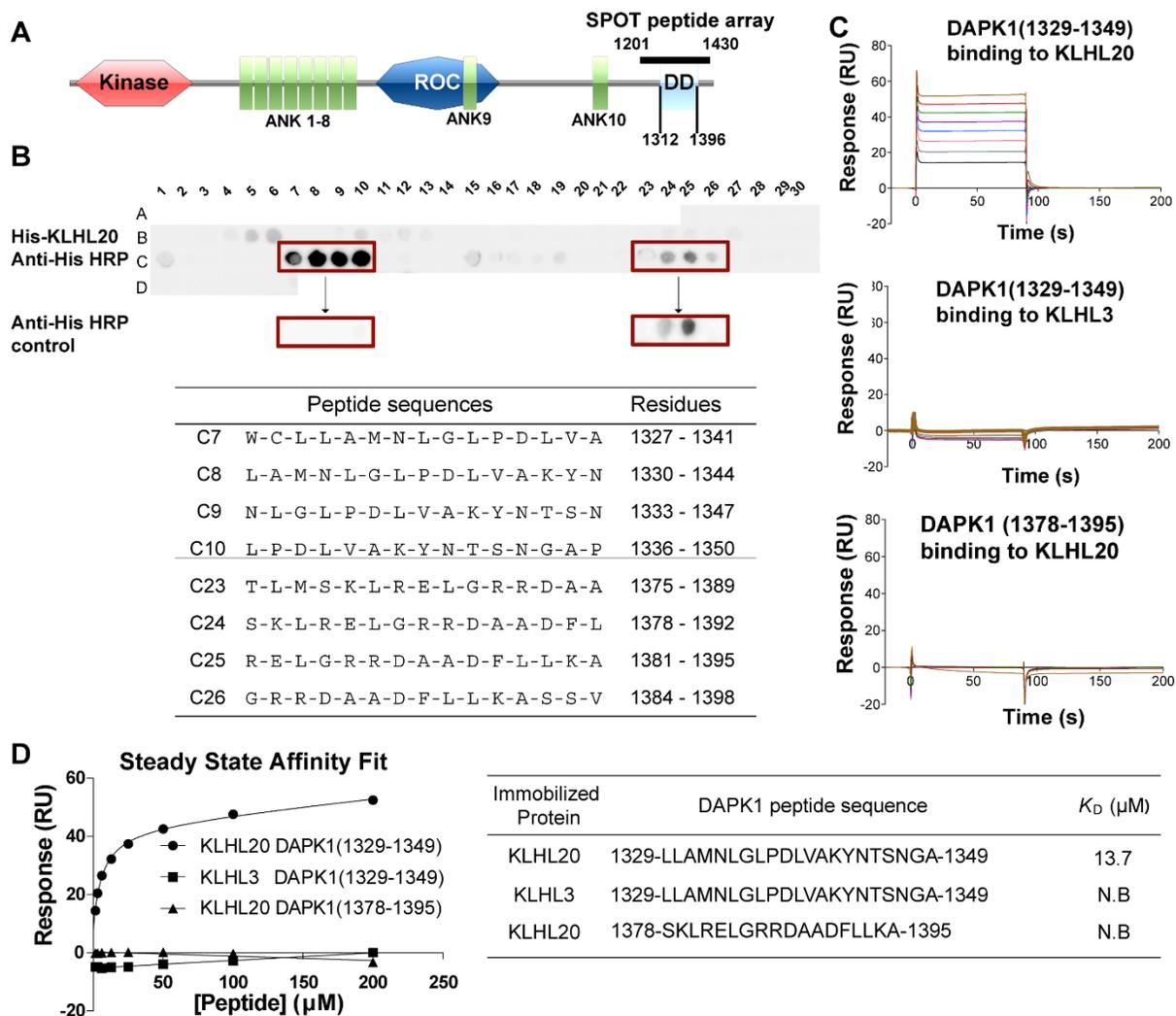
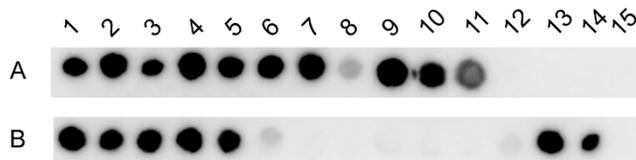


Figure 3.1 Mapping of the DAPK1 binding motif for KLHL20 recruitment. (A) Domain organization of human DAPK1 (ank, ankyrin repeat; DD, death domain comprising residues 1312 to 1396). Solid bar denotes the extended region explored for KLHL20 interaction (DAPK1 residues 1201-1430). (B) SPOT peptide array. Each spot was printed as a 15-mer DAPK1 peptide with a 3 residue frameshift at each consecutive position. Arrays were incubated with purified 6xHis-KLHL20 Kelch domain, washed and then KLHL20 binding detected using anti-His HRP-conjugated antibody. Binding was observed at two sites spanning DAPK1 residues 1327-1350 and 1378-1395, respectively. As a control, duplicate spots were probed with antibody alone and revealed non-specific antibody binding to DAPK1 residues 1378-1395. (C) For SPR experiments, KLHL20 and KLHL3 Kelch domains were immobilized by amine coupling on different flow cells of a CM5 sensor chip. Indicated DAPK1 peptides were injected subsequently at concentrations of 1.6 μM , 3.1 μM , 6.2 μM , 12.5 μM , 25 μM , 50 μM , 100 μM and 200 μM . Binding was monitored at a flow rate of 30 $\mu\text{L}/\text{min}$. (D) SPR binding data shown are representative of three independent experiments. Data were fitted using a steady state affinity equation. DAPK1 residues 1329-1349 bound to KLHL20 Kelch domain with $K_D = 13.7$ μM (N.B., no binding detected). The three independent experiments determined the $K_D = 14.8 \pm 2.8$ μM .

To validate these putative interaction sites, I designed peptides for the two DAPK1 regions and performed surface plasmon resonance (SPR) experiments to measure their respective binding affinities for KLHL20 (Figure 3.1C). A DAPK1 peptide spanning the N-terminal site residues 1329-1349 bound robustly to the Kelch domain of KLHL20 with $K_D = 13.7 \mu\text{M}$ (Figure 3.1D). The same peptide showed no apparent binding to the Kelch domain of KLHL3 demonstrating that the interaction was specific to KLHL20 (Figures 3.1D). A DAPK1 peptide spanning the C-terminal site residues 1378-1395 also failed to bind to KLHL20 confirming that this downstream region was a false positive (Figure 3.1D). Together these data identified a single epitope within the death domain of DAPK1 that showed both potency and specificity for interaction with KLHL20.

3.2.2. An 'LPDLV' motif in DAPK1 is critical for KLHL20 interaction

Attempts to crystallize KLHL20 either alone or in complex with the identified 21-mer peptide from DAPK1 produced only microcrystalline material yielding poor diffraction. Therefore, I sought to refine the minimal DAPK1 epitope by using the SPOT technology for peptide truncation experiments, as well as alanine scanning to probe the sequence determinants of binding. The results from these experiments were in excellent agreement and identified DAPK1 residues Leu1336 to Val1340 as critical for KLHL20 interaction (Figure 3.2). N-terminal deletion or mutation of Leu1336 drastically reduced KLHL20 binding, whereas C-terminal deletion or mutation of Val1340 abolished all detectable binding (Figure 3.2). Other deletions and mutations outside of this region appeared well tolerated mapping the critical binding region to a '1336-LPDLV-1340' motif in DAPK1.



Truncation series

	Peptide sequences	Binding
A1	L-A-M-N-L-G-L-P-D-L-V-A-K-Y-N	+++
A2	A-M-N-L-G-L-P-D-L-V-A-K-Y-N	+++
A3	M-N-L-G-L-P-D-L-V-A-K-Y-N	+++
A4	N-L-G-L-P-D-L-V-A-K-Y-N	+++
A5	L-G-L-P-D-L-V-A-K-Y-N	+++
A6	G-L-P-D-L-V-A-K-Y-N	+++
A7	L-P-D-L-V-A-K-Y-N	+++
A8	P-D-L-V-A-K-Y-N	+
A9	L-A-M-N-L-G-L-P-D-L-V-A-K-Y	+++
A10	L-A-M-N-L-G-L-P-D-L-V-A-K	+++
A11	L-A-M-N-L-G-L-P-D-L-V	++
A12	L-A-M-N-L-G-L-P-D-L	-
A13	L-A-M-N-L-G-L-P-D	-
A14	L-A-M-N-L-G-L-P	-
A15	L-A-M-N-L-G-L	-

Alanine scanning

	Peptide sequences	Binding
B1	L-A-M-N-L-G-L-P-D-L-V-A-K-Y-N	+++
B2	A-A-A -N-L-G-L-P-D-L-V-A-K-Y-N	+++
B3	L- A-A-A -L-G-L-P-D-L-V-A-K-Y-N	+++
B4	L- A-A-A -A-G-L-P-D-L-V-A-K-Y-N	+++
B5	L-A-M- A-A-A -L-P-D-L-V-A-K-Y-N	+++
B6	L-A-M-N- A-A-A -P-D-L-V-A-K-Y-N	+
B7	L-A-M-N-L- A-A-A -D-L-V-A-K-Y-N	-
B8	L-A-M-N-L-G- A-A-A -L-V-A-K-Y-N	-
B9	L-A-M-N-L-G-L- A-A-A -V-A-K-Y-N	-
B10	L-A-M-N-L-G-L-P- A-A-A -A-K-Y-N	-
B11	L-A-M-N-L-G-L-P-D- A-A-A -K-Y-N	-
B12	L-A-M-N-L-G-L-P-D-L- A-A-A -Y-N	+
B13	L-A-M-N-L-G-L-P-D-L-V- A-A-A -N	+++
B14	L-A-M-N-L-G-L-P-D-L-V-A- A-A-A	+++

Figure 3.2 An 'LPDLV' motif in DAPK1 is critical for KLHL20 interaction. DAPK1 peptide variants were printed in SPOT peptide arrays. Row A peptides explored N and C-terminal truncations, whereas row B explored triple-alanine scanning mutagenesis. Arrays were incubated with purified 6xHis-KLHL20 Kelch domain, washed and then binding detected with anti-His antibody. KLHL20 binding was abrogated upon deletion or mutation of a central 'LPDLV' sequence motif in DAPK1.

3.2.3. High resolution structure of the KLHL20 Kelch domain bound to DAPK1 peptide

For further co-crystallization trials, I tried an 11-residue DAPK1 peptide (LGLPDLVAKYN) in order to capture interactions of the central 'LPDLV' motif, while allowing for local conformational preferences and potential flanking interactions. The purified KLHL20 protein was concentrated to 12 mg/mL using a 10 kDa molecular-mass cut-off centrifugal concentrator in 50 mM HEPES pH 7.5, 300 mM NaCl and 5 mM TCEP buffer. The 11-residue DAPK1 peptide (LGLPDLVAKYN) was purchased from LifeTein and added in the same buffer to a final concentration of 3 mM. The protein-peptide mixture was incubated on ice for 1 hour prior to setting up sitting-drop vapour-diffusion crystallization plates. Micro-seed stocks were prepared from small KLHL20 crystals grown during previous rounds of crystal optimization. Seed stocks were diluted 500 fold before use. Viable crystals were obtained with strong diffraction after a combination of microseeding from initial hits and fine matrix screening for further optimization (Figure 3.3). The best-diffracting crystals of the KLHL20 complex were obtained at 20°C by mixing 75 nL protein, 20 nL diluted seed stock and 75 nL of a reservoir solution containing 2 M sodium chloride and 0.1 M acetate buffer pH 4.5. Prior to vitrification in liquid nitrogen, crystals were cryoprotected by direct addition of reservoir solution supplemented with 25 % ethylene glycol.

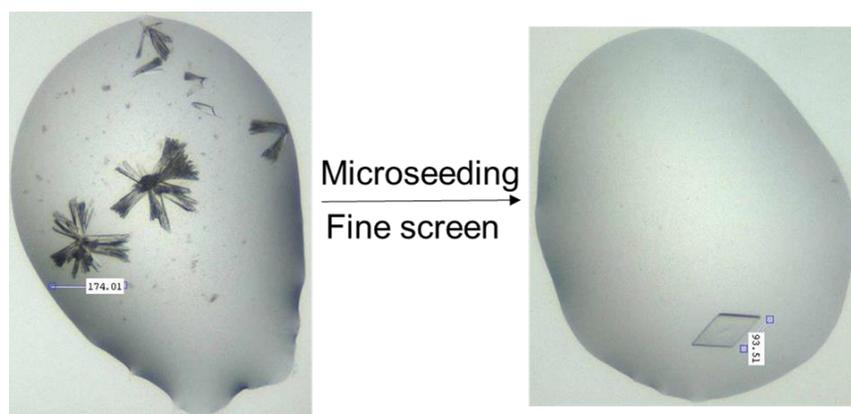


Figure 3.3 Crystal optimization for KLHL20-DAPK1 complex. Initial crystals were obtained at 20°C by mixing 75 nL protein and 75 nL of a reservoir solution containing 25% PEG3350 and 0.1 M citrate buffer pH 3.5. A combination of microseeding and fine screening optimization method was employed and higher quality crystals were obtained. Figure showing a representative crystal after optimization.

Subsequently, we were able to determine a high resolution structure for the complex of KLHL20 and DAPK1 peptide (Table 3.1). Diffraction data for the KLHL20 Kelch domain-DAPK1 complex were collected on beamline I03 at Diamond Light Source, Didcot, U.K. Data were processed in PHENIX version1.9 [31]. Molecular replacement was performed with PHENIX.Phaser-MR using KLHL12 (PDB 2VPJ chain A) as the search model. PHENIX.Autobuild was used to build the initial structural model. COOT [32] was used for manual model building and refinement, whereas PHENIX.REFINE was used for automated refinement. TLS parameters were included at later stages of refinement. Tools in COOT, PHENIX and MolProbity [33] were used to validate the structures.

Table 3.1 Data collection and refinement statistics

Structure of human KLHL20-DAPK1 complex, PDB 6GY5	
Data collection	
Beamline	Diamond Light Source, I04
Wavelength (Å)	0.9763
Resolution range (Å)	40.19 - 1.086 (1.125 - 1.086)
Space group	P 2 ₁ 2 ₁ 2 ₁
Unit cell dimensions	
<i>a, b, c</i> (Å)	40.5, 47.4, 151.9
α, β, γ (°)	90, 90, 90
Total reflections	242242 (19836)
Unique reflections	122957 (11329)
Completeness (%)	99 (92)
Mean <i>I</i> / σ (<i>I</i>)	12.89 (2.75)
CC1/2	0.998 (0.902)
R-merge	0.02995 (0.2073)
Refinement	
Reflections used in refinement	122957 (11329)
Reflections used for R-free	6234 (582)
R-work	0.1536 (0.2205)
R-free	0.1726 (0.2164)
Number of non-hydrogen atoms	2798
RMS deviation (bonds, Å)	0.01
RMS deviation (angles, °)	1.32
Ramachandran favored (%)	98
Ramachandran allowed (%)	2.2
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.39
Average B-factor (Å ²)	15.83

*Values in brackets show the statistics for the highest resolution shells. RMS indicates root-mean-square.

The structure was refined at 1.1 Å resolution and traced the full KLHL20 Kelch domain from residues 317 to 601 (Figure 3.4A). The complete DAPK1 peptide was also clearly defined in the electron density map (Figure 3.4B) allowing its binding interactions to be mapped in atomic detail.

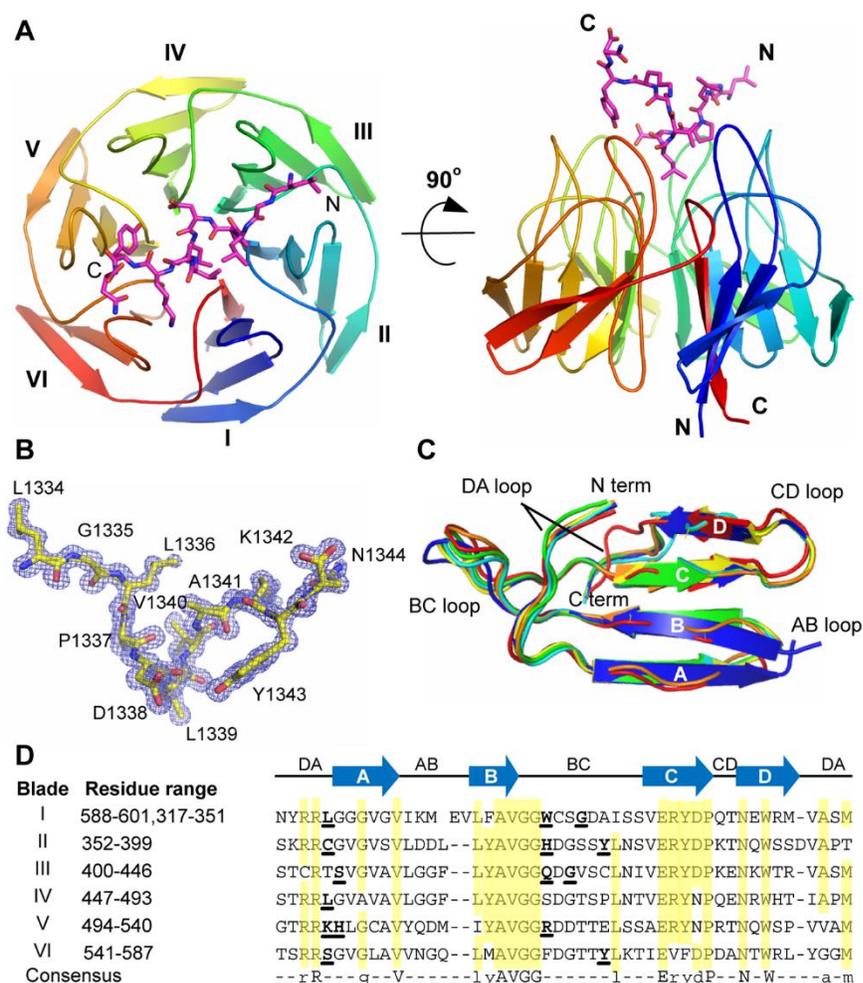


Figure 3.4 High resolution structure of KLHL20 Kelch domain bound to DAPK1 peptide. (A) Overview of the structure of KLHL20 Kelch domain in complex with DAPK1 peptide (purple sticks). Kelch repeats forming blades I to VI are labelled. (B) 2Fo-Fc electron density map (blue mesh) for the DAPK1 peptide contoured at 1.0 σ . (C) Superposition of Kelch domain blades I to VI colored from blue to red. Each blade is composed of four antiparallel β -strands (labelled A-D) and connecting loops. (D) Sequence alignment of the six Kelch repeats in KLHL20. Conserved residues are highlighted in yellow. DAPK1-interacting residues are shown in bold and underlined.

The Kelch domain structure shows a canonical β -propeller fold. The six Kelch repeats form the six blades (I-VI) of the propeller arranged radially around a central axis. Each repeat is folded into a twisted β -sheet consisting of four antiparallel β -strands (A-D, Figure 3.4C). A final C-terminal β -strand is observed to close the β -propeller and inserts into blade I as the innermost β A strand. Blade I is therefore comprised of a C-terminal β A strand and N-terminal β B, β C, and β D strands. Packing between each blade is mediated by a number of conserved hydrophobic positions as well as several buried charged residues that recur within each kelch repeat (Figure 3.4D).

The substrate binding surface on KLHL20 is shaped by the long BC loops, which protrude outwards from the Kelch domain surface, and the largely buried DA loops which link adjacent blades and contribute to the protein core. Notably, the six BC loops in KLHL20 are all of equal length comprising 11 residues (Figure 3.4D), whereas other Kelch domain structures have shown more varied loop lengths across the different blades [16].

3.2.4. Extended interactions of the DAPK1 peptide

The bound DAPK1 peptide shows an extended conformation that packs between Kelch domain blades II and III at its N-terminus and blades V and VI at its C-terminus (Figure 3.5A). At its centre, the peptide adopts a single loose helical turn that is stabilized by intramolecular hydrogen bonds between the carbonyl of Pro1337 and the amides of Val1340 and Ala1341. Here, the peptide inserts deeply into the central cavity of the Kelch domain β -propeller where it is anchored in the complex by Leu1339, the second leucine in the 'LPDLV' motif (Figure 3.5B). Binding at this central region allows the peptide to form additional contacts with blades I and IV. Thus, DAPK1 forms interactions with all six Kelch repeats, including interactions with all six DA loops and all BC loops with the exception of the BC loop in blade IV (Figures 3.4D and 3.5C).

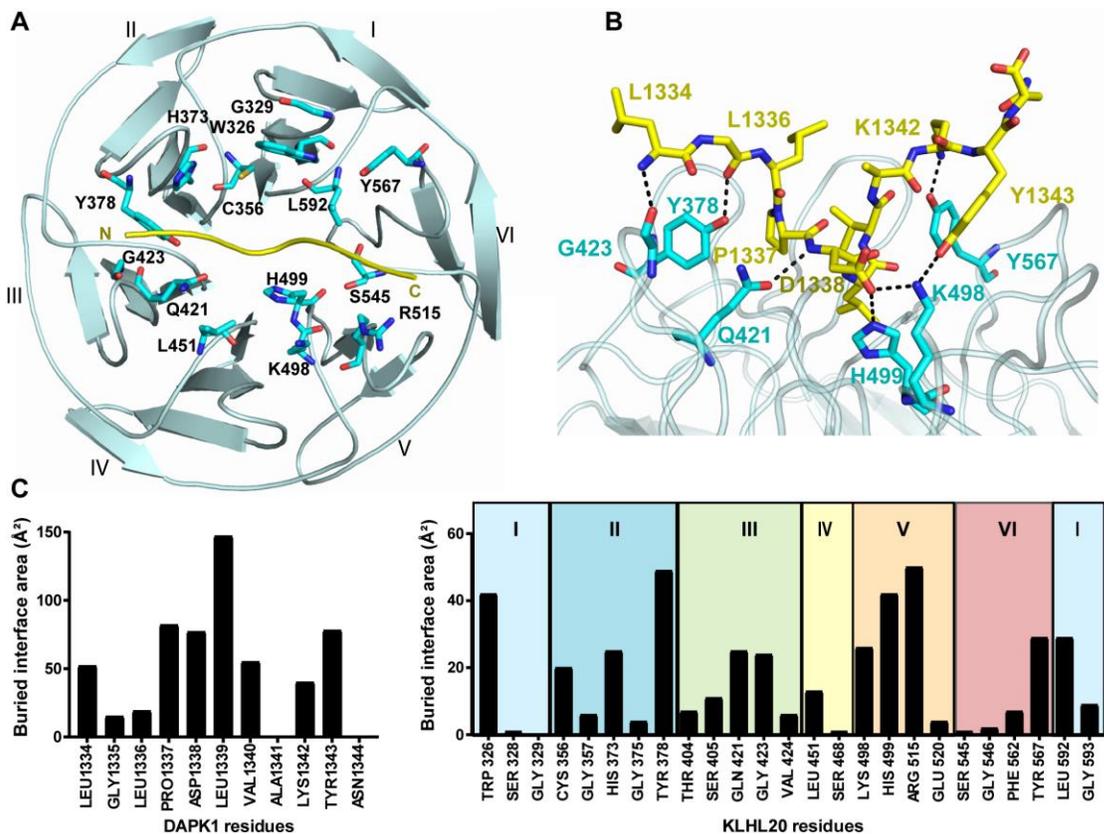


Figure 3.5. Interactions in the KLHL20 Kelch domain-DAPK1 complex. (A) An overview of the DAPK1-binding residues in KLHL20 coloured cyan with DAPK1 peptide shown as yellow ribbon. **(B)** Salt bridge and hydrogen bond interactions in the complex interface are shown by dashed lines. **(C)** Buried interface surface areas for interacting residues in the KLHL20 Kelch domain-DAPK1 complex.

3.2.5. Interactions of the ‘LPDLV’ motif

The ‘LPDLV’ motif of DAPK lies at the core of the protein-peptide interface. Here, the hydrophobic side chains pack against Kelch domain blades I and II and make notable van der Waals contacts with KLHL20 Trp326, His373 and Leu592, respectively (Figure 3.5A). Somewhat surprisingly the first leucine in the ‘LPDLV’ motif, Leu1336, is oriented away from the binding interface and has only minor interaction with KLHL20, mostly through its main chain atoms. In the SPOT peptide arrays, changes at this position reduced KLHL20 binding

significantly, but did not abolish it (Figure 3.2). The importance of this position likely stems from the conformational constraints of the following DAPK1 residue Pro1337. By contrast, the second leucine, Leu1339, is the most buried DAPK1 residue in the complex (Figure 3.5C). This side chain lies sandwiched between KLHL20 His499 and Leu592 (Figures 3.6A-B), but forms interactions across all the Kelch repeats, except for blade IV, by virtue of its central binding position. The final residue in the ‘LPDLV’ motif, DAPK1 Asp1338, is oriented away from the hydrophobic side chains to face Kelch domain blade V, where it forms a salt bridge with KLHL20 Lys498, as well as a hydrogen bond to His499 (Figure 3.6B). Residues across the mapped DAPK1 binding motif are well conserved across vertebrate species (Figure 3.6C).

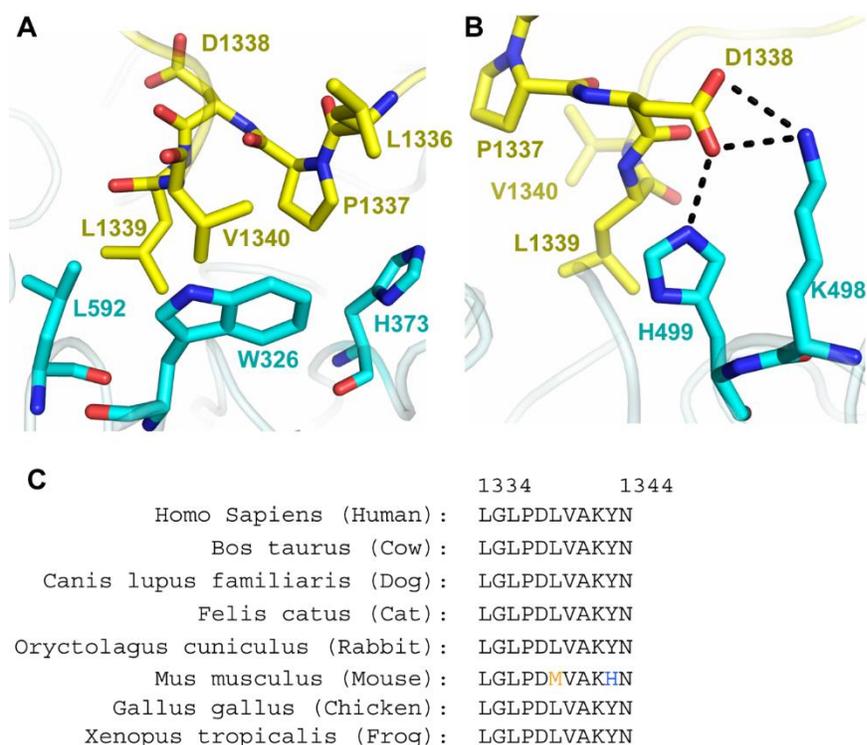


Figure 3.6 Interactions of the DAPK1 ‘LPDLV’ motif. (A) Hydrophobic interactions between the DAPK1 ‘LPDLV’ motif and KLHL20 Kelch domain blades I and II. (B) DAPK1 Asp1338 forms a salt bridge with KLHL20 Lys498, as well as a hydrogen bond to H499. (C) The sequence of the crystallized DAPK1 peptide is conserved across species.

3.2.6. Direct and water-mediated hydrogen bonding in the KLHL20 Kelch domain-DAPK1 complex

In total, the complex between KLHL20 and DAPK1 includes 8 direct hydrogen bond or salt bridge interactions (Figure 3.5B), as well as a number of water-mediated interactions (Figure 3.7). The N-terminal three residues of the DAPK1 peptide are oriented away from the KLHL20 surface. Their binding interactions are mediated by their main chain atoms, which form hydrogen bonds with KLHL20 residues Tyr378, Gln421 and Gly423, respectively (Figure 3.5B). Gln421 forms an additional hydrogen bond with the backbone amide of DAPK1 Asp1338, one of the critical residues within the 'LPDLV' motif. Towards the C-terminus of the peptide, interactions are formed through DAPK1 Lys1342 and Tyr1343, while Ala1341 and Asn1344 are oriented to solvent. Lys1342 folds towards Kelch domain blade VI where it forms a direct hydrogen bond to the BC loop residue Tyr567 (Figure 3.5B). DAPK1 Tyr1343 folds instead against blade V to hydrogen bond with KLHL20 Lys498 (DA loop, Figure 3.5B) and forms additional hydrophobic packing with the BC loop residue Arg515. Water molecules in the complex help to bridge more distant contacts or to satisfy other nitrogen and oxygen atoms that otherwise lack direct hydrogen bonds (Figure 3.7).

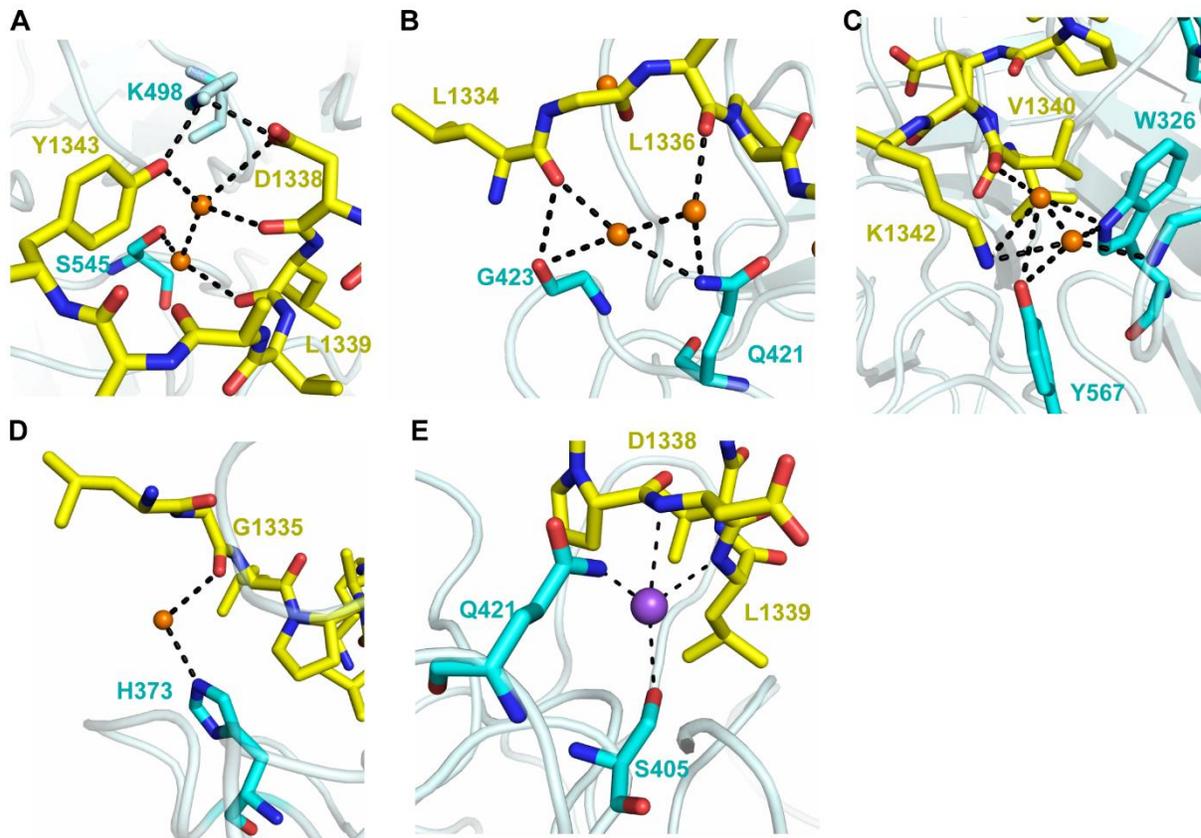


Figure 3.7 Indirect contacts between KLHL20 and DAPK1. (A-D) Waterbridged hydrogen bonds in the protein-peptide interface. Waters are shown as orange spheres. **(E)** A sodium ion (purple sphere) is proximal for electrostatic interactions with the side chains of KLHL20 Ser405 and Gln421 as well as the backbone atoms of DAPK1 D1338 and L1339

3.2.7. KLHL20-induced degradation of DAPK1 is dependent on the ‘LPDLV’ motif

To confirm the identified ‘LPDLV’ motif as a regulatory site for DAPK1 interaction and degradation I performed immunoprecipitation and cycloheximide (CHX) chase experiments in HEK293T cells. To disrupt the interaction site we generated a full length DAPK1 mutant in which the critical 1336-‘LPDLV’ motif was mutated to ‘LPAAV’. Immunoprecipitation (Figure 3.8A) of Flag-KLHL20 Kelch domain and HA-DAPK1 full length variants showed that the wild-type (WT) DAPK1 was robustly bound to KLHL20 whereas the DAPK1 mutant was only recovered at a low background level also observed with anti-Flag agarose beads alone. I then performed a cycloheximide chase assay to compare the degradation of DAPK1 variants in the

presence or absence of full length KLHL20. As shown in Figures 3.8B and C, co-expression of full length KLHL20 and DAPK1 WT caused a striking reduction in the half-life of DAPK1 compared to expressing DAPK1 WT alone. However, the DAPK1 mutant appeared strikingly resistant to KLHL20 co-expression consistent with its disrupted protein interaction. Indeed, the half-life of the DAPK1 mutant exceeded that of DAPK1 WT under any condition above. Taken together, these data indicated that the 'LPDLV' motif was required for both DAPK1 recruitment and degradation by KLHL20.

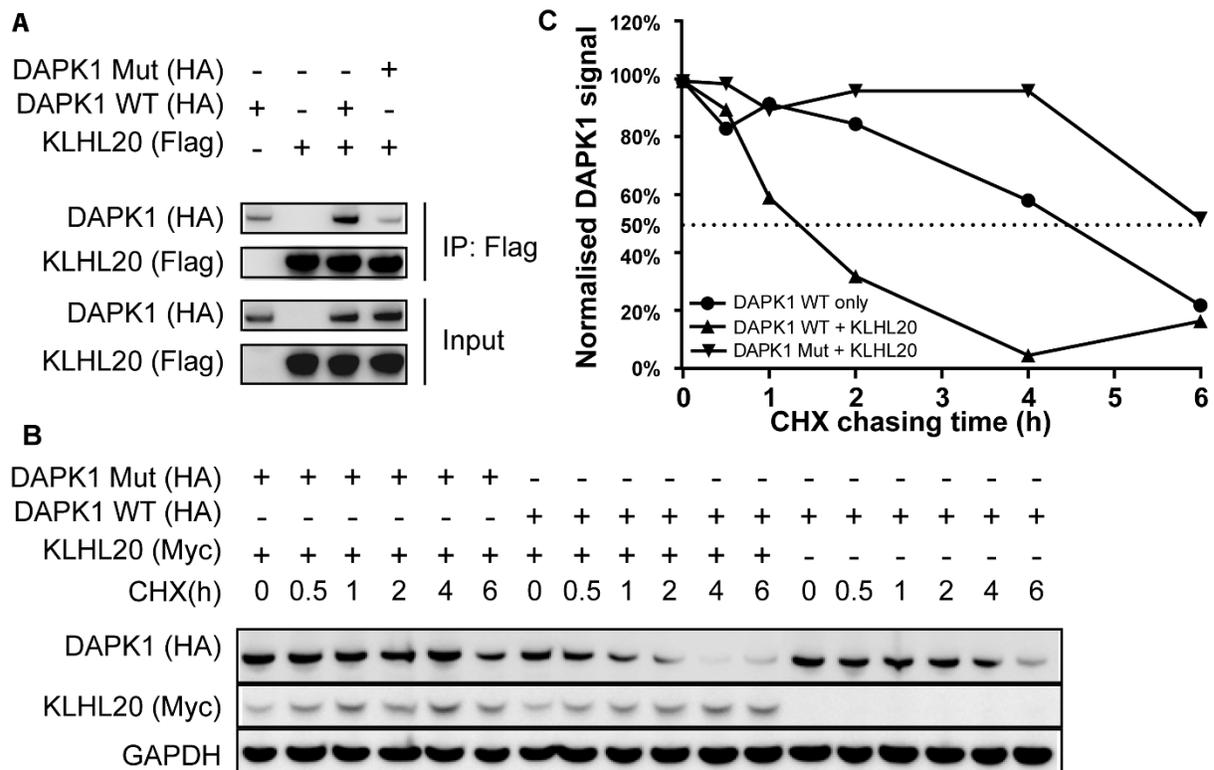


Figure 3.8 Mutations in the DAPK1 'LPDLV' motif impair DAPK1 binding and degradation by KLHL20. (A) Full length DAPK1 variants and KLHL20 Kelch domain were co-transfected into HEK293T cells as indicated. Flag-KLHL20 Kelch domain was immunoprecipitated with anti-Flag antibody. DAPK1 WT was robustly co-purified with KLHL20, whereas DAPK1 mutant was only recovered at the background level of the beads alone. (B) DAPK1 variants were transfected into HEK293T cells with or without full length KLHL20 as indicated. After 24 hours post transfection, cells were incubated with 100 μ g/mL cycloheximide (CHX) and harvested at different time points as indicated. DAPK1 protein levels were detected by Western blot and normalized to GAPDH. (C) Quantitation of (B), showing results from a single experiment.

3.2.8. A putative KLHL20-binding site in EPAS1 contains a 'GPDVL' motif

The KLHL20-DAPK1 structure identifies a central 'LPDLV' motif in the DAPK1 epitope that inserts into the central pocket of the Kelch β -propeller as a loose helical turn. The interface in KLHL20 complements this motif with a hydrophobic core supported by a salt bridge interaction. A similar pattern was also observed in a SPOTs array exploring KLHL20 binding sites within C terminal region of EPAS1. EPAS1 is another reported interaction partner of KLHL20 [18] and contains a 691-'GPDVL' motif in its C-terminal region for which I detected binding to the KLHL20 Kelch domain (Figure 3.9).

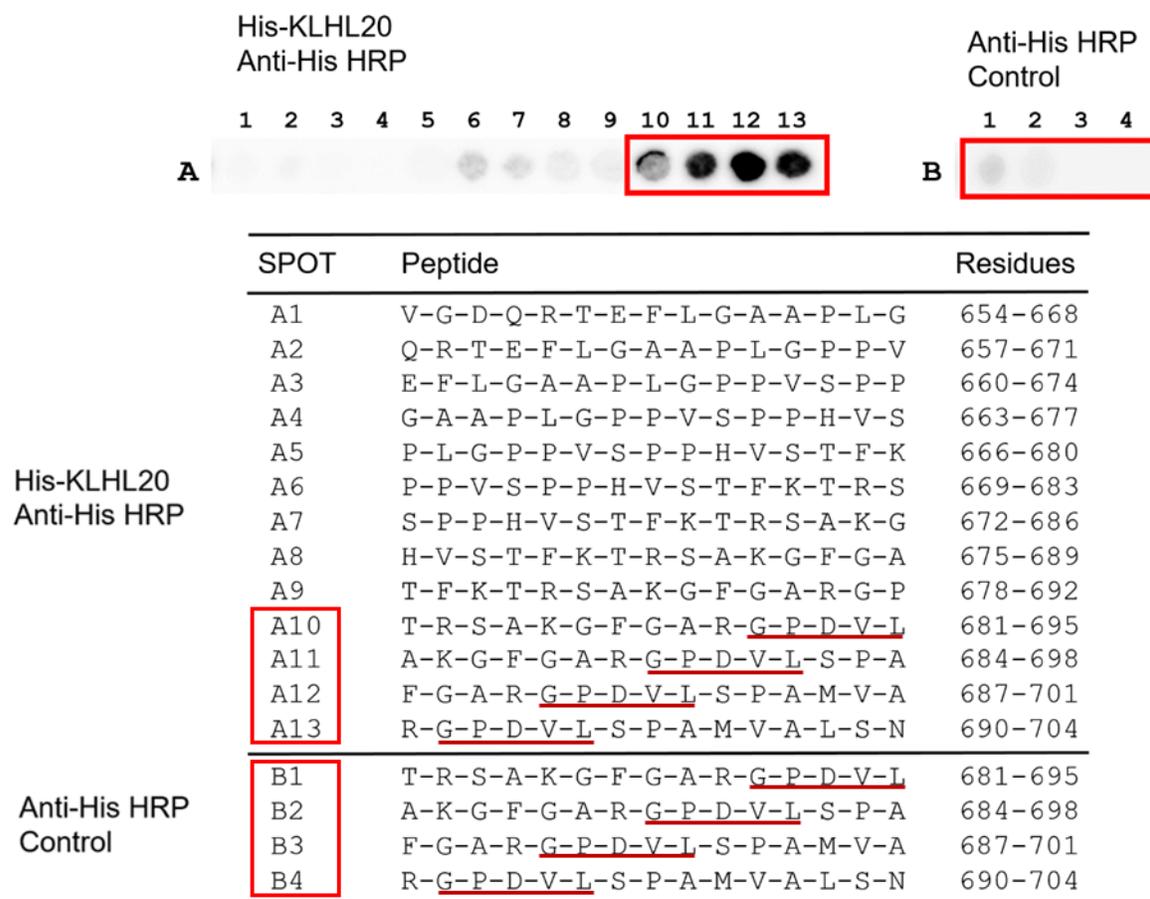


Figure 3.9 SPOTs peptide array for EPAS1. Each spot was printed as a 15-mer EPAS1 peptide with a 3 residue frameshift at each consecutive position. SPOTs array A1-13 was incubated with purified 6xHis-KLHL20 Kelch domain, washed and then KLHL20 binding detected using anti-His HRP-conjugated antibody. Binding was observed at peptides containing a ‘GPDVL’ motif. As a control, duplicate SPOTs array B1-B4 was probed with antibody alone. The GPDVL motif is highlighted with red underlines in the table. Peptides that contained the entire motif have highlighted with red boxes in both the arrays and the table.

3.2.9. Small molecule inhibitor screening for KLHL20

KLHL20 has emerged as an interesting target for drug development with potential application in both oncology and Alzheimer’s disease. Inhibition of KLHL20 would help to stabilize the tumour suppressor proteins DAPK1 and PML [1, 5]. It could also stabilize ULK1 to prolong autophagy allowing greater clearance of potentially toxic misfolded proteins [8]. The interaction surface in KLHL20 provides a notable hydrophobic contribution that should make it favourable

for inhibitor development (Figure 3.10). Moreover, the identified DAPK1 peptide provides a valuable reagent for drug screening assays based on peptide displacement.

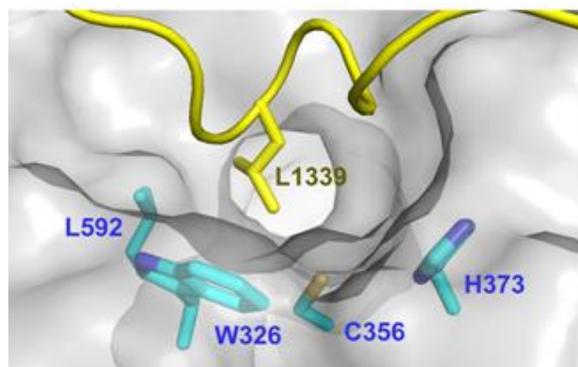
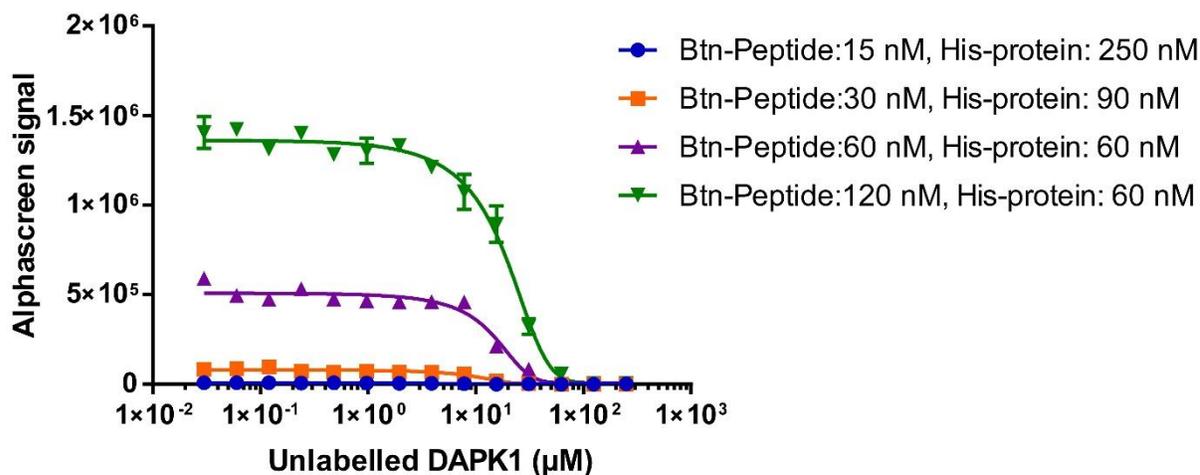


Figure 3.10 Surface representation of KLHL20 substrate binding pocket. The figure highlights the potentially druggable pocket bound by DAPK1 Leu1339. Accessible hydrophobics such as KLHL20 Trp326 and a proximal KLHL20 Cys356 are shown with stick representation and labelled with residue names and numbers.

I established an AlphaScreen (Perkin Elmer) assay using biotinylated DAPK1 peptide (Biotin-LLAMNLGLPDLVAKYNTSNGA, Btn-peptide) and N-terminally hexahistidine tagged KLHL20 protein (His-protein) to screen KLHL20 inhibitors against a 10K-compound Manchester Institute Diversity Set (MIDaS) library. Prior to screening the library, preliminary experiments for dose response with the unlabelled (non-biotinylated) DAPK1 peptide were performed at four different concentration ratios of Btn-peptide to His-protein (15 nM/250 nM, 30 nM/90 nM, 60 nM/60 nM and 120 nM/60 nM)(Figure 3.11A). The interaction of Btn-peptide with His-protein was detected and measured by the Alphascreen signal. By increasing the concentration of the unlabelled DAPK1 peptide in the reaction, the interaction was interrupted, which results in Alphascreen signal decrease in a dose dependent manner. Experiments set up with Btn-peptide to His-protein ratios of 60 nM/60 nM and 120 nM/60 nM both revealed low micromolar IC₅₀ values (13.07 μ M and 13.61 μ M, respectively) which are consistent with the K_D value determined by SPR experiments. (Figures 3.1 and 3.11A). Alphascreen signals measured with 15 nM and 30 nM Btn-peptide were too close to the background level to observe

signal for the inhibition effect. Therefore, I performed the compound screening against the MIDAS library using 60 nM Btn-peptide and 60 nM His-protein. The screen was designed and performed in a 384-well plate format as a single shot experiment with the compound concentration at 33 μ M in the final reaction. (Detailed methods can be found in section 2.7). To analyse the specificity of the inhibition effect, the inhibition percentages of KLHL20 in the single concentration screening are plotted against the inhibition percentages of a counter target screening against the same MIDAS library at the Alzheimer's Research UK Oxford Drug Discovery Institute, University of Oxford, UK (Figure 3.11B). The off-target analysis suggests promising specificity for KLHL20 using this non-covalent inhibitor collection and mitigates the risk of compounds causing interference in the assay which therefore generate false positives.

A Inhibition by unlabelled DAPK1 peptide



B

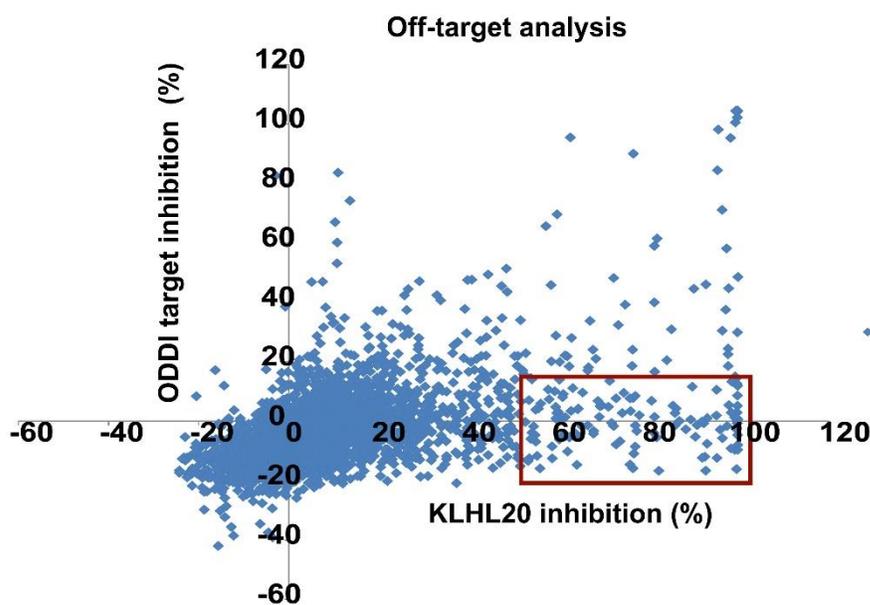


Figure 3.11 Alphascreen assay for inhibitor screening of KLHL20 using peptide-displacement strategy. (A) The interaction of His-protein with Btn-peptide is interrupted by unlabelled DAPK1 peptide in Alphascreen assay. Data points represent mean value of three technical replicates, error bars represent standard deviation. **(B)** Single shot inhibitor screening of the 10K MIDAS library of non-covalent compounds. Hits for KLHL20 are plotted against hits for a counter target screening at the Alzheimer's Research UK Oxford Drug Discovery Institute (ODDI), University of Oxford, UK. The hits that show above 50% inhibition to KLHL20 and below 20% inhibition to the counter target are marked in the red box.

For further validation, I collaborated with Dr Oleg Fedorov and Mr James Bennett (from the Biophysical and biochemical screening group, SGC Oxford) to determine the dose-response relationship of the top hits that were selected from the screening results on the basis of high KLHL20 inhibition (above 50%) and low counter target inhibition (below 20%) (Figure 3.11B). Due to time pressures completing this thesis, the dose-response experiments were performed by Mr James Bennett and showed single digit micromolar IC₅₀ values for the best performing hits using two orthogonal assays Alphascreen and HTRF (Cisbio) (Figure 3.12). Structural analyses of these initial hits are underway to provide more information for drug design. The chemical scaffolds cannot be disclosed at this preliminary stage due to the nature of the collaboration.

Compound ID	AlphaScreen				HTRF			
	IC ₅₀ uM	% I (50 uM)	% I (50 uM) SS Screen	Hill slope	IC ₅₀ uM	% I (50 uM)	% I (50 uM) SS Screen	Hill slope
XS061116e	2.5	99.8	99.9	1.5	2.5	96.7	92.2	1.1
UB007699a	3.7	99.7	99.4	4.6	5.8		-29.0	3.0
UB007700a	8.7	98.0	98.6	1.5	18.7	88.1	77.2	1.7
UB007744a	9.4	98.5	96.1	1.6	11.4	90.2	83.2	1.5
UB007759a	15.1	89.1	97.4	2.0		14.7	62.2	
UB007718a	15.2	70.5	98.4	1.0	32.8	81.8	83.9	3.0
UB007670a	18.2	99.7	99.8	2.0	34.3	92.2	65.9	3.4
UB007693a	18.7	59.2	69.3	0.9		56.0	45.2	
UB007725a	24.3	61.4	87.2	0.4		56.5	65.9	
UB007735a	29.0	67.4	83.6	1.5	51.4	44.6	58.0	1.5
UB007685a	36.5	68.2	97.7	2.5	44.8	53.2	49.5	2.4
UB007720a	40.3	49.6	94.0	1.7	42.1	57.2	57.5	1.2
UB007741a	49.7	52.7	98.2	3.1	52.2	48.4	53.3	3.2
UB007705a	66.8	44.8	52.6	1.0		28.8	46.9	
UB007722a		38.5	44.9			48.0	67.1	
UB007743a		8.4	27.3		25.9	66.5	60.0	1.0

Figure 3.12 Dose response experiments for the initial KLHL20 inhibitor hits. IC₅₀ values were determined by both Alphascreen and HTRF assays with good agreement. The data in this figure are produced by Mr James Bennett.

3.3. Discussion

The Cullin-RING E3 ligase KLHL20 has been shown to ubiquitinate some half a dozen protein targets that link its activities to diverse processes including autophagy, hypoxia, cancer and Alzheimer's disease [17]. In this chapter I performed the first structural and biochemical analyses of KLHL20 to elucidate how it engages its substrates through the Kelch β -propeller domain. Structural studies required the identification of a short DAPK1 peptide motif that subsequently enabled crystallization. As a result, I was able to solve the structure of the KLHL20 Kelch domain in complex with DAPK1 peptide at 1.1 Å resolution. The structural and biochemical studies identified a central 'LPDLV' motif in the DAPK1 epitope for KLHL20 interaction. A similar 'GPDVL' motif was also identified for KLHL20 interaction in EPAS1, another reported KLHL20 binding partner. The interacting motifs within other KLHL20 substrates remain to be defined at the same level, but are likely to form a similar pattern of hydrophobic and charge-charge interactions.

The low micromolar binding of the DAPK1 peptide to KLHL20 is comparable in affinity to substrates of the SPOP E3 ligase [19], which similarly assembles into a Cullin-RING ligase complex through CUL3 [20]. However, this is weaker than the low nanomolar binding observed for NRF2 interaction with the Kelch domain of KEAP1 [21]. These differences may reflect the strict regulation of NRF2, which ensures its constitutive degradation, or the requirement for substrate co-adaptors as found for KLHL12 [22]. Alternatively, there may be differences in affinity between the binding of DAPK1 peptide and the death domain in the context of the full length DAPK1 or KLHL20 proteins. Death domains are well known protein interaction modules that fold as a bundle of six α -helices [23]. While the isolated death domain of DAPK1 appears to be intrinsically disordered, it is possible that other DAPK regions contribute to its proper folding [15]. The critical 'LPDLV' motif of DAPK1 maps to the predicted α 3 helix. To understand how this might interact with KLHL20 in the context of the full death domain I built a homology model of human DAPK1 using the MyD88 protein structure (PDB 3MOP) as a template and

ICM-Pro software (Molsoft) (Figures 3.13A-B) [24]. Superposition of the 'LPDLV' motifs revealed good agreement between the model, the MyD88 template and the crystal structure (Figures 3.13B-C). Overall, the helical turn of the DAPK1 peptide was a good match to the folding of the α 3 helix (Figures 3.13B-C). Consequently, residues in the 'LPDLV' motif were closely aligned in the different structures (Figure 3.13C). However, structural deviations in the flanking peptide residues suggest that their interactions are less certain in the context of the folded death domain. Most importantly, the key interacting residues of DAPK1 were exposed providing a surface epitope for KLHL20 to bind. The model suggests that the α 3 helix of the death domain can insert into the relatively wide pocket of KLHL20 to recapitulate the observed peptide interaction without steric hindrance. There is some precedent for such an arrangement from the structure of KEAP1 bound to the DLG motif of NRF2, which also formed an extended helical structure [25].

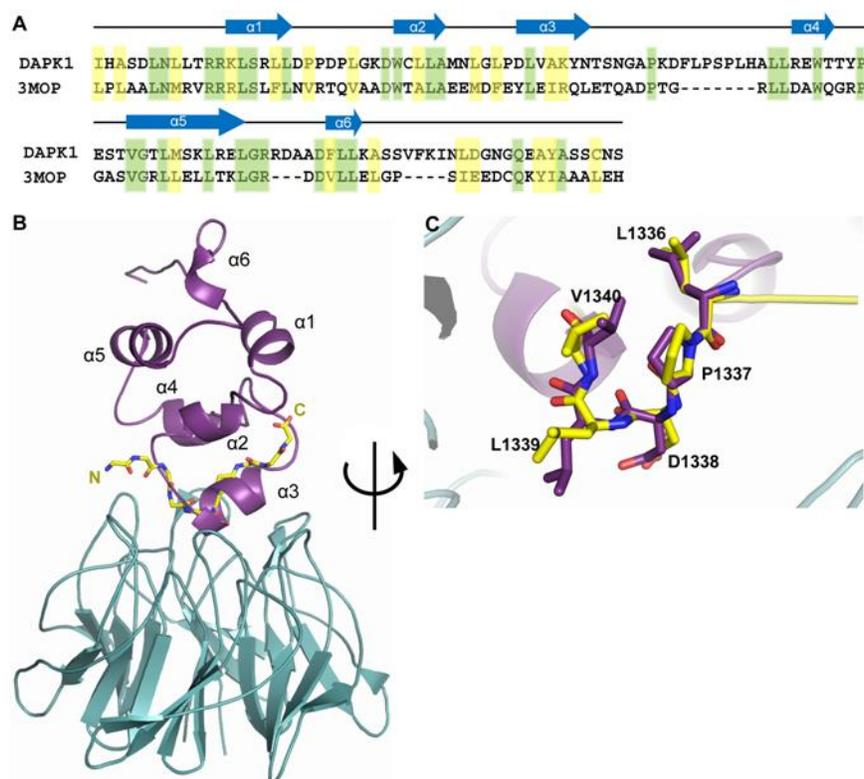


Figure 3.13. Homology model of DAPK1 death domain in complex with KLHL20 Kelch domain. (A) Sequence alignment of the death domains of DAPK1 and MyD88 (PDB 3MOP, chain A). **(B)** Superposition of the KLHL20 Kelch domain-DAPK1 structure (cyan/yellow) and a homology model of the DAPK1 death domain (purple; template PDB 3MOP) based on the critical ‘LPDLV’ motifs. A similar comparison with the template structure is shown in Figure S3. **(C)** Close-up view showing good agreement between the helical conformation of the ‘LPDLV’ motif in the crystallized DAPK1 peptide and the homology model ($\alpha 3$).

Of note, previous studies of KEAP1 have characterized the binding of both unmodified and phosphorylated peptides (for example, the ‘ETGE’-containing motif from NRF2 and the ‘phospho-STGE’-containing motif of sequestosome-1/p62)[13, 14, 26]. To date, no posttranslational modifications have been reported for the death domain of DAPK1 (www.phosphosite.org; [27]) and no phosphorylated substrate motif for KLHL20 has been identified yet. Nonetheless, other substrates of KLHL20 may similarly substitute a phosphorylated residue for the aspartate found in the ‘LPDLV’ motif of DAPK1. It is known for

example that KLHL20 binds specifically to the activated pool of ULK1 to terminate autophagy [8].

The structure of KLHL20 at atomic resolution provides a robust template for structure-based drug design. The structure identifies a hydrophobic substrate pocket that appears attractive for small molecule inhibitor development. Indeed, a few best non-covalent hits screened from 10K MIDAS library exhibited low micromolar IC50s for KLHL20. Of note, the hydrophobic interaction surface in KLHL20 includes an exposed cysteine residue (Cys356) that lies within 4 Å of the bound peptide (Figure 3.10). This cysteine is accessible for modification and proximal to KLHL20 Trp326, another key DAPK1 interacting residue (Figures 3.6A and 3.10). Thus, KLHL20 may be a promising target for screening against covalent inhibitor libraries as well as non-covalent inhibitor libraries.

3.4. References

1. Lee, Y.R., et al., *The Cullin 3 substrate adaptor KLHL20 mediates DAPK ubiquitination to control interferon responses*. EMBO J, 2010. **29**(10): p. 1748-61.
2. Yuan, W.-C., et al., *A Cullin3-KLHL20 Ubiquitin Ligase-Dependent Pathway Targets PML to Potentiate HIF-1 Signaling and Prostate Cancer Progression*. Cancer Cell, 2011. **20**(2): p. 214-228.
3. Chen, H.Y., et al., *KLHL39 suppresses colon cancer metastasis by blocking KLHL20-mediated PML and DAPK ubiquitination*. Oncogene, 2015. **34**(40): p. 5141-51.
4. Gómez Ravetti, M., et al., *Uncovering molecular biomarkers that correlate cognitive decline with the changes of hippocampus' gene expression profiles in Alzheimer's disease*. PloS one, 2010. **5**(4): p. e10153-e10153.
5. Yuan, W.C., et al., *A Cullin3-KLHL20 Ubiquitin Ligase-Dependent Pathway Targets PML to Potentiate HIF-1 Signaling and Prostate Cancer Progression*. Cancer Cell, 2011. **20**(2): p. 214-28.
6. Inbal, B., et al., *DAP kinase links the control of apoptosis to metastasis*. Nature, 1997. **390**(6656): p. 180-184.
7. Dai, L., et al., *DAPK Promoter Methylation and Bladder Cancer Risk: A Systematic Review and Meta-Analysis*. PloS one, 2016. **11**(12): p. e0167228-e0167228.
8. Liu, C.C., et al., *Cul3-KLHL20 Ubiquitin Ligase Governs the Turnover of ULK1 and VPS34 Complexes to Control Autophagy Termination*. Mol Cell, 2016. **61**(1): p. 84-97.
9. Woik, N., et al., *Kelch-like ECT2-interacting protein KLEIP regulates late-stage pulmonary maturation via Hif-2 α in mice*. Disease models & mechanisms, 2014. **7**(6): p. 683-692.
10. Higashimura, Y., et al., *Kelch-like 20 up-regulates the expression of hypoxia-inducible factor-2 α through hypoxia- and von Hippel-Lindau tumor suppressor protein-*

- independent regulatory mechanisms*. Biochemical and Biophysical Research Communications, 2011. **413**(2): p. 201-205.
11. Cohen, O., et al., *DAP-kinase participates in TNF-alpha- and Fas-induced apoptosis and its function requires the death domain*. The Journal of cell biology, 1999. **146**(1): p. 141-148.
 12. Schumacher, F.R., et al., *Structural and biochemical characterization of the KLHL3-WNK kinase interaction important in blood pressure regulation*. Biochem J, 2014. **460**(2): p. 237-46.
 13. Lo, S.C., et al., *Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling*. EMBO J, 2006. **25**(15): p. 3605-17.
 14. Padmanabhan, B., et al., *Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer*. Mol Cell, 2006. **21**(5): p. 689-700.
 15. Dioletis, E., A.J. Dingley, and P.C. Driscoll, *Structural and functional characterization of the recombinant death domain from death-associated protein kinase*. PLoS One, 2013. **8**(7): p. e70095.
 16. Canning, P., et al., *Structural basis for Cul3 protein assembly with the BTB-Kelch family of E3 ubiquitin ligases*. J Biol Chem, 2013. **288**(11): p. 7803-14.
 17. Chen, H.Y., C.C. Liu, and R.H. Chen, *Cul3-KLHL20 ubiquitin ligase: physiological functions, stress responses, and disease implications*. Cell Div, 2016. **11**: p. 5.
 18. Higashimura, Y., et al., *Kelch-like 20 up-regulates the expression of hypoxia-inducible factor-2alpha through hypoxia- and von Hippel-Lindau tumor suppressor protein-independent regulatory mechanisms*. Biochem Biophys Res Commun, 2011. **413**(2): p. 201-5.
 19. Zhuang, M., et al., *Structures of SPOP-substrate complexes: insights into molecular architectures of BTB-Cul3 ubiquitin ligases*. Mol Cell, 2009. **36**(1): p. 39-50.
 20. Errington, W.J., et al., *Adaptor protein self-assembly drives the control of a cullin-RING ubiquitin ligase*. Structure, 2012. **20**(7): p. 1141-53.
 21. Tong, K.I., et al., *Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model*. Mol Cell Biol, 2006. **26**(8): p. 2887-900.
 22. McGourty, C.A., et al., *Regulation of the CUL3 Ubiquitin Ligase by a Calcium-Dependent Co-adaptor*. Cell, 2016. **167**(2): p. 525-538 e14.
 23. Ferrao, R. and H. Wu, *Helical assembly in the death domain (DD) superfamily*. Curr Opin Struct Biol, 2012. **22**(2): p. 241-7.
 24. Abagyan, R., et al., *Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search*. Proteins, 1997. **Suppl 1**: p. 29-37.
 25. Fukutomi, T., et al., *Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1*. Mol Cell Biol, 2014. **34**(5): p. 832-46.
 26. Ichimura, Y., et al., *Phosphorylation of p62 activates the Keap1-Nrf2 pathway during selective autophagy*. Mol Cell, 2013. **51**(5): p. 618-31.
 27. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations*. Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20.
 28. Hao, B., et al., *Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases*. Mol Cell, 2007. **26**(1): p. 131-43.
 29. Uljon, S., et al., *Structural Basis for Substrate Selectivity of the E3 Ligase COP1*. Structure, 2016. **24**(5): p. 687-696.
 30. Picaud, S. and P. Filippakopoulos, *SPOTing Acetyl-Lysine Dependent Interactions*. Microarrays (Basel), 2015. **4**(3): p. 370-88.

31. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
32. Emsley, P., et al., *Features and development of Coot*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
33. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.

Chapter 4

Structural basis for recruitment of DVL1 to KLHL12

4.1. Introduction

Several BTB-Kelch members play critical roles in embryonic development. Among them, KLHL12 is well-known for its inhibitory effect on the Wnt pathway.

The Wnt pathway regulates cell polarity development via a cascade of signalling transduction. In the Wnt signalling pathway, a GPCR receptor Frizzled is stimulated by a class of secreted glycoproteins, so called Wnt ligands. Upon activation, Frizzled facilitates Wnt signalosome assembly and ultimately leads to β -catenin stabilization (canonical Wnt signalling) or a β -catenin-independent effect (non-canonical Wnt signalling) [1, 2].

Dishevelled (DVL) proteins take part in both β -catenin dependent or independent pathways. DVL proteins comprise three domains – an N-terminal DIX domain, a central PDZ, and a C-terminal DEP domain [3] (Figure 4.1). The DIX domain mediates its self-polymerization and interaction with Axin to assist Wnt/ β -catenin signalosome assembly [4, 5]. The central PDZ domain directly binds to Frizzled, which is essential to activate Wnt pathway by recruiting DVL proteins to plasma membrane [6]. Translocation of Axin in the signalosomes to the periphery of the inner cell membrane inhibits β -catenin degradation and results in β -catenin accumulation in the nucleus [7, 8]. The DEP domain interacts with DAAM1 (dishevelled associated activator of morphogenesis 1), and then activates β -catenin independent pathway [9, 10].

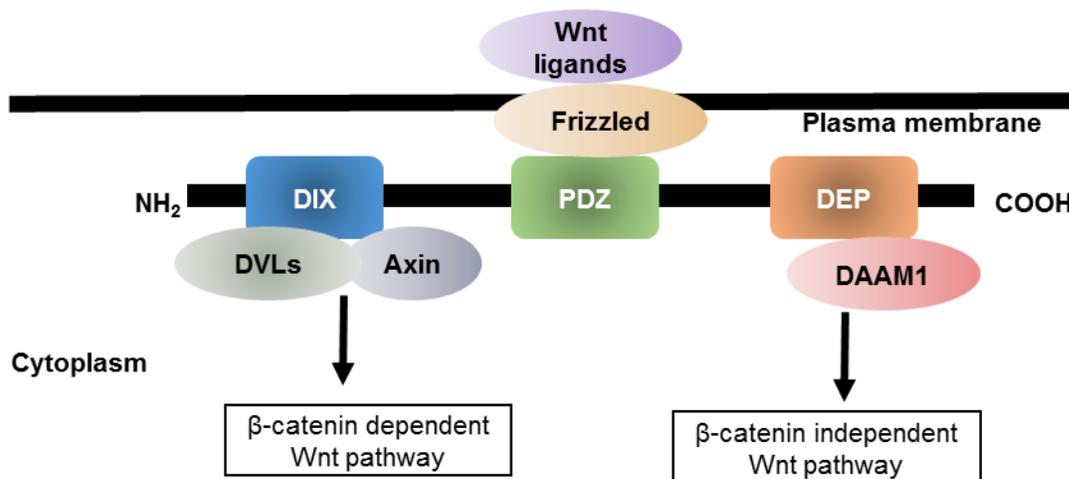


Figure 4.1 A schematic diagram showing the domain composition of DVL proteins. DVL is recruited to the plasma membrane via its PDZ-Frizzled interaction. The DIX domain facilitates Wnt signalosome formation by recruiting Axin and self-polymerization. The signalosome transduces Wnt signalling through the β -catenin-dependent pathway.

Vertebrates possess three conserved paralogs DVL1-3. All three paralogs function similarly in Wnt signal transduction. However, temporal and spatial diversity in expression pattern is also observed in various species [11]. As a result, the three paralogs exhibit only partial redundancy and some distinct functions. For example, DVL2 and 3 play critical roles in the axial skeleton and heart development, whereas DVL1 plays a specific role cardiac remodelling and regeneration [2, 12, 13].

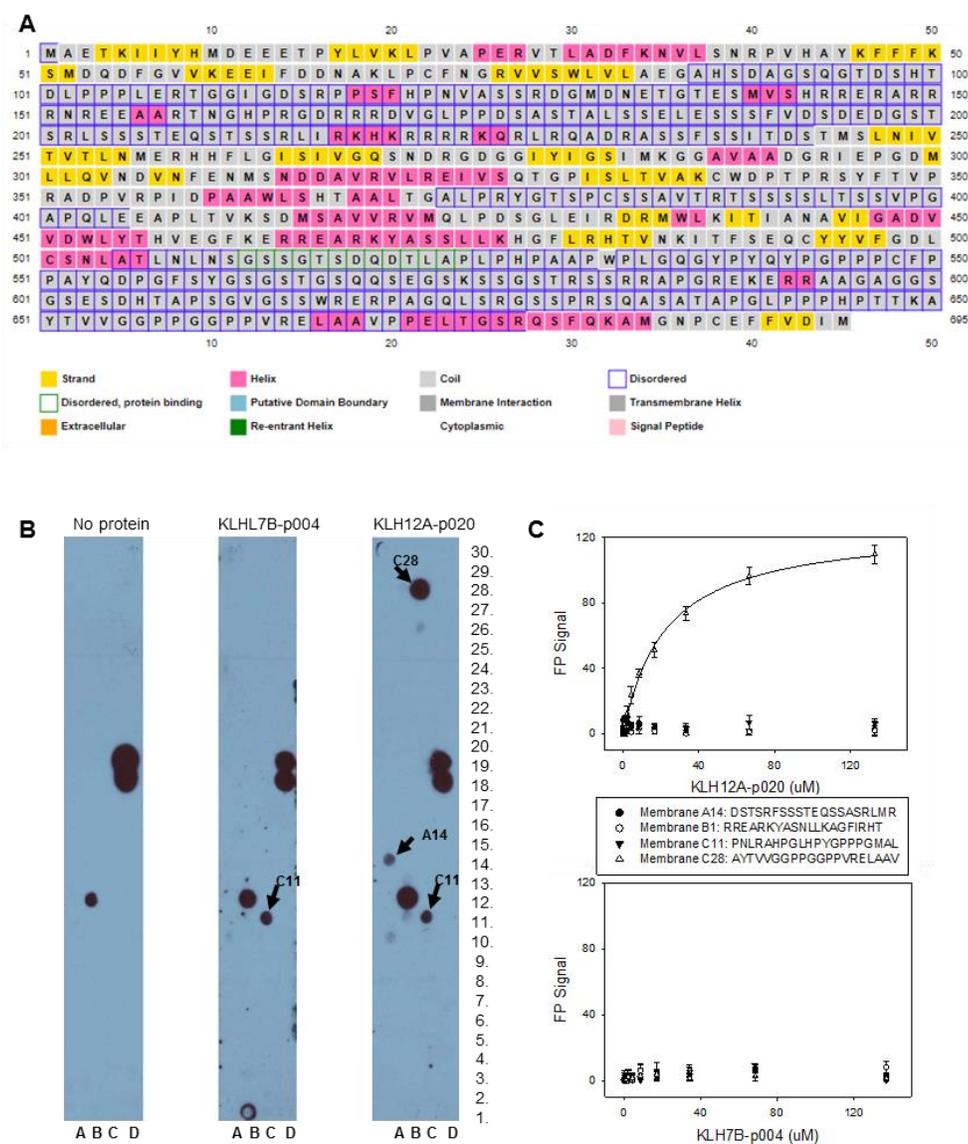
DVL proteins are regulated by the ubiquitin proteasome system. Mass spectrometry proteomics and domain deletion immunoprecipitation experiments identified KLHL12 as a direct binder of DVL2 and 3. KLHL12 binds to DVL C-terminal region (492–716 a.a. in DVL1) and recruits DVL to Cullin3 E3 ligase complex. The Cullin3-KLHL12 complex then poly-ubiquitinates DVL for proteasomal degradation, thereby inhibiting Wnt signalling pathway. In vivo study in zebrafish confirmed KLHL12 as an antagonist of DVL-mediated convergent-extension during gastrulation [14]. There are additional inhibitory mechanisms of this process

to prevent DVL poly-ubiquitination and therefore promote Wnt signalling. For example, NRX expels KLHL12 from DVL complexes [15] and PLEKHA4 sequesters the KLHL12 in plasma membrane clusters [16, 17].

KLHL12 was first discovered in *Drosophila* and subsequently in *Homo sapiens* as an amplified gene under apoptosis-inducing protein treatment [18]. To date, several substrates of KLHL12 have been identified in addition to DVLS. Interestingly, KLHL12 modifies them with different ubiquitin chain types. KLHL12 targets the Dopamine D4 receptor for poly-ubiquitination, yet the subsequent process remains elusive [19]. The Ub K29R K48R K63R triple mutant largely blocked the D4 ubiquitination by KLHL12, however, none of the single mutants phenocopied the loss of ubiquitination [20]. Notably, KLHL12 was also shown to promote non-Lysine ubiquitination of the Dopamine Receptors D4.2 and D4.4, but not of the D4.7 variant [21]. In another case, CUL3–KLHL12 catalyses the mono-ubiquitination of SEC31, a COPII component, and drives large COPII coat assembly [22]. In addition, the SEC31 recruitment to Cullin3-KLHL12 complex requires PEF/ALG2 as target-specific co-adaptors which are regulated by calcium. These findings suggest that the bound substrate may influence the ubiquitination step, perhaps by promoting recruitment of different E2-ubiquitin pairs that favour different linkages.

Among all the KLHL12 substrates, DVLS appear the most extensively studied in terms of the substrate-E3 interaction and down-stream ubiquitination-degradation axis. The C-terminal region in DVL1 where the KLHL12-interacting motif resides is predicted to be structurally disordered (Figure 4.2A). Therefore, our collaborator in SGC Toronto mapped the DVL1 recognition epitope using SPOT peptide technology and a fluorescence polarization assay. The DVL1 peptide 650-669a.a. was mapped for KLHL12 recruitment and was shown to bind to KLHL12 Kelch domain with $K_D = 22 \mu\text{M}$ (Figure 4.2B-C).

With this knowledge, I employed the SPOT peptide technology to refine the minimal epitope through peptide truncation and alanine scanning experiments before structural studies to define the binding mode. DVL1, as well as the two paralogs DVL2 and DVL3, were also addressed in ubiquitination and stability assays in this chapter. By providing both structural and functional data, I aim to unveil the substrate recruitment mechanism of KLHL12 and discuss the potential consensus in recruiting other substrates.



(Experiments were performed by Gregory Wasney from SGC Toronto)

(Figure legend on the next page)

Figure 4.2 Mapping of the DVL1 binding motif for KLHL12 recruitment. **(A)** Secondary structure prediction of DVL1. **(B)** SPOT peptide array. Each spot was printed as a 20-mer DVL1 peptide with a 15 residue frameshift at each consecutive position. Spots D28 and D19 were printed as poly-His peptides as positive controls for the antibody detection. Arrays were incubated with purified 6xHis-KLHL12 Kelch domain, washed and then KLHL12 binding detected using anti-His HRP-conjugated antibody. Binding was observed at four sites – A14, B12, C11 and C28. As negative controls, replicate spots were probed with either antibody alone or KLHL7, a protein of the same family. The negative controls revealed non-specific antibody binding to B12, and a unique putative KLHL7 binding site at spot B1. **(C)** Fluorescence polarisation assay. Four peptides of A14, B1, C11 and C28 from the SPOT array were labelled with FITC. KLHL12 and KLHL7 Kelch domains were assayed for binding to each of these peptides, respectively. Spot C28 (DVL1 residues 650-669) bound to KLHL12 with $K_D = 22 \pm 2 \mu\text{M}$, but not to KLHL7.

4.2. Results

4.2.1. A 'PGGPP' motif in DVL1 is critical for KLHL12 interaction.

To map the minimal DVL1 epitope for KLHL12 interaction, I utilised the SPOT technology for alanine scanning mutagenesis and peptide truncation experiments. A peptide array was printed on my behalf by Dr Sarah Picaud (SGC, University of Oxford) based on the 20-mer peptide previously identified (650-AYTVVGGPPGGPPVRELAHV). The results from these experiments were in excellent agreement and identified DVL1 residues Pro658 to Pro662 as critical for KLHL12 interaction (Figure 4.3). Mutation of these residues and C-terminal deletion of Pro662 drastically reduced KLHL12 binding, whereas N-terminal deletion of Pro658 abolished all detectable binding (Figure 4.3). Deletion of Val663 and Arg664 also reduced KLHL12 binding, whereas the mutation of them did not, suggesting that the backbone atoms at these positions may suffice for interaction. Other deletions and mutations outside of this region were well tolerated.

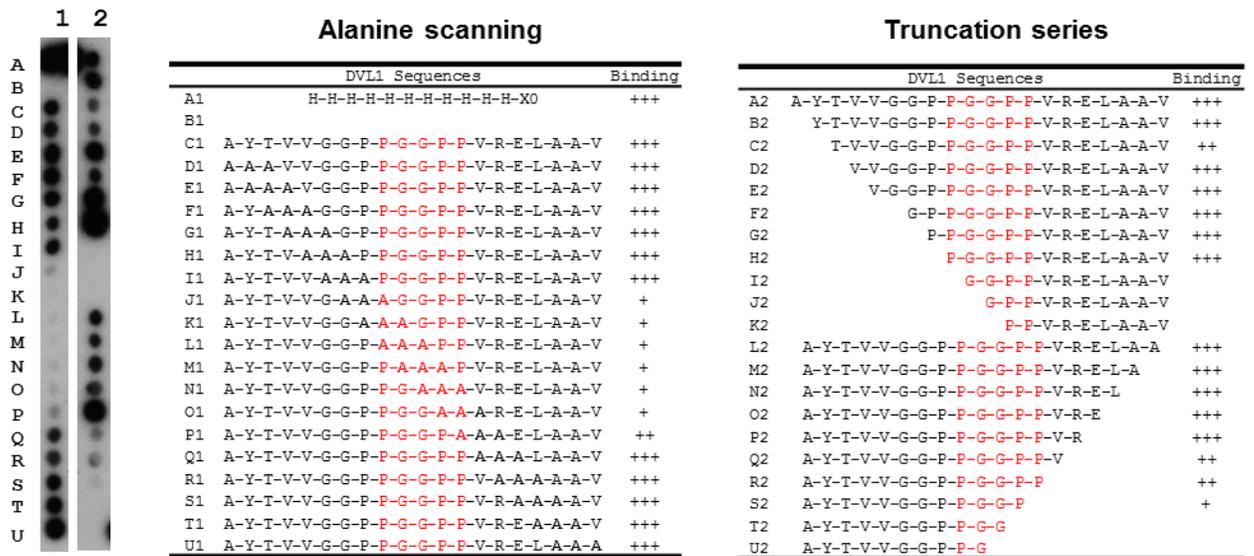


Figure 4.3 A 'PGGPP' motif in DVL1 is critical for KLHL12 interaction. DVL1 peptide variants were printed in SPOT peptide arrays. Column 1 peptides explored triple-alanine scanning mutagenesis, whereas column 2 explored N and C-terminal truncations. Arrays were incubated with purified 6xHis-KLHL12 Kelch domain, washed and then binding detected with anti-His antibody. KLHL12 binding was disrupted upon mutation or deletion of the 'PGGPP' sequence motif in DVL1.

4.2.2. Crystallisation trials and structure determination for a KLHL12-DVL1 complex

The 20-mer DVL1 peptide (650-659 a.a.) identified in the original SPOT assay and a shorter 15-mer peptide (650-664 a.a.) based on the 'PGGPP' epitope were used for crystallisation trials with the KLHL12 Kelch domain, respectively.

This KLHL12-DVL1 complex buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP, was concentrated to 8 mg/mL. Crystal trials were performed using the vapour diffusion method with sitting drops with a variety of coarse crystallisation screens as described in section 2.11. Initial crystals were obtained with the 15-mer DVL1 peptide, but not with the 20-mer peptide, at 4°C using precipitant solutions containing 0.2 M ammonium acetate, 30% PEG4000 or 8000,

pH ranging from 4.5 to 7.5. Initial crystals exhibited as needle-like objects and diffracted to 2.4 Å with streaky spots, suggesting misalignment of lattice planes due to plastic deformation. I attempted to improve crystal quality by microseeding with coarse screens and fine screens employing a matrix of salt/pH/PEG variation based on the initial crystallisation conditions. Variable crystals were obtained as shown in Figure 4.4.

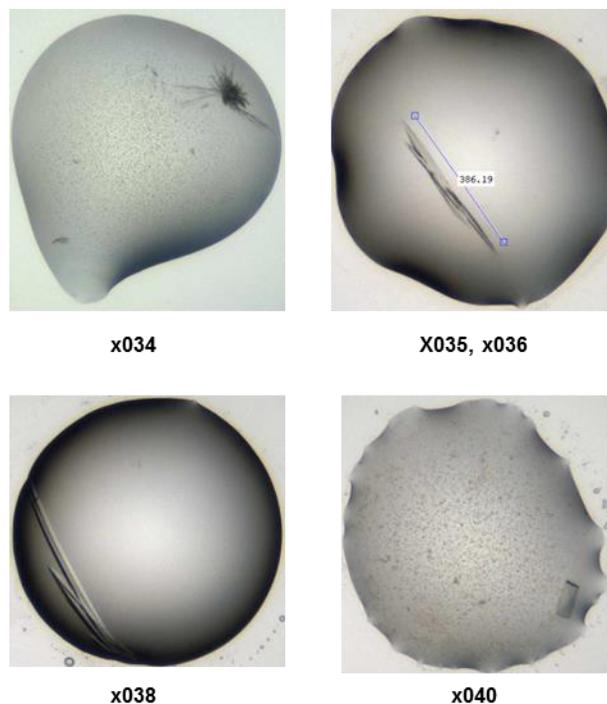


Figure 4.4 Crystallization trials of KLHL12-DVL1. Representative crystals from various crystallisation trials. Identified crystals were named sequentially with those shown being named x034 to x040.

Diffraction quality crystals in space group $P 1 2_1 1$ or $P 1$ were obtained for the KLHL12-DVL1 complex (Table 4.1). After initial processing, a diffraction dataset from crystal x034 was selected for further structure refinement due to its relatively high resolution and the quality of the electron density map for the bound DVL1 peptide (Table 4.1). Data anisotropy was corrected prior to molecular replacement in Aimless scaling. The resulting structure of the KLHL12-DVL1 was solved by molecular replacement using Phaser with KLHL12 (PDB ID:

2VPJ) as search model and then refined at 2.38 Å resolution with four complexes in the asymmetric unit. Group B factor was restrained in the early refinement cycles. TLS and NCS were employed at the later stage in the refinement. Data collection and refinement statistics are shown in Table 4.2.

Table 4.1 Summary of KLHL12-datasets

Identifiers	Conditions	Space groups	Cell Dimesions- a,b,c (Å); α , β , γ (°)	Comments
PDB: 2VPJ	30% PEG4000, 0.2M ammonium acetate, 0.1M acetate pH 4.6	P 1 2 ₁ 1	44.64, 61.55, 45.52; 90, 111.75, 90	1.85 Å resolution, deposited Apo structure
x034	30% PEG4000, 0.2M ammonium acetate, 0.1M acetate pH 4.5	P 1 2 ₁ 1	80.23, 73.15, 101.85; 90, 94.50, 90	2.38 Å resolution, KLHL12-DVL1 complex, refined to Rfree/Rwork 0.26/0.21
x035	20% PEG3350, 10% ethylene glycol, 0.1M bis-tris-propane pH 6.5, 0.2M sodium bromide	P 1 2 ₁ 1	80.99, 72.80, 102.93; 90, 90.82, 90	2.78 Å resolution, MR fine, Apo structure
x036	20% PEG3350, 10% ethylene glycol, 0.1M bis-tris-propane pH 6.5, 0.2M sodium bromide	P 1 2 ₁ 1	81.36, 74.13, 101.65; 90, 95.51, 90	2.23 Å resolution, KLHL12-DVL1 complex, refined to Rfree/Rwork 0.28/0.23, map quality for DVL1 is poorer than that is in x034 dataset
x038	20% PEG3350, 10% ethylene glycol, 0.1M bis-tris-propane pH 6.5, 0.2M sodium nitrate	P 1	68.516, 73.344, 73.779; 66.538, 64.84, 69.531	2.34 Å resolution, intensity moments suggest significant twinning (>5%)
x040	0.15M ammonium sulfate, 25% PEG4000, 15% glycerol	P 1 2 ₁ 1	80.237, 73.62, 101.489; 90, 94.923, 90	2.54 Å resolution, MR fine, map quality for DVL1 almost as good as that in x034 dataset

Table 4.2 Data collection and refinement statistics

Structure of human KLHL12-DVL complex, under refinement	
Data collection	
Beamline	Diamond Light Source, I24
Wavelength (Å)	0.9686
Resolution range (Å)	79.98 - 2.383 (2.469 - 2.383)
Space group	P 1 2 ₁ 1
Unit cell dimensions	
<i>a,b,c</i> (Å)	80.225 73.145 101.845
α, β, γ (°)	90 94.501 90
Total reflections	212254 (31109)
Unique reflections	47094 (4664)
Completeness (%)	99.57 (99.59)
Mean I/sigma(I)	5.8 (2.2)
CC1/2	0.984 (0.816)
R-merge	0.185(0.679)
Refinement	
Reflections used in refinement	47015 (4659)
Reflections used for R-free	2359 (223)
R-work	0.2114 (0.2599)
R-free	0.2591 (0.3168)
Number of non-hydrogen atoms	9629
RMS deviation (bonds, Å)	0.011
RMS deviation (angles, °)	1.27
Ramachandran favored (%)	94.79
Ramachandran allowed (%)	5.12
Ramachandran outliers (%)	0.09
Rotamer outliers (%)	3.69
Average B-factor (Å ²)	25.77

*Values in brackets show the statistics for the highest resolution shells. RMS indicates root-mean-square.

4.2.3. Structure overview of KLHL12 Kelch domain bound to DVL1 peptide.

The co-structure traced the full KLHL12 Kelch domain from residues 279 to 567 (Figure 4.5A). The structure of the KLHL12 Kelch domain resembles those of previously determined Kelch structures. Superimposing the four complexes in the asymmetric unit reveals nearly identical folding and conformations in the KLHL12-DVL1 interface, despite some variabilities in the flexible loops away from the KLHL12 binding pocket (Figure 4.5B).

The DVL1 peptide was traced from G656 to P662 in chain E, G655 to P662 in chain F, G656 to V663 in chain G and H (Figure 4.5C-D), allowing structural analysis of the key 'PGGPP' motif. The four DVL1 chains, though defined to various extents, exhibit backbone conformations that are in perfect agreement. The electron density was not resolved for other N and C-terminal residues, suggesting lack of contacts to stabilise these flanking residues.

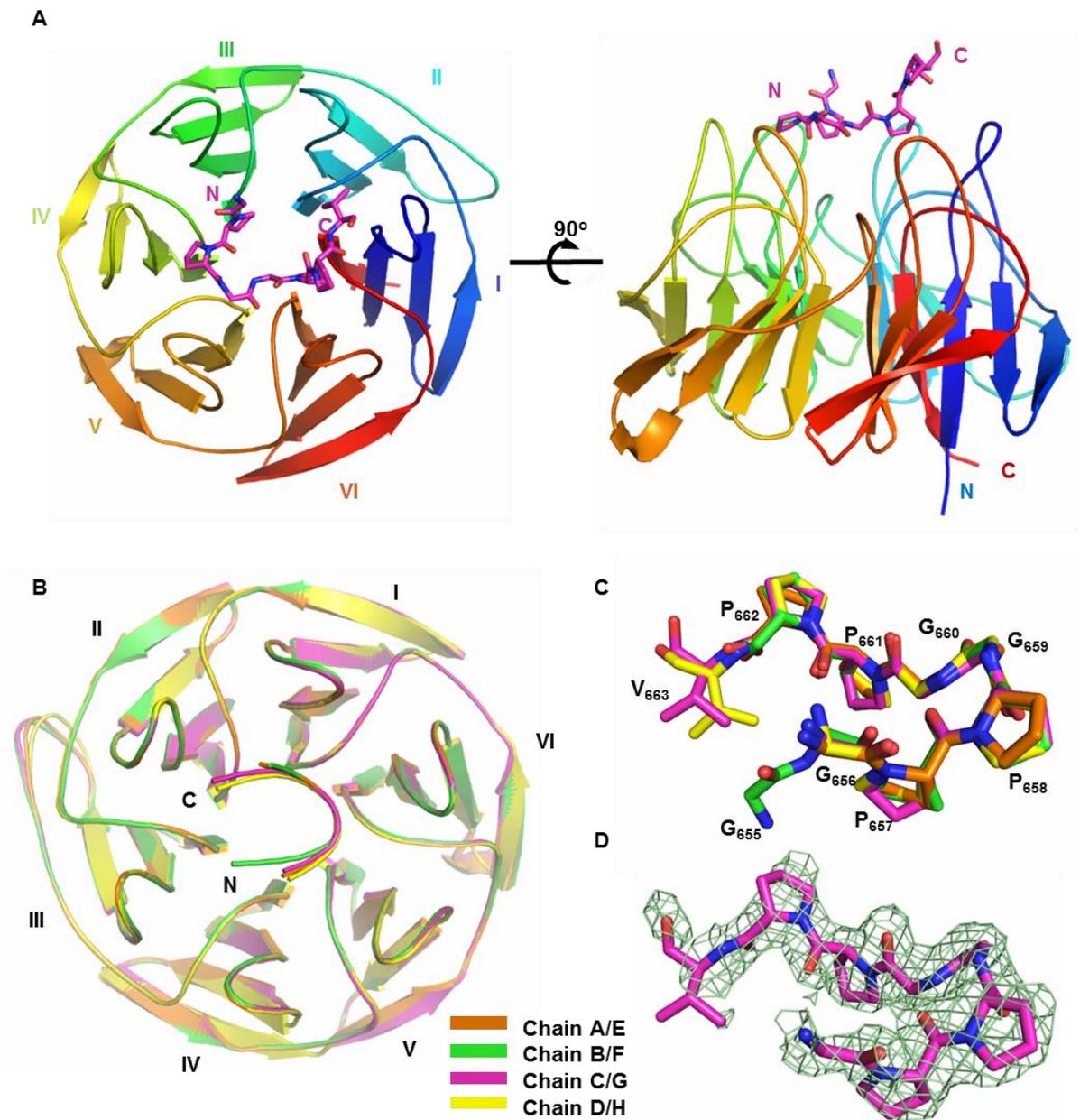


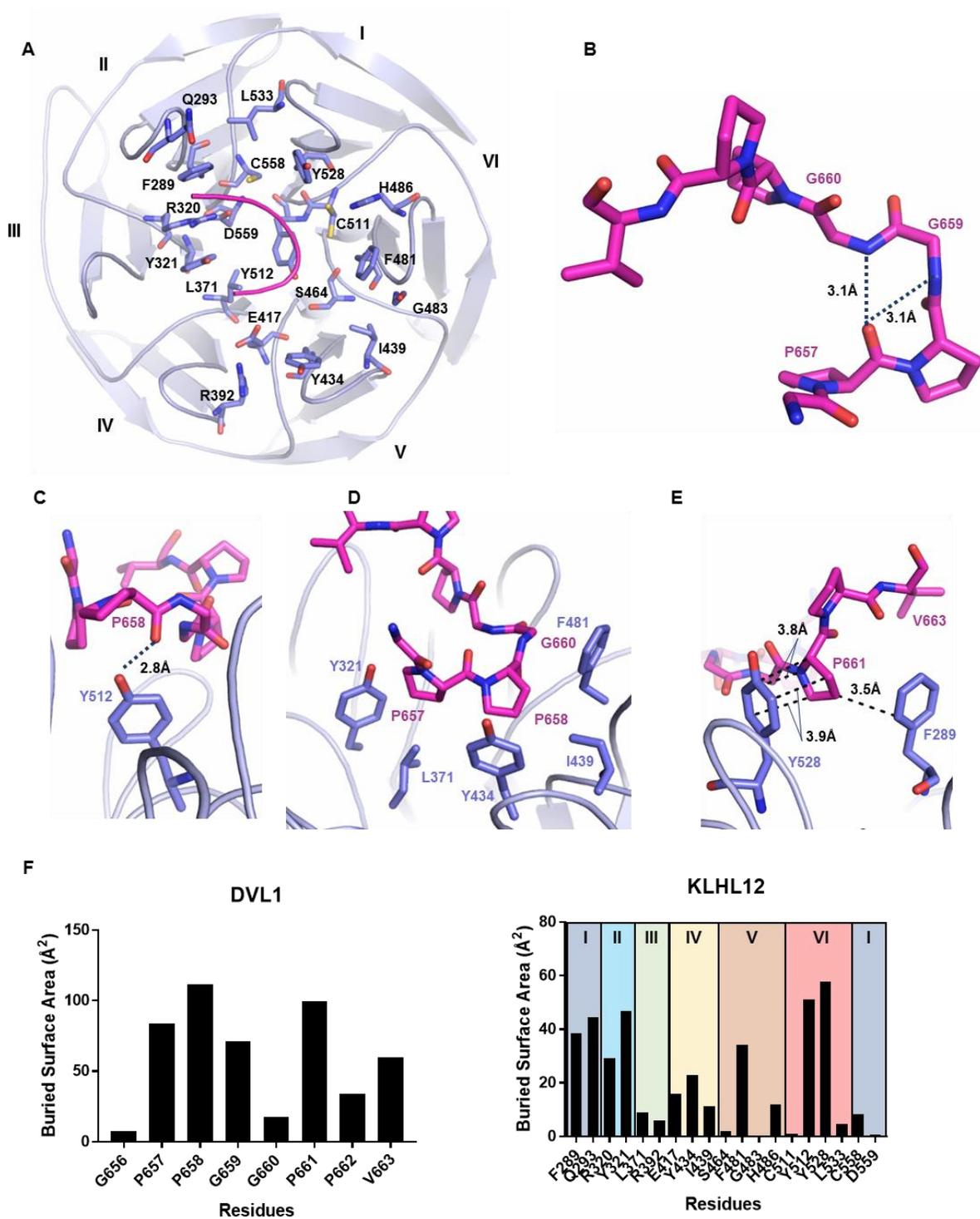
Figure 4.5 A 2.38Å structure of KLHL12 Kelch domain bound to DVL1 peptide. (A) Overview of the structure of KLHL12 Kelch (rainbow ribbon) in complex with DVL1 peptide (purple sticks). Kelch repeats forming blades I to VI are labelled. N/C termini for both KLHL12 and DVL1 are labelled. (B) Superposition of four KLHL12-DVL1 complexes in the same asymmetric unit cell, ribbon representation. (C) Superposition of four DVL1 chains in the same asymmetric unit, stick representation. (D) 2Fo-Fc electron density map (green mesh) for the DVL1 peptide contoured at 0.8 σ

4.2.4. Interactions in the KLHL12 Kelch domain-DVL1 complex.

The bound DVL1 peptide packed between Kelch domain blades III and IV at the resolved peptide N-terminus and blades II and III at its C-terminus (Figure 4.6A). The peptide adopts a U-shaped turn conformation that is stabilised by the weak intramolecular hydrogen bonds between the carbonyl of Pro657 and the amides of the following di-Gly (Gly659 and Gly660) (Figure 4.6B). The conformation allows the DVL1 peptide to form contacts with all six blades of KLHL12 as summarised in Figure 4.6F.

The 'PGGPP' motif accounts for the majority of the resolved DVL1 residues and is responsible for most of the interactions with KLHL12. The first proline in the motif, Pro658, forms the only hydrogen bond in the complex interface with KLHL12 Tyr512 (Figure 4.6C). Pro658 and Gly659, together with Pro657, make extensive hydrophobic interactions through van der Waals contacts with blade III, IV, V and VI. Of note, these three residues pack against the hydrophobic grooves supported by KLHL12 Tyr321, Leu371 and Tyr434, Ile439, Phe481. The formation of these two hydrophobic cores are evidenced by the buried surface areas of these residues (Figure 4.6F). The second proline in the motif, Pro661, is sandwiched by KLHL12 Phe289 and Tyr528. In addition to the hydrophobic interactions between DVL1 Pro661/Val663 and KLHL12 Phe289/Val533, DVL1 Pro661 forms a prolyl-aromatic ring-ring stacking interaction with KLHL12 Tyr528 (Figure 4.6E). Due to the nitrogen heteroatom, the pyrrolidine ring exhibits a partially positive π face, whereas the π face in the aromatic ring is electron rich [23]. Hence this proline-tyrosine interaction is favoured not only because of the hydrophobic ring stacking driving force, but also due to the electrostatic force in the presence of dipoles. The C-terminal proline in the motif, Pro662, however, is oriented away from the binding interface. In the SPOT peptide arrays (Figure 4.3), mutation or deletion of this residue reduced the KLHL12 binding drastically but did not completely abolish it. It is likely that Pro662

fulfils a conformational role by facilitating a slight turn in the DVL1 peptide to avoid its steric hindrance with blade II of KLHL12.



(Figure legend on next page)

Figure 4.6 Interactions in the KLHL12 Kelch domain-DVL1 complex. (A) An overview of the DVL1-binding residues in KLHL12. The interacting residues in KLHL12 are labelled. KLHL12 Kelch is coloured cyan and DVL1 peptide is shown as pink ribbon. (B) Peptide intramolecular interactions within the DVL1 backbone. (C) The hydrogen bond interaction in the complex interface shown in dotted lines. (D) Hydrophobic interactions and (E) proline-aromatic interactions in the complex interface. Distances are shown as labels alongside with dashed lines for each interaction. (F) Buried interface surface areas for interacting residues in the KLHL12 Kelch domain-DVL1 complex. The KLHL12 residues are colour coded to indicate the blades they reside in.

4.2.5. KLHL12-induced degradation of DVL1 is dependent on the ‘PGGPP’ motif

To validate the identified ‘PGGPP’ motif as a site regulating DVL1 recognition and degradation by KLHL12, I tested the DVL1 interaction and stability in HEK293T cells. To disrupt the interaction site I generated two full length DVL1 mutants in which the critical 657-‘PPGGPP’ motif was either mutated to ‘AAAAAA’ or deleted. Immunoprecipitation (Figure 4.7A) of full length Flag-KLHL12 and HA-DVL1 full length variants showed that the wild-type (WT) DVL1 was robustly bound to KLHL12 whereas the DVL1 mutants were only recovered at background level.

I then investigated whether DVL1 mutants were stabilised due to lack of the ‘PGGPP’ motif. Equal amounts of DVL1 constructs were transfected into HEK293T cells in the presence or absence of full length KLHL12. As is shown in Figure 4.7B, co-expression of full length KLHL12 and DVL1 WT caused a striking reduction in the DVL1 level compared to expressing DVL1 WT alone. However, the DVL1 mutants exhibited resistance to KLHL12 co-expression.

Taken together, these data indicated that the ‘PGGPP’ motif is critical for both DVL1 recruitment and degradation by KLHL12.

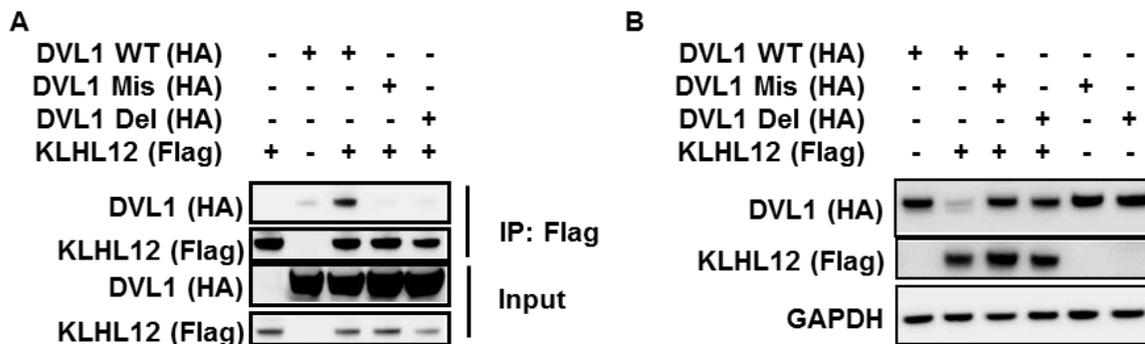


Figure 4.7 Mutations in the DVL1 ‘PGGPP’ motif impair DVL1 binding and degradation by KLHL12. (A) Full length DVL1 variants and full length KLHL12 were co-transfected into HEK293T cells as indicated. Flag-KLHL12 was immunoprecipitated with anti-Flag antibody. DVL1 WT was robustly co-purified with KLHL12, whereas DVL1 mutants were only recovered at the background level of the beads alone. (B) DVL1 variants were transfected into HEK293T cells with or without full length KLHL12 as indicated. Cells were treated with cycloheximide (CHX) for 1 hour before harvesting. DVL1 protein levels were detected by Western blot. GAPDH was probed as a loading control. The experiment was performed once.

4.2.6. ‘PGXPP’ is a consensus motif in DVL1, 2 and 3 for KLHL12 interaction.

The ‘PGGPP’ motif is conserved across the three DVL paralogs, except for an alanine substitution at the second glycine in DVL2 and 3 (Figure 4.8A). Notably, DVL1 Gly660 at this position had the lowest buried surface area in the DVL1 co-structure among the ‘PGGPP’ motif (Figure 4.6F). To investigate the effect of this alanine substitution, I modelled a G660A mutation in the DVL1 structure (Figure 4.8B). The methyl side chain of alanine faces the KLHL12 substrate binding pocket, yet does not form any significant contacts within a 4 Å distance. A SPOT peptide array also confirmed that the Gly to Ala substitution is tolerated for KLHL12 binding (Figure 4.8C). To conclude, ‘PGXPP’ is a consensus site for KLHL12 to bind and degrade DVL1/2/3, where ‘X’ refers to Ala or Gly.

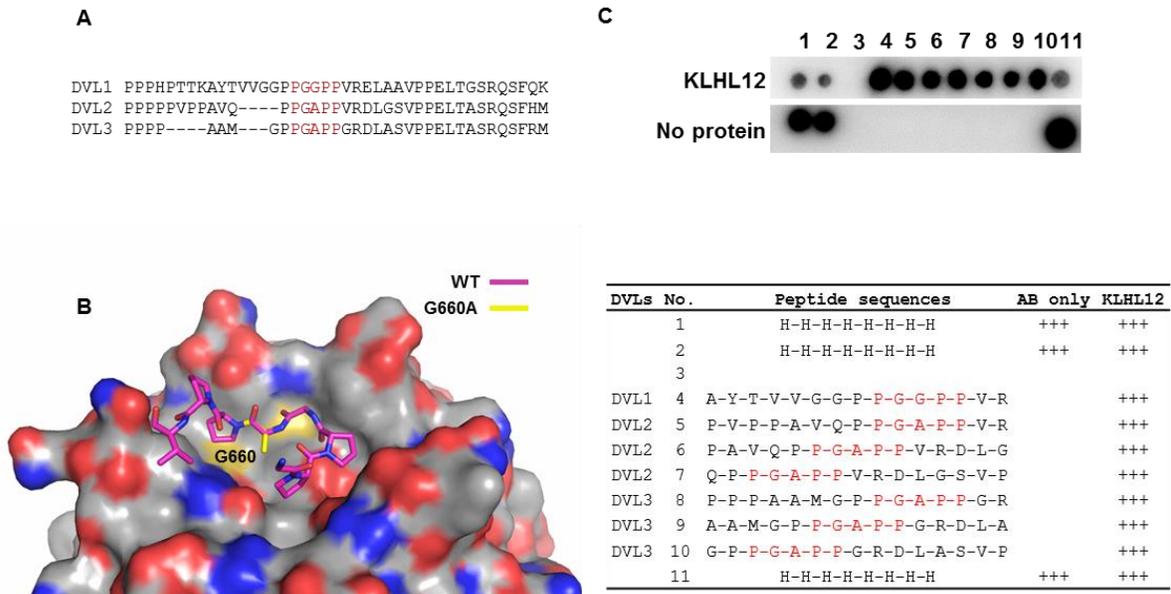


Figure 4.8 ‘PGXPP’ is a consensus motif in DVL1, 2 and 3 for KLHL12 interaction. (A) The PGXPP motif is conserved across three DVL1 paralogs. (B) Structural model incorporating the G660A mutation. Wild type DVL1 peptide is shown as purple sticks. The mutated alanine is shown as yellow sticks. (C) DVL1, 2 and 3 peptides containing ‘PGXPP’ were printed in SPOT peptide arrays. Arrays were incubated with purified 6xHis-KLHL12 Kelch domain, washed and then binding detected with anti-His antibody.

4.3. Discussion

KLHL12 inhibits Wnt signalling by ubiquitination and degrading DVL proteins. Here I mapped the 658-PGGPP sequence as the key KLHL12-interacting motif in DVL1 via SPOT arrays. Mutating or deleting the ‘PGGPP’ motif in full length DVL1 caused its stabilisation due to loss of KLHL12 interaction. Crystallisation of KLHL12-DVL1 complex was enabled with a short DVL1 peptide motif (650-664 a.a.). The structure of this complex was then determined at 2.4 Å resolution.

The structures of bound KLHL12 and apo KLHL12 (PDB: 2VPJ) are overall similar, with 0.99 Å RMSD for all atoms. Superposition of the apo and bound KLHL12 reveals only minor movement of side chains in the substrate binding pocket (Figure 4.9), suggesting minimal conformation dynamics in the substrate binding event.

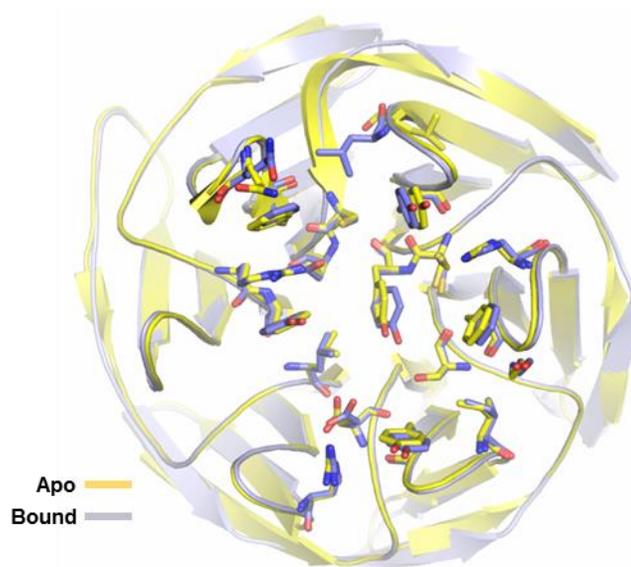


Figure 4.9 Superposition of KLHL12 Kelch domain in apo form and bound form. Apo structure – 2VPJ. Apo structure is coloured in yellow and bound structure in cyan.

Of note, though both the apo and bound crystals formed in space group $P 1 2_1 1$, the crystal-packing contacts are distinct in these two states. In the apo state, the $\beta 3$ - $\beta 4$ loops in blades I and VI of KLHL12 chain A form hydrogen bonds and hydrophobic interactions with the BC loops in blades I, IV, V, and VI of KLHL12 chain B (Figure 4.10A). Interestingly, these crystal-packing contacts engaged KLHL12 residues that also interact with the DVL1 peptide in the bound structure, for example, Q293, Y434, Y528 and L533 (Figure 4.10B-D). Given protein crystallization is predominantly dependent on entropic effects and the driving force appears to be the release of ordered water from the sites of crystal contacts [24], the KLHL12 substrate binding pocket appears low entropy, i.e. low conformational freedom of side chains. In the bound state, the KLHL12 substrate binding pocket is occupied by the DVL1 peptide, which does not provide available side chains for crystal packing. As a result, the KLHL12 molecules in this complex formed crystal-packing contacts mainly through their AB and BC loops (Figure 4.10E). The reduction in low-entropy patches in KLHL12 surface may result in plasticity in crystal packing, which explains the streaky spots in X-ray diffraction of KLHL12-DVL1 crystals.

On the other hand, the crystal packing adopted by apo KLHL12 crystals hinders the accessibility of the substrate binding pocket, therefore is not suitable for fragment soaking for drug screening. The bound KLHL12 crystals provide an alternative packing, which can be used for peptide displacement based fragment soaking.

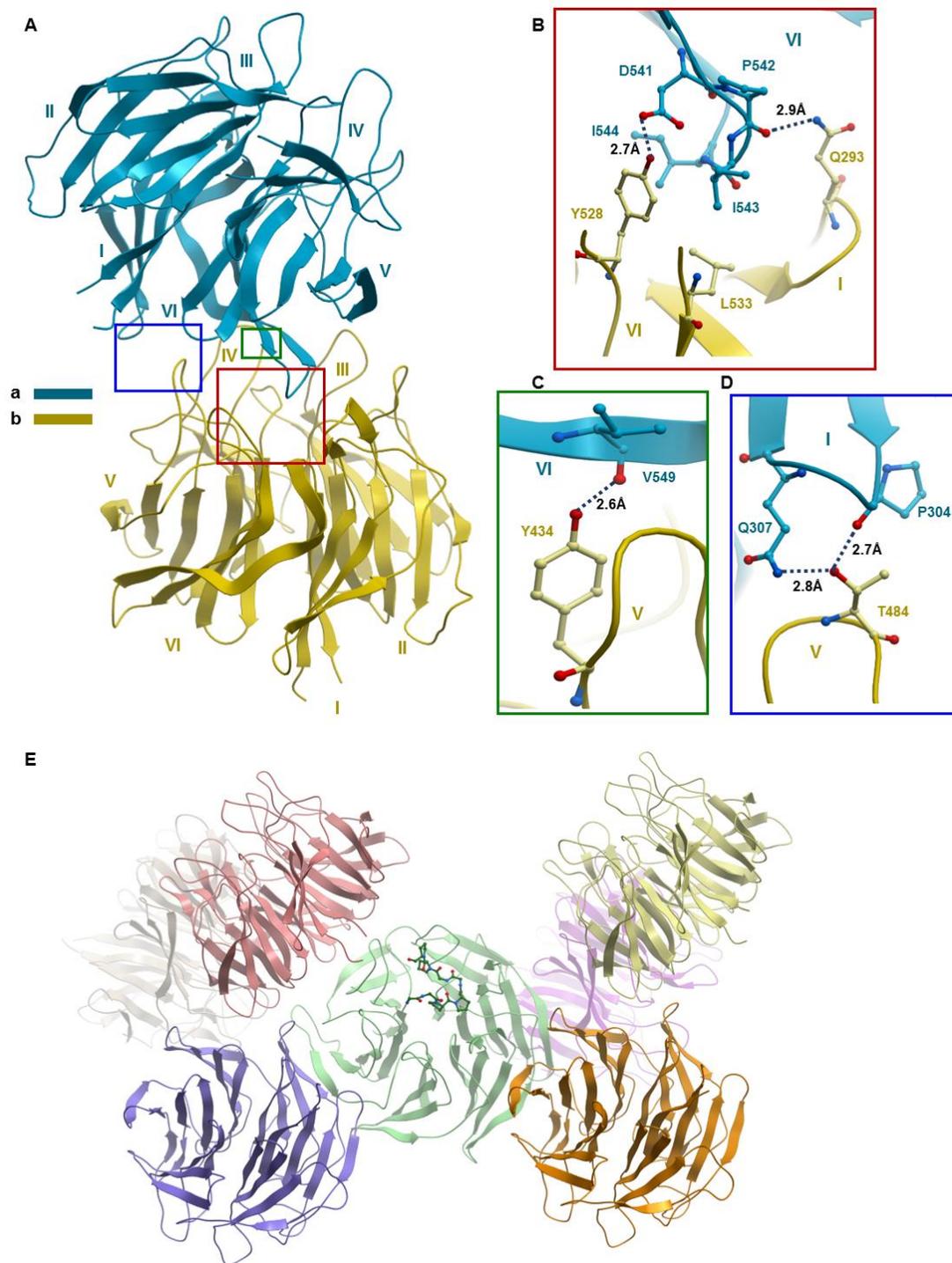


Figure 4.10 Crystal-packing contact analysis for apo and bound KLHL12. (A) Crystal-packing contacts of apo KLHL12. Blades are numbered for chains A and B, respectively. Contacts are marked with red, green and blue boxes. (B-D) Zoom-in snapshots of the contacts in red, green and blue boxes marked in (A). Hydrogen bonds are labelled with dashed lines and distances. (E) Crystal packing for bound KLHL12, showing the crystal packing does not bury the substrate binding pocket.

Not only does the KLHL12 substrate binding pocket show low apparent entropy, the DVL1 peptide in the co-structure also shows geometrical rigidity due to two di-prolines in close proximity (657-PPGGPP). Proline has a limited Phi value of about $-63 \pm 15^\circ$. As is shown in table 4.3, in the DVL1 peptide, all four prolines adopt a *trans* conformation, in which the theoretical energy minimum is at Phi = -75° and Psi = 145° . Together these features suggest the binding of KLHL12 and DVL1 is driven by strict conformational constraints.

Table 4.3 Dihedral angles of prolines in the bound DVL1 peptide

DVL1	Phi (°)	Psi(°)	Omega(°)
P657	-69.2	142.3	-177.8
P658	-56.8	128.4	173.4
P661	-76.6	153.4	171.4
P662	-61.9	148.1	176.5

The 'PGXPP' motif is conserved across DVL1/2/3 for KLHL12 binding. In addition to one hydrogen bond formed between KLHL12 Tyr512 and DVL1 Pro658, hydrophobic interactions are the dominant force stabilising the KLHL12-DVL1 complex interface (Figure 4.6). In 2012, Jin et al performed a mutagenesis study when investigating SEC31 ubiquitination by KLHL12. They found that alanine mutations at positions 289FG, 369RGL, 434YDG and 510RCY in KLHL12 causes loss of function [22]. Interestingly, these mutations are also predicted to disrupt the hydrophobic interactions between KLHL12 and DVL1 based on my structure (Figure 4.6D-E). Therefore I speculate that the recognition motifs within other KLHL12 substrates/binding partners are likely to bind to the same pocket and exhibit a similar pattern of 'PGXPP'. Indeed, PEF, the co-adaptor to mediate SEC31 recruitment to KLHL12, contains a 18-'PGAPP' motif in its N-terminal domain (1-109 a.a.) which is a domain known to bind to KLHL12 [25]. A recent study on a KLHL12 natural inhibitor PLEKHA4 indicates that it binds to KLHL12 through its proline-rich domain (167-357 a.a.) [16]. This proline-rich domain also contains a 177-'PGGPP' motif which exactly matches the recognition motif in DVL1. Beside my data with all three DVL

paralogs, whether 'PGXPP' is a consensus motif in more KLHL12 Kelch domain binding partners remains to be validated with additional experiments on PEF1 and PLEKHA4.

To confirm the 'PGGPP' motif is critical for DVL1 ubiquitination by KLHL12, I also attempted to perform an ubiquitination assay in HEK293T cells (data not shown). However the data were elusive due to high background and subtle differences in the ubiquitination level of DVL1 WT and mutants, indicating interference from other E3s that regulate DVLS. To date, there are three additional E3s found to ubiquitinate DVLS. ITCH, an E3-ubiquitin ligase belonging to the HECT-type E3 subfamily, specifically targets phosphorylated DVLS for degradation [26]. Wnt-5 α activation of JNK phosphorylates NEDDL4, which in turn ubiquitinates DVLS for degradation via polyubiquitination at K-6, K-27 and K-29. Thus, NEDDL4 acts as a negative feedback regulator of Wnt pathway activation [27]. Upon starvation-induced metabolic stress, DVLS are ubiquitinated and degraded by the von-Hippel-Lindau protein (pVHL) [28]. To optimise my ubiquitination assay, more stringent controls and adjustment on DNA transfection amount need to be considered to increase signal to background contrast.

4.4. References

1. Logan, C.Y. and R. Nusse, *The Wnt signaling pathway in development and disease*. Annu Rev Cell Dev Biol, 2004. **20**: p. 781-810.
2. Malekar, P., et al., *Wnt signaling is critical for maladaptive cardiac hypertrophy and accelerates myocardial remodeling*. Hypertension, 2010. **55**(4): p. 939-45.
3. Wong, H.C., et al., *Structural basis of the recognition of the dishevelled DEP domain in the Wnt signaling pathway*. Nature Structural Biology, 2000(1072-8368 (Print)).
4. Schwarz-Romond, T., et al., *The DIX domain of Dishevelled confers Wnt signaling by dynamic polymerization*. Nat Struct Mol Biol, 2007. **14**(6): p. 484-92.
5. Kishida, S., et al., *DIX domains of Dvl and axin are necessary for protein interactions and their ability to regulate beta-catenin stability*. Molecular and Cellular Biology, 1999(0270-7306 (Print)).
6. Wong, H.C., et al., *Direct binding of the PDZ domain of Dishevelled to a conserved internal sequence in the C-terminal region of Frizzled*. Molecular Cell, 2003(1097-2765 (Print)).
7. Angers, S. and R.T. Moon, *Proximal events in Wnt signal transduction*. Nat Rev Mol Cell Biol, 2009. **10**(7): p. 468-77.
8. Nusse, R. and H. Clevers, *Wnt/beta-Catenin Signaling, Disease, and Emerging Therapeutic Modalities*. Cell, 2017. **169**(6): p. 985-999.

9. Gao, C. and Y.G. Chen, *Dishevelled: The hub of Wnt signaling*. Cell Signal, 2010. **22**(5): p. 717-27.
10. Pan, W.J., et al., *Characterization of function of three domains in dishevelled-1: DEP domain is responsible for membrane translocation of dishevelled-1*. Cell Research, 2004(1001-0602 (Print)).
11. Gentzel, M. and A. Schambony, *Dishevelled Paralogs in Vertebrate Development: Redundant or Distinct?* Front Cell Dev Biol, 2017. **5**: p. 59.
12. Hamblet, N.S., et al., *Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure*. Development, 2002. **129**(24): p. 5827-38.
13. Gentzel, M., et al., *Distinct functionality of dishevelled isoforms on Ca²⁺/calmodulin-dependent protein kinase 2 (CamKII) in Xenopus gastrulation*. Molecular Biology of the Cell, 2015. **26**(5): p. 966-977.
14. Angers, S., et al., *The KLHL12-Cullin-3 ubiquitin ligase negatively regulates the Wnt-beta-catenin pathway by targeting Dishevelled for degradation*. Nat Cell Biol, 2006. **8**(4): p. 348-57.
15. Funato, Y., et al., *Nucleoredoxin sustains Wnt/beta-catenin signaling by retaining a pool of inactive dishevelled protein*. Curr Biol, 2010. **20**(21): p. 1945-52.
16. Shami Shah, A., et al., *PLEKHA4/kramer Attenuates Dishevelled Ubiquitination to Modulate Wnt and Planar Cell Polarity Signaling*. Cell Rep, 2019. **27**(7): p. 2157-2170 e8.
17. Funato, Y., et al., *Nucleoredoxin Sustains Wnt/ β -Catenin Signaling by Retaining a Pool of Inactive Dishevelled Protein*. Current Biology, 2010. **20**(21): p. 1945-1952.
18. Mai, A., S.K. Jung, and S. Yonehara, *hDKIR, a human homologue of the Drosophila kelch protein, involved in a ring-like structure*. Exp Cell Res, 2004. **300**(1): p. 72-83.
19. Rondou, P., et al., *BTB Protein KLHL12 targets the dopamine D4 receptor for ubiquitination by a Cul3-based E3 ligase*. J Biol Chem, 2008. **283**(17): p. 11083-96.
20. Rondou, P., et al., *KLHL12-mediated ubiquitination of the dopamine D4 receptor does not target the receptor for degradation*. Cell Signal, 2010. **22**(6): p. 900-13.
21. Skieterska, K., et al., *KLHL12 Promotes Non-Lysine Ubiquitination of the Dopamine Receptors D4.2 and D4.4, but Not of the ADHD-Associated D4.7 Variant*. PLoS One, 2015. **10**(12): p. e0145654.
22. Jin, L., et al., *Ubiquitin-dependent regulation of COPII coat size and function*. Nature, 2012. **482**(7386): p. 495-500.
23. Hobza, P., *Stacking interactions*. Phys Chem Chem Phys, 2008. **10**(19): p. 2581-3.
24. Derewenda, Z.S. and P.G. Vekilov, *Entropy and surface engineering in protein crystallization*. (0907-4449 (Print)).
25. McGourty, C.A., et al., *Regulation of the CUL3 Ubiquitin Ligase by a Calcium-Dependent Co-adaptor*. Cell, 2016. **167**(2): p. 525-538 e14.
26. Wei, W., et al., *The E3 ubiquitin ligase ITCH negatively regulates canonical Wnt signaling by targeting dishevelled protein*. Mol Cell Biol, 2012. **32**(19): p. 3903-12.
27. Ding, Y., et al., *HECT domain-containing E3 ubiquitin ligase NEDD4L negatively regulates Wnt signaling by targeting dishevelled for proteasomal degradation*. J Biol Chem, 2013. **288**(12): p. 8289-98.
28. Ma, B., et al., *The Wnt Signaling Antagonist Dapper1 Accelerates Dishevelled2 Degradation via Promoting Its Ubiquitination and Aggregate-induced Autophagy*. J Biol Chem, 2015. **290**(19): p. 12346-54.

Chapter 5

Characterisation of the KLHL3-WNK3 complex

5.1. Introduction

KLHL3 controls mammalian blood pressure through regulation on With No Lysine (WNK) kinases. WNKs, together with their downstream targets, SPS1-related proline/alanine-rich kinase (SPAK)/ Oxidative stress-responsive kinase 1 (OSR1), regulate cation-chloride channels to achieve ion homeostasis in the Kidney and Neurons through a cascade of phosphorylation process [1]. Four WNK isoforms (WNK1, 2, 3 and 4) are activated by autophosphorylation on Serine at T-loop (S382 in isoform 1) when the cells are exposed to hypotonic and low $[Cl^-]$ conditions [2]. Activated WNK kinases then stimulate the kinase activity of SPAK/OSR1 by phosphorylating a conserved threonine residue in their kinase activation segments (SPAK Thr233, OSR1 Thr185)[3], which is facilitated by interaction between the SPAK/OSR1 CCT (Conserved C-terminal) domain and WNK RFXV/I peptide motifs [2]. Similarly, the SPAK/OSR1 CCT domain is recruited to RFXV/I motifs in the N-terminus of N[K]CC cation-chloride channels, and thereby phosphorylates conserved threonine residues in their cytoplasmic domains to activate the channels activity [4]. Conversely, phosphorylation on KCC cation-chloride channels by the WNK-SPAK/OSR1 axis presents an inhibiting function [5]. Excessive activity of the WNK-SPAK/OSR1 cascade is the primary cause of PHAI1 (pseudohypoaldosteronism type II) hypertension and hyperkalemia syndrome

KLHL3 degrades WNKs through the ubiquitination-proteolysis process. KLHL3 directly binds WNKs through the Kelch domain and recruits WNKs to the Cullin3^{KLHL3} complex. Mutations in KLHL3 that disrupt interactions with either WNKs or Cullin3 lead to WNK upregulation and cause PHAI1 hypertension and hyperkalemia syndrome [6]. Mutations within a WNK4 non-

catalytic motif (residues 557-567) also cause PHAI hypertension syndrome [4] and have been found to occur within an acidic degron motif for recognition by KLHL3 [7]. The structural basis for the interaction of KLHL3 and WNK4 has been recently determined and clearly defines the disruptive effects of PHAI-associated mutations in the complex interface [8].

The acidic degron motif in WNK4 is conserved among WNK1-4 allowing all four WNK family proteins to bind to KLHL3 [8]. However, WNK3 is notably distinguished by the presence of four amino acid substitutions in its acidic degron motif compared to the 11-mer WNK4 degron that was previously crystallized in complex with KLHL3 (Figure 5.1). These substitutions include two proline residues with potential to alter the peptide backbone. In addition, the WNK3 degron uniquely contains a threonine residue which potentially allows for phosphorylation-dependant regulatory mechanism.

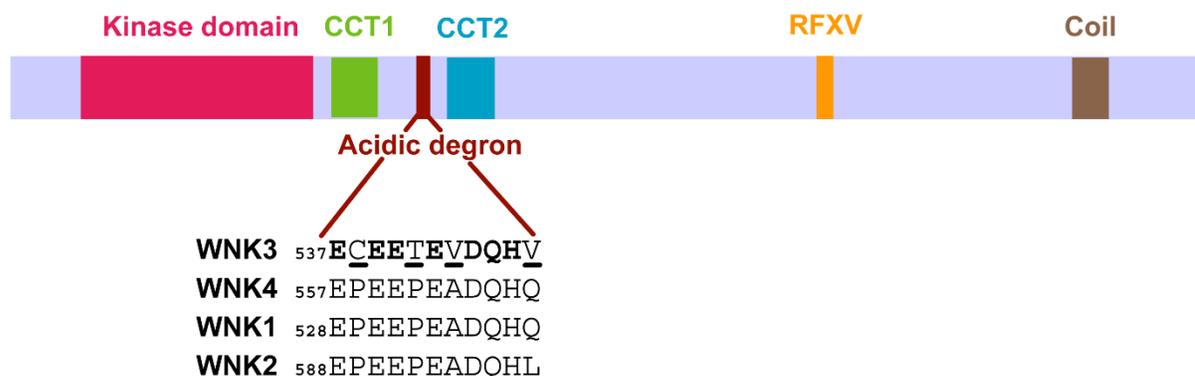


Figure 5.1 Schematic diagram of WNKs. WNKs bind to KLHL2/3 through their highly conserved acidic motifs.

To address the differences, in this chapter I characterised the interaction of KLHL3 with WNK3 using Isothermal Titration Calorimetry and determined a 2.8 Å structure of KLHL3-WNK3 complex. The structural and biochemical studies revealed the rigid substrate-binding pattern of KLHL3. Together with the Fluorescence Polarisation (FP) binding assays performed by our collaborators, we conclude that the threonine phosphorylation in the WNK3 degron would disrupt the complex interface as hypothesized.

5.2. Results

5.2.1. The WNK3 degron peptide binds to the KLHL3 Kelch domain with a K_D in micromolar range.

Firstly I analysed the complex assembly of KLHL3-WNK3 using ITC. The WNK3 degron peptide (537-547a.a. ECEETEVDQHV) was purchased from SEVERN Biotechnology LTD. KLHL3 Kelch domain (298-587 a.a.) was purified from E.coli utilising the protocol described in 2.3. The KLHL3 Kelch domain and WNK3 peptide were buffered in 50 mM HEPES pH 7.5, 300 mM NaCl and 1 mM TCEP. The WNK3 peptide was titrated into KLHL3 protein or buffer alone at 15°C in VP-ITC instrument. A binding mode of 1:1 was chosen for curve fitting after blank data subtraction, using the ITC package in Origin software (Figure 5.2A). The data characterised the binding of KLHL3 and WNK3 as an enthalpy driven event ($\Delta H = -1.329E4$ J mol⁻¹, $\Delta S = -21.7$ J mol⁻¹ K⁻¹) with a K_D of 4.95 μ M ($K_A = 2.02E5$ M⁻¹). To date, no ITC analysis has been reported on the complex assembly of KLHL3-WNK4. To assess the binding of KLHL3 and WNK4 degron in the same system, I titrated the WNK4 degron peptide (549-567a.a.) into the KLHL3 Kelch domain and observed a similar enthalpy driven binding event with a $K_D = 1.1$ μ M (Figure 5.2B). Together, these results suggest WNK3 could bind to KLHL3 with similar thermodynamic features to WNK4 with a low micromolar K_D .

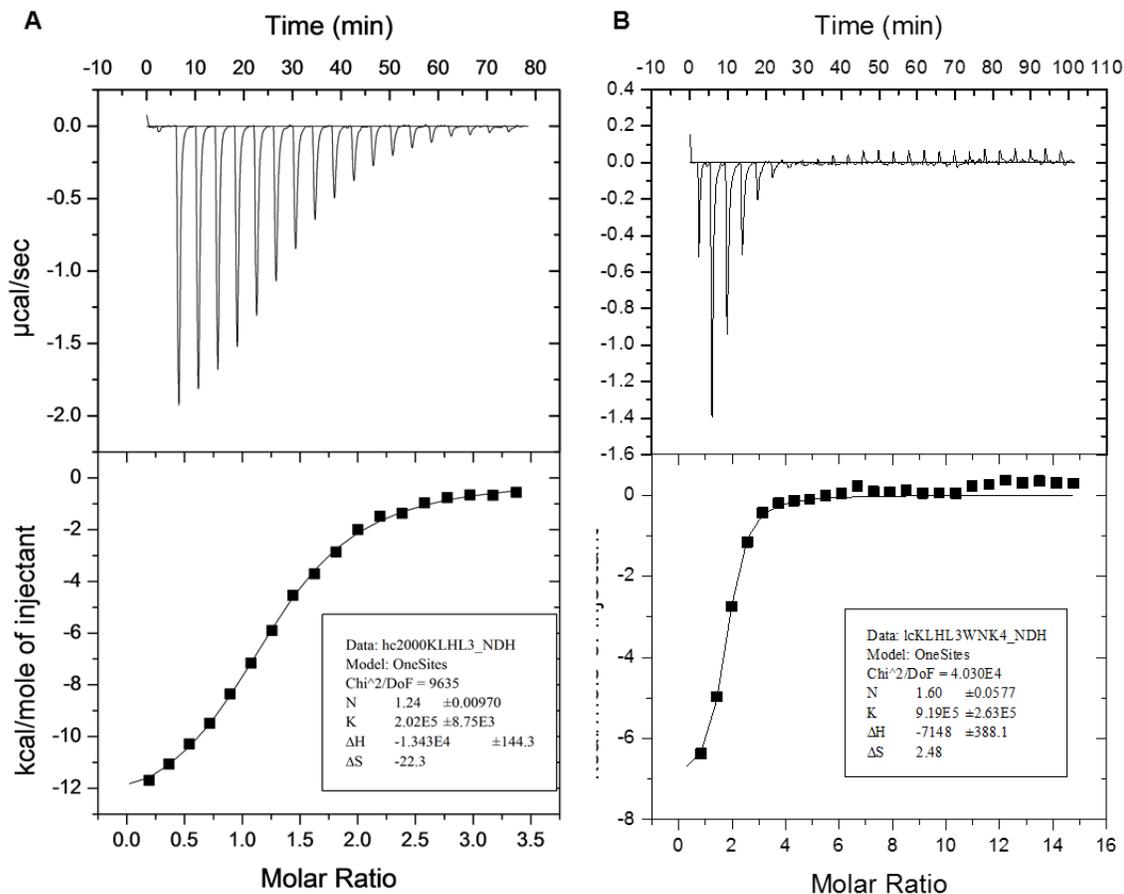


Figure 5.2 ITC showing that KLHL3 binds to WNK3 and WNK4 degrons with low micromolar K_D values. The KLHL3 Kelch domain, WNK3 and WNK4 peptide were prepared in 50 mM HEPES pH 7.5, 300 mM NaCl and 1 mM TCEP. Experiments were performed at 15°C in VP-ITC instrument. Data were replotted after blank (WNK3 to buffer and WNK4 to buffer titration respectively) subtraction. A single site 1:1 binding mode was selected for curve fitting and values are shown in the bottom right panel. (A) 500 μ M WNK3 peptide was titrated into 30 μ M KLHL3 Kelch. (B) 1.2 mM WNK3 peptide was titrated into 12 μ M KLHL3 Kelch.

5.2.2. Structure determination of the WNK3 acidic degron motif in complex with KLHL3

To investigate whether the divergent sequence of the WNK3 degron impacts upon its binding mode in structural aspect, I set up co-crystallisation trials for the KLHL3 Kelch domain and

WNK3 degron peptide. Purified KLHL3 Kelch domain was concentrated to 9 mg/mL (290 μ M) using a 10 kDa MWCO Amicon Ultra concentrator in buffer containing 50 mM HEPES pH7.5, 300 mM NaCl and 1 mM TCEP. 10 mM WNK3 peptide was prepared in the same buffer as the primary stock. WNK3 peptide was added to the KLHL3 protein solution aiming for a final concentration of 1.5 mM. The mixture was incubated on ice for an hour to allow binding before setting up crystal plates.

Initial microcrystals were only obtained at 4°C, and from three conditions – (1) 20% PEG3000 – 0.1M citrate pH 5.5, protein and precipitant volume ratio at 2:1; (2) 8% PEG4000 – 0.1M acetate pH 4.5, protein and precipitant volume ratio at 1:1; (3) 20% PEG3350 – 0.2M ammonium nitrate, protein and precipitant volume ratio at 1:2.

Seed stocks were made out of the microcrystals from above conditions. A fine screen for each condition was designed adopting the matrix of gradient PEG concentration and pH or salt concentration based on the condition composition. A second round of crystallisation trials employing microseeding and fine screens was carried out. Such optimisation increased crystal size and improved crystal morphology (Figure 5.3). The best diffraction crystals were obtained from 6% PEG4K -- 0.1M acetate pH 5.1 at 4°C, protein with a reservoir volume ratio of 2:1.

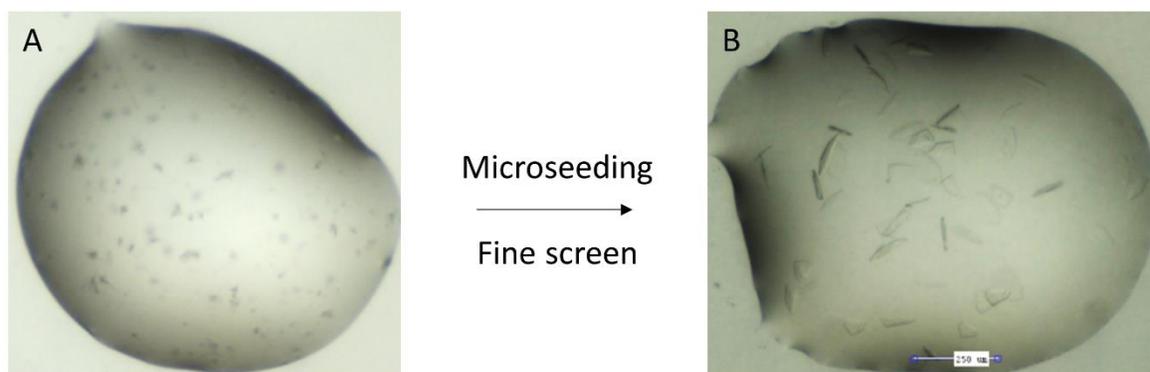


Figure 5.3 Crystallisation optimisation employing microseeding and fine screens. (A) Examples of microcrystals obtained from coarse screens. (B) Crystals yielded in the best diffraction condition after optimisation.

I processed the diffraction data, refined the model and determined a 2.8 Å crystal structure of the 11-mer WNK3 degron in complex with the KLHL3 Kelch domain (PDB: 5NKP) (Figure 5.4). Data processing and refinement statistics are presented in Table 1.

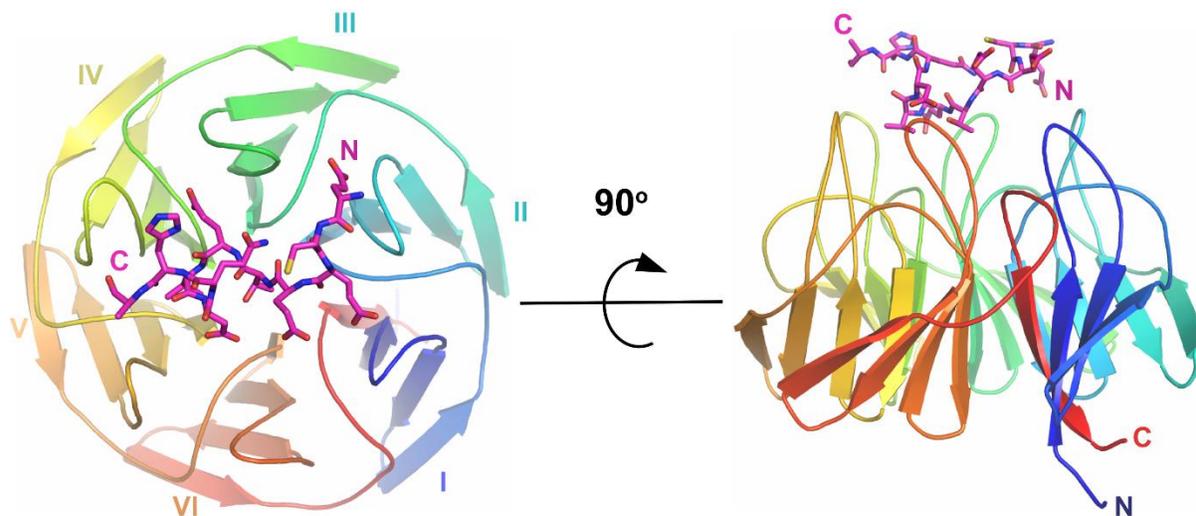


Figure 5.4 Structure overview of KLHL3 Kelch and WNK3 degron peptide complex. KLHL3 Kelch domain is represented with rainbow ribbon and WNK3 degron is represented with purple sticks. KLHL3 Kelch domain structure resembles previously solved kelch structures in general. WNK3 degron peptide sits on the substrate binding pocket which is built by flexible intra-blade loops of KLHL3.

Table 5.1 Data collection and refinement statistics

Structure of human KLHL3-WNK3 complex, PDB 5NKP	
Data collection	
Beamline	Diamond Light Source, I04
Wavelength (Å)	0.9763
Resolution range (Å)	50.02 - 2.8 (2.87 - 2.8)
Space group	C 2 2 2 ₁
Unit cell dimensions	
<i>a,b,c</i> (Å)	84.66, 169.72, 123.86
α, β, γ (°)	90, 90, 90
Total reflections	44732 (4394)
Unique reflections	22366 (2197)
Completeness (%)	99.98 (100.00)
Mean <i>I</i> / σ (<i>I</i>)	8.18 (1.51)
CC1/2	0.983 (0.490)
R-merge	0.092(0.4966)
Refinement	
Reflections used in refinement	22358 (1642)
Reflections used for R-free	1101 (148)
R-work	0.2288 (0.3224)
R-free	0.2472 (0.3653)
Number of non-hydrogen atoms	4566
RMS deviation (bonds, Å)	0.007
RMS deviation (angles, °)	0.75
Ramachandran favored (%)	96.26
Ramachandran allowed (%)	3.57
Ramachandran outliers (%)	0.17
Rotamer outliers (%)	0.43
Average B-factor (Å ²)	37.6

*Values in brackets show the statistics for the highest resolution shells. RMS indicates root-mean-square.

The co-structure was solved in space group $C 2 2 2_1$, with two WNK3 complexes in the asymmetric unit. The entire WNK3 peptide was traced in chain C (Figure 5.5A), but electron density was not resolved for the C-terminal residue Val547 in chain D (Figure 5.5B). In both WNK3 chains, Cys538 was observed to form an intermolecular disulphide-bond that potentially contributed to the observed crystal lattice. Despite distinct crystal contacts, the KLHL3 structure and WNK3 binding mode were highly conserved with the KLHL3-WNK4 complex (PDB 4CH9) suggesting a true representation of the physiologically-relevant assembly.

Consistent with other family structures, the KLHL3 Kelch domain consists of six propeller blades arranged along a central axis. The WNK3 peptide sits atop the substrate binding pocket, which is shaped by the six loops connecting the $\beta 2$ - $\beta 3$ strands, as well as the extended $\beta 4$ - $\beta 1$ loops that connect adjacent blades. Near identical conformations are observed for the peptide backbone in the two WNK3 chains, as well as side chains in the core interface, but several solvent-exposed side chains appear free to adopt alternative conformations, including Glu537, Cys538, Glu540 and His546 (Figures 5.5A and B).

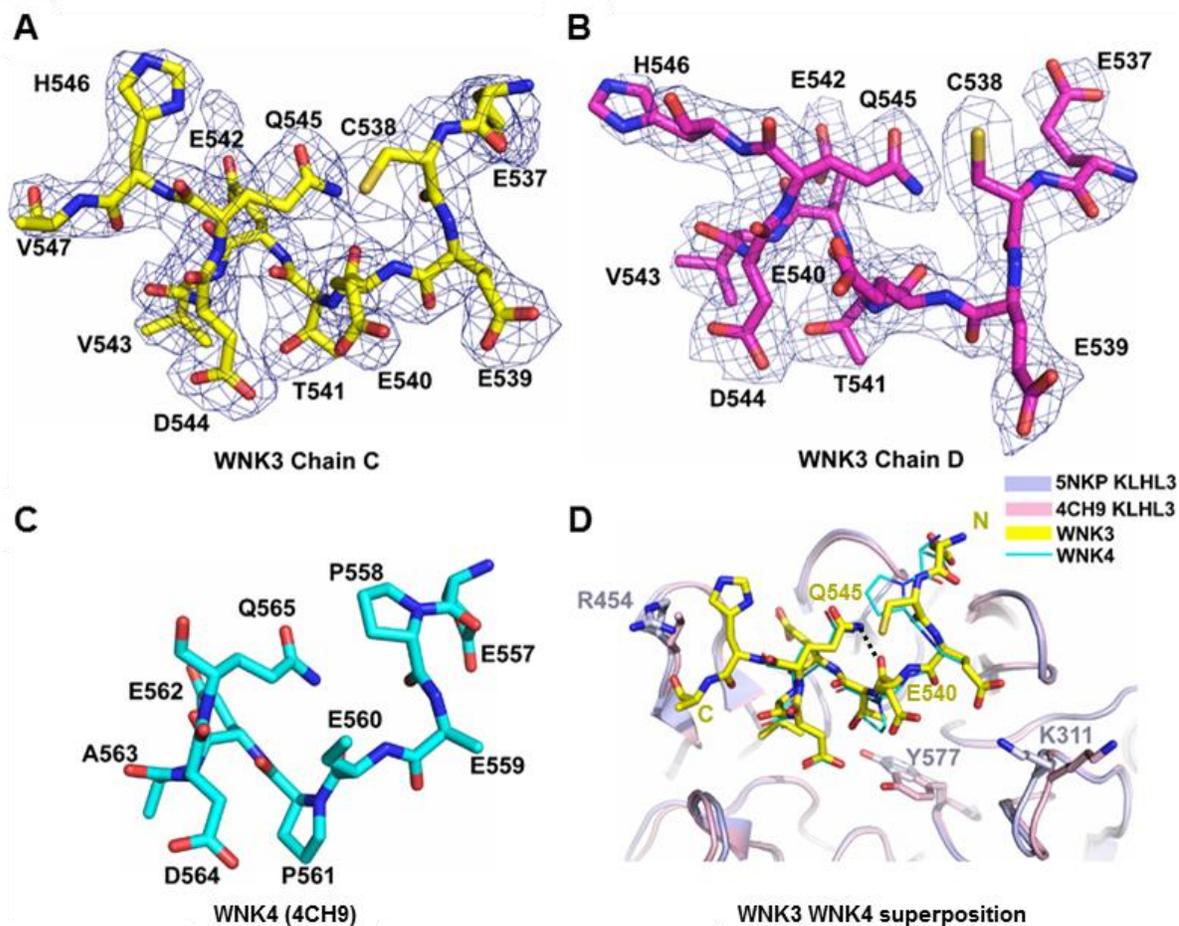


Figure 5.5 Structural comparison of WNK degnon peptides. (A)(B) Stick representation and 2Fo-Fc electron density map contoured at 1.0 σ of 5NKP chain C and D respectively. WNK3 residues and their numbers are labelled. (C) Stick representation of WNK4 degnon peptide in 4CH9. Residues are labelled as shown. (D) Superposition of 5NKP and 4CH9. The WNK3 and WNK4 are aligned. Substrate binding pockets of KLHL3 in both 5NKP and 4CH9 are illustrated in ribbon representation. KLHL3 residues that adopt different conformations are highlighted with stick representation. Intramolecular hydrogen bond between WNK3 Q545 and E540 is labelled with a dashed line.

5.2.3. Structural comparison of WNK3 and WNK4 binding to KLHL3

Of note for structural comparisons, the WNK3 and previous WNK4 co-structures were both determined under acidic conditions (pH 5.1 for WNK3 and pH 4.3 for WNK4), providing similar environments for electrostatic interactions. Overall, the bound peptides adopt a conserved conformation, containing an extended N-terminal segment that spans the two proline positions

in WNK4, and a C-terminal segment that folds into a single helical turn (Figures 5.4 and 5.5). The conserved glutamine in the WNK family degron motif acts to stabilize this turn through an intramolecular hydrogen bond, which in WNK3 is mediated between the side chain of Gln545 and the backbone carbonyl of Glu540 (Figure 5.5D).

Key differences in the two structures include the substitutions of WNK3 Cys538 and Thr541 at the two proline positions of WNK4. While the proline phi-psi angles are near ideal for extended structure (near -60° and $+135^\circ$, respectively), the substitutions in WNK3 are associated with a subtle shift (up to 1 Å) in the peptide backbone position. This shift appears favourable for the binding interface to accommodate the bulky WNK3 Val543 substitution for WNK4 A563, as well as the branched Thr541 side chain for WNK4 Pro561 (Figure 5.5D).

A number of polar and hydrophobic interactions in the KLHL3-WNK3 structure are strongly conserved with the equivalent WNK4 complex (Figure 5.6A). Perhaps the most important interaction is the salt bridge formed between WNK3 Asp544 and KLHL3 Arg528 (Figure 5.6D). The substitution R528H in KLHL3 is the most frequent mutation in Gordon's hypertension syndrome and is sufficient to abolish KLHL3 function [8, 9]. WNK3 Val543 is a conservative change from WNK4 Ala563 and packs adjacent to this salt bridge to form equivalent van der Waals interactions with KLHL3 Y449 and H498. The preceding WNK3 residue Glu542 is also buried within the KLHL3 surface between KLHL3 Y449 and F402, where it can hydrogen bond to KLHL3 Ser432 (Figure 5.6A). Finally, the backbone carbonyls of WNK3 Cys538 and Glu539 form conserved hydrogen bond interactions between KLHL3 side chains Arg360 and Arg339, respectively (Figure 5.6A).

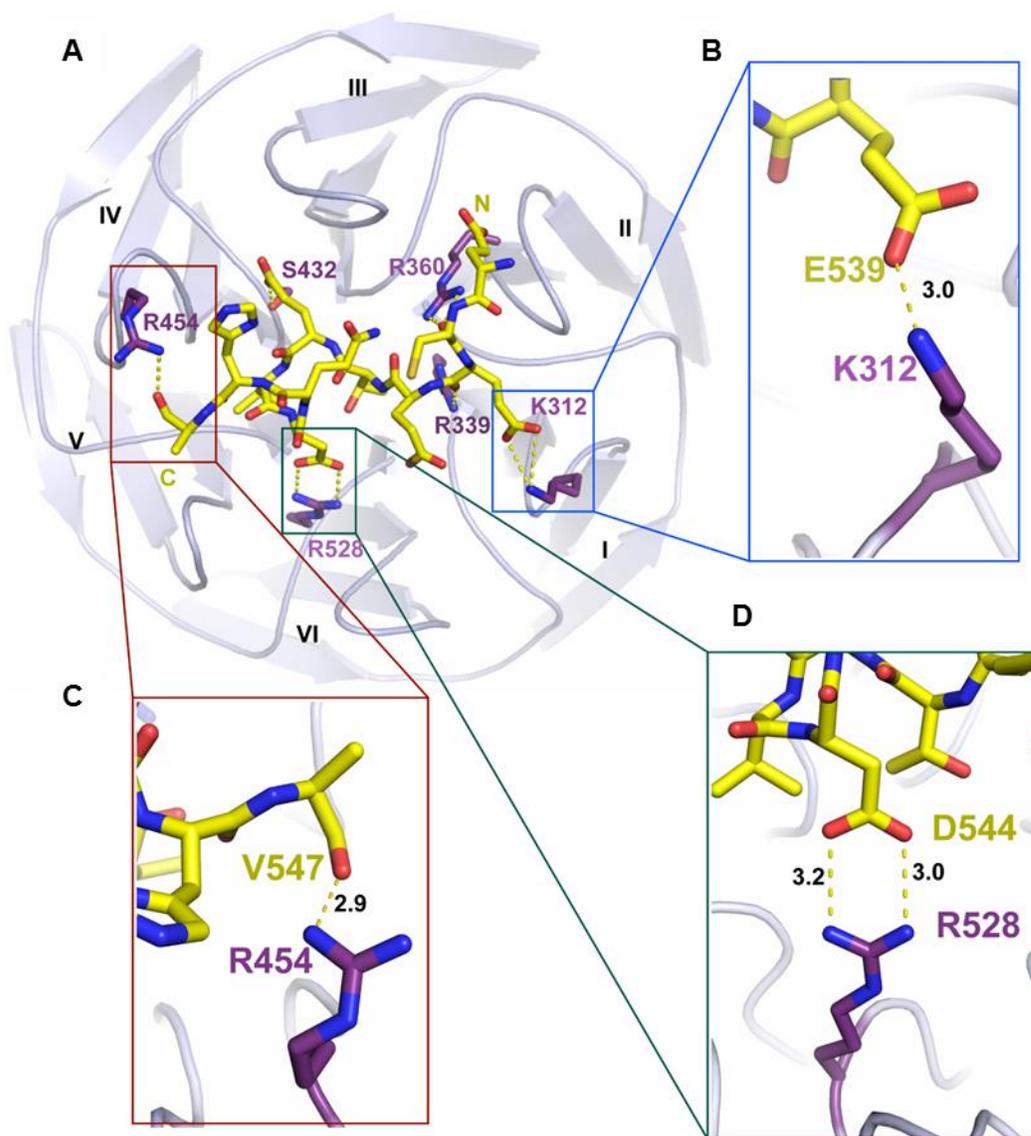


Figure 5.6 Interactions in KLHL3-WNK3 interface. (A) Overview of the polar contacts in the complex interface. KLHL3 residues are labelled in purple and WNK3 peptide is shown in yellow. Hydrogen bonds are demonstrated using dashed lines. (B-D) Close-up views of the featured polar contacts. KLHL3 residues are labelled in purple and WNK3 residues are labelled in yellow. The distances between atoms are measured in pymol (\AA).

5.2.4. Newly observed interactions in the WNK3 binding interface

The new structure reveals several features of the WNK family degron interaction that were not previously defined. First, the full 11-mer degron motif can be traced in the WNK3 structure, whereas the two C-terminal residues were not observed in the WNK4 complex. Surprisingly,

truncation of either of these residues was not tolerated in WNK4 showing their importance for the KLHL3 interaction [8]. The newly defined C-terminal residues, WNK3 His546 and Val547, appear flexible and able to pack on the surface of the Kelch domain, forming van der Waals and electrostatic interactions with KLHL3 Tyr449 and Arg454, including a hydrogen bond between the backbone carbonyl of WNK3 Val547 (chain C) and KLHL3 Arg454 (Figure 5.6C). They also contribute to the peptide's helical turn conformation through the backbone interactions of His546. Second, the new structure reveals a salt bridge between the conserved WNK3 residue Glu539 and KLHL3 Lys312 (Figure 5.6B). The side chain atoms of this glutamate residue appeared disordered in the WNK4 structure (Figure 5.5C), possibly resulting from its solvent exposed position and the additional flexibility of the associated lysine.

5.2.5. Steric constraints would disfavour phosphothreonine in the WNK3 interface

WNK3 Thr541 adopts a central buried position in the complex interface and would appear to offer a unique potential phosphorylation site for regulation among the WNK family degron motifs (Figures 5.7A and B). The threonine side chain forms van der Waals interactions with several KLHL3 residues, including Arg339 and Tyr577 (Figures 5.7C and 5.7D). To understand how post-translational modification on Thr541 might affect WNK3 binding to KLHL3, I modelled a phosphoryl group onto the threonine side chain using the ICM-Pro software package [10]. When modelled in its crystallized conformation a severe steric clash was observed with WNK3 Asp544 that would break the critical salt bridge between this aspartate and KLHL3 Arg528 (Figure 5.7C). I therefore performed an energy minimization step to explore other potential side chain conformations. The resulting alternative conformation positioned the phosphoryl group of pThr541 between Arg339 and Tyr577 in a sterically crowded environment that would still bury the negative charge unfavourably without satisfying the required hydrogen bonding (Figure 5.7D).

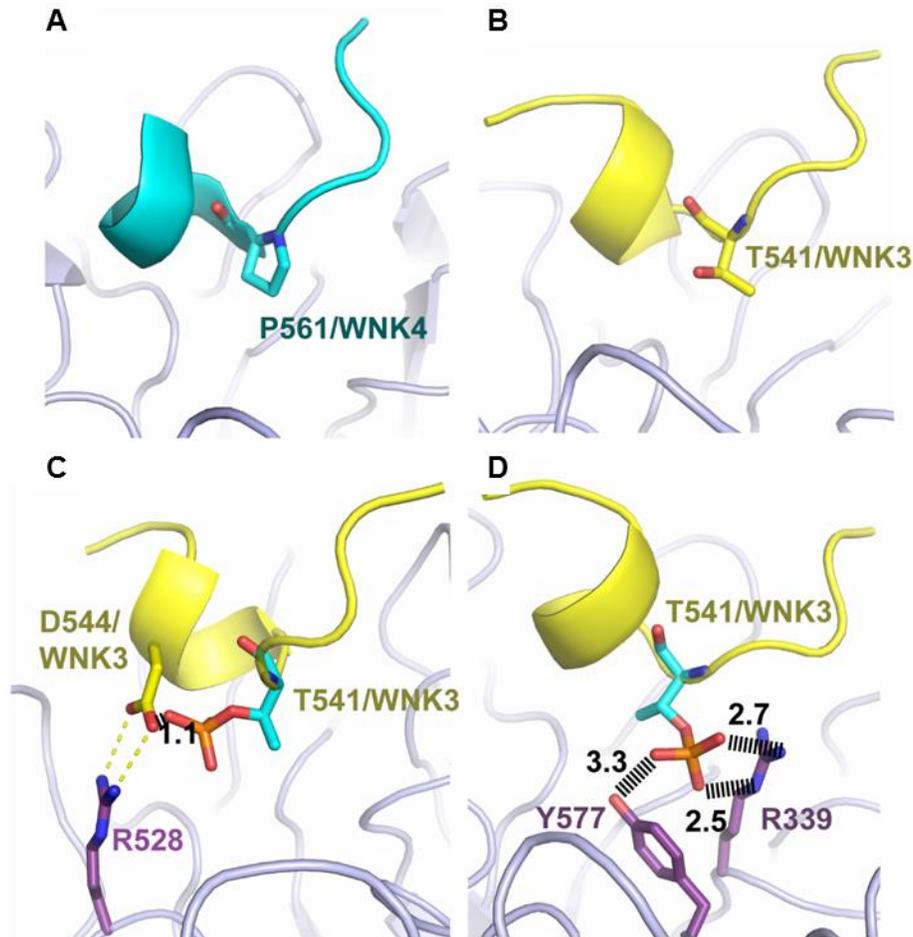
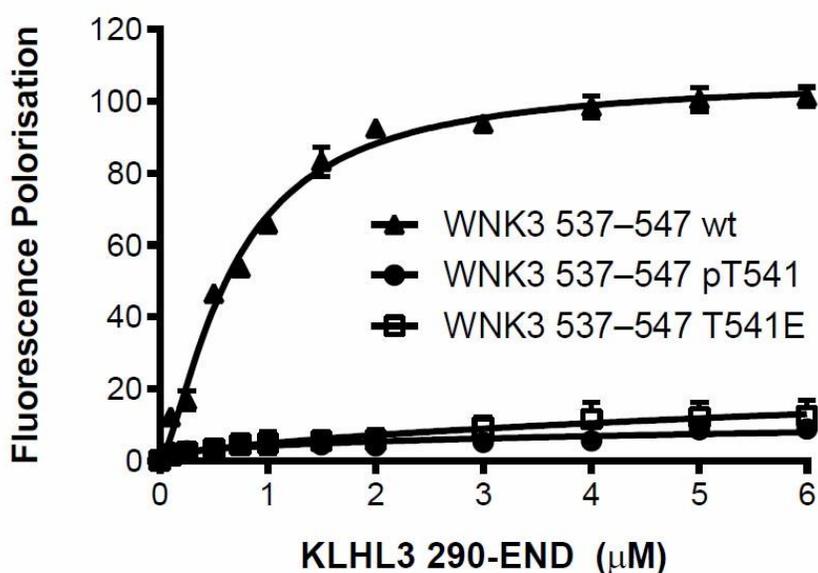


Figure 5.7 Steric constraints would disfavour phosphothreonine in the WNK3 interface. (A) Structure of KLHL3-WNK4 complex (4CH9) featuring WNK4 Pro561. KLHL3 is shown in light purple and WNK4 is shown in blue. (B) Structure of KLHL3-WNK3 complex (5NKP) featuring WNK3 Thr541. WNK3 is shown in yellow. (C)(D) Two conformations of phosphorylation modelling at WNK3 Thr541. Residues clashed with the phosphothreonine are labelled. Distance is measured in pymol (Å)

To investigate the binding of WNK3 variants to the KLHL3 Kelch domain, our collaborators Dr Gerrit M. Daubner and Dr Dario R. Alessi from University of Dundee employed a fluorescence polarisation assay developed previously to map the WNK4 degron motif (⁵⁵⁷EPEEPEADQHQ) [8]. They observed that the 11-mer WNK3 degron motif (ECEETEVDQHV) bound to KLHL3 with $K_D = 0.67 \mu\text{M}$ (Figure 5.8). Strikingly, the binding of an equivalent WNK3 peptide carrying a single phosphorylation at Thr541 was severely

destabilized to measure a dissociation constant ($K_D > 50 \mu\text{M}$, Figure 5.8). A similar loss of KLHL3 interaction was observed using a WNK3 T541E mutant degron (Figure 5.8). Together, these results indicated that WNK3 binding was abrogated upon modification of WNK3 Thr541. Thus, both the fluorescence polarisation assay and structure indicate that Thr541 phosphorylation would disrupt the WNK3-KLHL3 interaction and therefore serve as a potential regulatory mechanism.



	WNK3 537-547 wt	WNK3 537-547 pT541	WNK3 537-547 T541E
Bmax	106.5	~ 3253	~ 240.4
h	1.435	~ 0.3596	0.5707
Kd	0.6689	~ 1.106e+008	~ 899.0

Figure 5.8 FP assay showing the WNK3 acidic degron motif binds potently to KLHL3 but does not bind if phosphorylated at Thr541. Peptides containing the WNK family degron motif were conjugated to a Lumio green fluorophore and titrated with recombinant KLHL3 Kelch domain. Fluorescence polarisation measurements were recorded using a BMG PheraStar plate reader, with an excitation wavelength of 485 nm and an emission wavelength of 538 nm, and measurements were corrected to the fluorescent probe alone. One Site Specific binding with Hill Slope was assumed (model $Y=B_{max} \cdot X^h / (K_d^h + X^h)$) and the disassociation constant, and associated standard error was obtained. The mean values of data measured from three replicates were plotted in the graph. The error bars represent the standard deviation. The experiment is performed by Dr Gerrit M. Daubner.

5.2.6. Phosphorylation mapping of WNK3 under different stimuli.

The WNK3 degron motif most closely matches the preferred substrate sites of acidophilic kinases such as casein kinase II (CK2) [11]. However, to date there have been no reports of phosphorylation on WNK Thr541 either in the literature or in public databases. The majority of proteomic analyses use trypsin digestion to restrict cleavage sites to basic arginine and lysine positions, which simplifies database searches for peptide matches.

WNK3 Thr541 lies in a particularly poor sequence region for this approach as trypsin digestion would yield a peptide 36 amino acids in length due to the paucity of surrounding basic residues. We reasoned therefore that a potential phosphorylation at this site may have been missed previously and decided to perform a new proteomic search using elastase cleavage, which targets the more common positions of small hydrophobic residues [12]. For these experiments, I transfected HEK293T cells with Flag-tagged WNK3 plasmid and after 36 hours treated the cells for 30 minutes with either isotonic or hypotonic buffer to allow for phosphorylation under different stimuli (Buffer compositions in 2.10.2.). Samples enriched for WNK3 were then prepared by anti-Flag immunoprecipitation, cleaved and analysed by MS/MS. Mass spectrometry analysis was performed at the Discovery Proteomics Facility (headed by Roman Fischer). Using this approach we obtained near complete coverage of the WNK3 sequence including the degron motif (Figures 5.9 and 5.10). Reassuringly, the experiment identified the known phosphorylation site at Ser308 in the kinase activation segment (Figure 5.10). However, no phosphorylation was detected at Thr541 on any of the 29 peptides recovered for this degron site. The most likely explanation is that the stoichiometric ratio of protein phosphorylated versus unphosphorylated does not favour its detection without specific phospho-peptide enrichment.

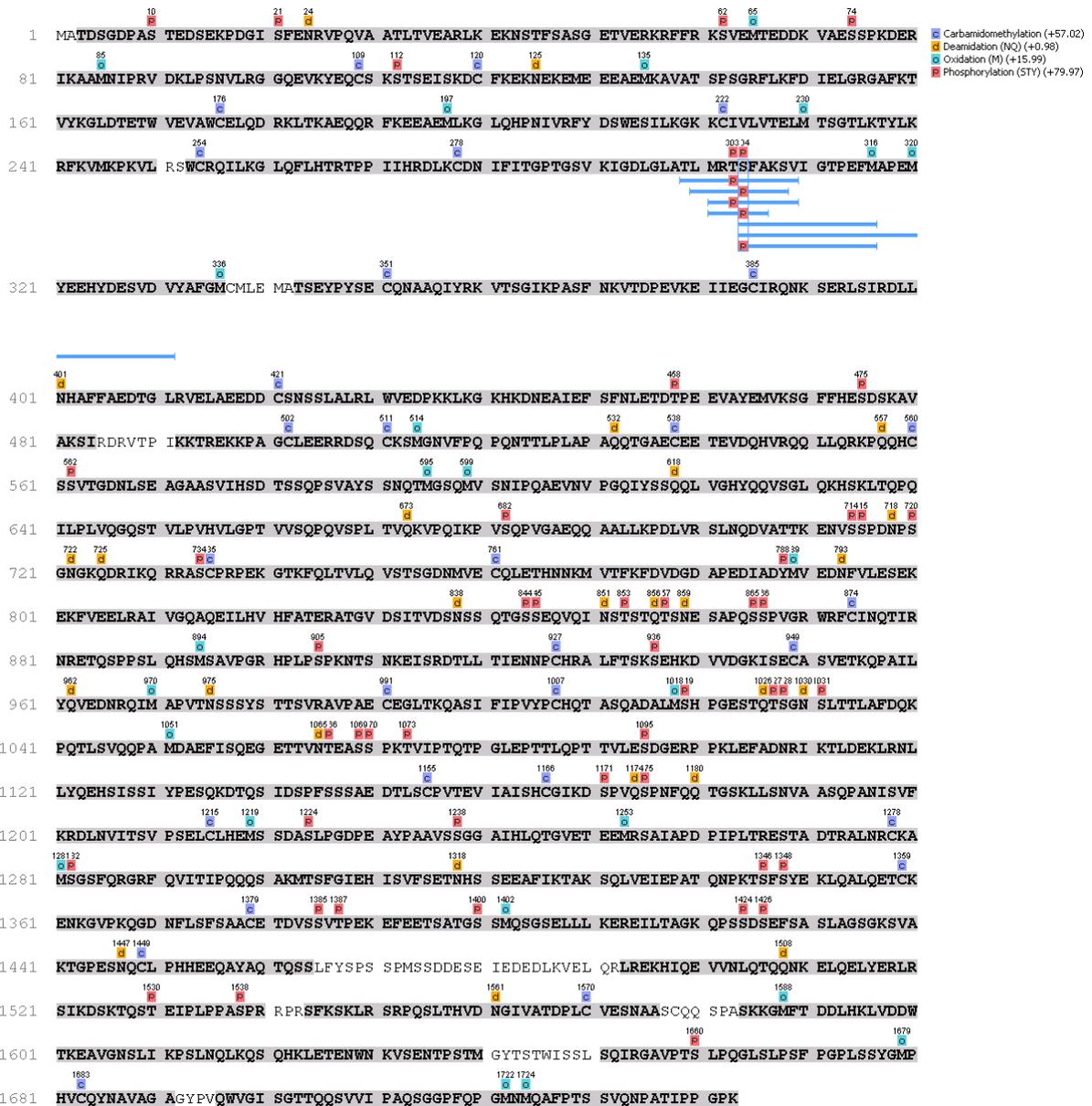


Figure 5.9 Overview of WNK3 peptide hits from hypotonic treatment. Residues recovered in elastase MS/MS are shaded in grey, showing 97% sequence coverage. Modifications are labelled above residues. Peptide hits containing Ser308 are listed.

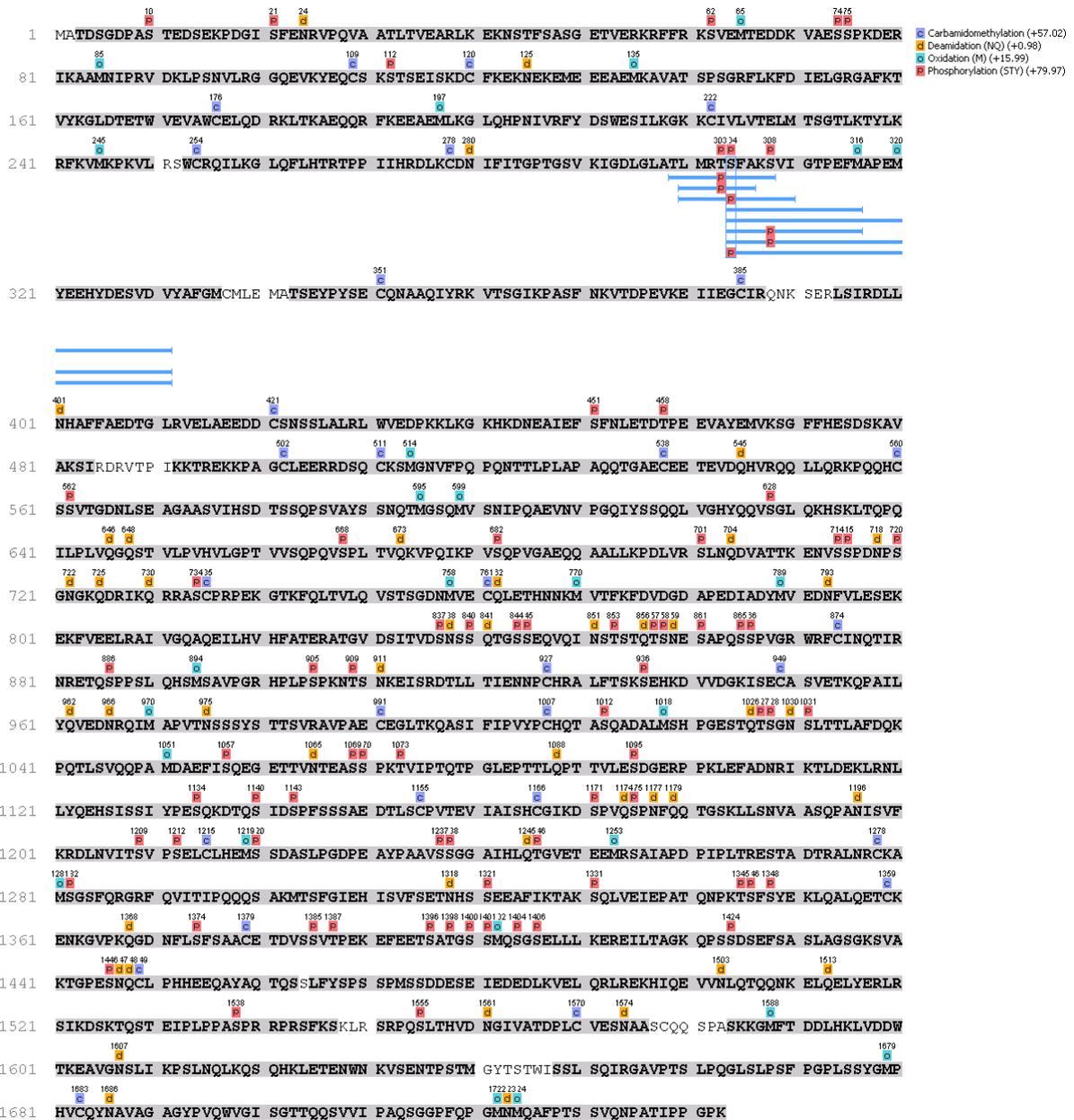


Figure 5.10 Overview of WNK3 peptide hits from isotonic treatment. Residues recovered in elastase MS/MS are shaded in grey, showing 99% sequence coverage. Modifications are labelled above residues. Peptide hits containing Ser308 are listed.

5.3. Discussion

Previously published data have shown that all WNK isoforms may be ubiquitinated by CUL3^{KLHL3} E3 ligase to regulate blood pressure [8]. Overall, the results in this chapter showed that despite four residue substitutions out of eleven, the WNK3 acidic degron motif still bound

to the KLHL3 Kelch domain with high affinity. The new structure of the KLHL3-WNK3 complex presented in this chapter (PDB 5NKP) reveals a conserved structural mechanism for WNK family interaction with KLHL3. The four amino acid substitutions within the WNK3 acidic degron motif compared to WNK4 are well tolerated with no significant changes in the bound peptide conformation. ITC measurements also reveal an enthalpy driving pattern shared by both KLHL3-WNK3 and KLHL3-WNK4 complexes. Despite the conserved features, the new structure reveals several features of the degron interaction that were not previously defined. The current model extended the understanding towards the termini of both molecules compared to the previous KLHL3-WNK4 structure.

However, the K_D values observed in the FP assay for both binding events were much higher than those observed in ITC ($K_D = 0.38 \mu\text{M}$ [8] versus $1.1 \mu\text{M}$ for WNK4 and $K_D = 0.67 \mu\text{M}$ versus $4.95 \mu\text{M}$ for WNK3) (Figures 5.2 and 5.8). Despite the difference in experimental techniques, the key factor that affected the measured affinity might be the buffer composition. In FP, the bindings were measured in a buffer containing 150mM NaCl and 2mM DTT, whereas in ITC, they were measured in a buffer containing 300mM NaCl and 1mM TCEP. It is likely that the binding of KLHL3 with WNKs is dependent on buffer ionic strength. Further single-factor experiments are required to investigate this hypothesis.

Degradation of WNKs can be impaired by Akt and PKA phosphorylation on KLHL3 S433. We were wondering if similar mechanism also occurs on WNKs side. Particularly, the WNK3 acidic degron motif contains a unique Thr which may allow phosphorylation. Both structural modelling and Fluorescence Polarisation assay indicated that phosphorylation on WNK3 Thr541 would disrupt KLHL3-WNK3 binding by stereo-hindrance effect. Taken together, phosphorylation on WNK3 Thr541 may regulation WNK3 protein level by disrupting KLHL3-WNK3 interaction under certain stimuli.

WNK3 was proposed to be the main driver of KCC3 T991 and T1048 phosphorylation under hypotonic high K^+ condition. To test if the same condition would stimulate Thr541

phosphorylation, we performed mass spectrometry phospho-mapping on elastase digested WNK3 samples with hypotonic or isotonic treatment. Unfortunately we didn't map a phosphorylated Thr541 in either sample. Further studies are required to observe whether this site is phosphorylated under physiological conditions in particular cell types.

5.4. References

1. Alessi, D.R., et al., *The WNK-SPAK/OSR1 pathway: master regulator of cation-chloride cotransporters*. *Sci Signal*, 2014. **7**(334): p. re3.
2. Thastrup, J.O., et al., *SPAK/OSR1 regulate NKCC1 and WNK activity: analysis of WNK isoform interactions and activation by T-loop trans-autophosphorylation*. *Biochem J*, 2012. **441**(1): p. 325-37.
3. Vitari, A.C., et al., *The WNK1 and WNK4 protein kinases that are mutated in Gordon's hypertension syndrome phosphorylate and activate SPAK and OSR1 protein kinases*. *Biochem J*, 2005. **391**(Pt 1): p. 17-24.
4. Richardson, C. and D.R. Alessi, *The regulation of salt transport and blood pressure by the WNK-SPAK/OSR1 signalling pathway*. *J Cell Sci*, 2008. **121**(Pt 20): p. 3293-304.
5. de Los Heros, P., et al., *The WNK-regulated SPAK/OSR1 kinases directly phosphorylate and inhibit the K⁺-Cl⁻ co-transporters*. *Biochem J*, 2014. **458**(3): p. 559-73.
6. Wilson, F.H., et al., *Human hypertension caused by mutations in WNK kinases*. *Science*, 2001. **293**(5532): p. 1107-12.
7. Shibata, S., et al., *Kelch-like 3 and Cullin 3 regulate electrolyte homeostasis via ubiquitination and degradation of WNK4*. *Proc Natl Acad Sci U S A*, 2013. **110**(19): p. 7838-43.
8. Schumacher, F.R., et al., *Structural and biochemical characterization of the KLHL3-WNK kinase interaction important in blood pressure regulation*. *Biochem J*, 2014. **460**(2): p. 237-46.
9. Ohta, A., et al., *The CUL3-KLHL3 E3 ligase complex mutated in Gordon's hypertension syndrome interacts with and ubiquitylates WNK isoforms: disease-causing mutations in KLHL3 and WNK4 disrupt interaction*. *Biochem J*, 2013. **451**(1): p. 111-22.
10. Abagyan, R., et al., *Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search*. *Proteins: Structure, Function, and Bioinformatics*, 1997. **29**(S1): p. 29-37.
11. Miller, M.L., et al., *Linear motif atlas for phosphorylation-dependent signaling*. *Science signaling*, 2008. **1**(35): p. ra2-ra2.
12. Davis, S., et al., *Expanding Proteome Coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) Combined with Broad Specificity Proteolysis*. *J Proteome Res*, 2017. **16**(3): p. 1288-1299.

Chapter 6

Identifying the novel substrates for KBTBD4, a subtype-specific cancer driver in Medulloblastoma

6.1. Introduction

The Cullin-RING E3 ligase KBTBD4 was recently reported as the ‘most compelling single-gene discovery’ in medulloblastoma [1]. Medulloblastoma (MB) is the most common childhood brain malignancy, comprising 63% of childhood intracranial embryonal tumours. Pathologically, MB is originated from aberrant cell differentiation during cerebellum development within a patient’s early life [2]. Standard therapy includes surgery, chemotherapy and cranio-spinal irradiation, resulting in 70–80% overall survival rates. However, the majority of survivors suffer from cognition defects in the long term and the prognosis is highly associated with the patients’ ages and the molecular profiles of the tumours [2].

Several genome-wide studies [3-7] defined four subgroups – WNT-MBs, SHH-MBs, Group 3 MBs and Group 4 MBs, with distinct molecular features based on gene expression profiling and DNA copy number alterations. Among the four types, Group 3-MBs have the worst outcome with high risk tumours and poor prognosis, while Group 4 MBs account for 40% of all patients. The mutations found in KBTBD4 in Group 3/4 MBs were predominantly in-frame insertions affecting the second Kelch repeat of the Kelch domain. Of note, different recurrent insertions were found in the different MB subgroups, specifically R313>PRR in Group 3 and P311>PP in Group 4, suggesting a potential gain of function. Interestingly, recurrent mutations at the same sites were discovered in Pineal Parenchymal Tumor of Intermediate Differentiation (PPTID)[8]. Similar to other BTB-Kelch family members, KBTBD4 potentially binds to Cullin3

as an E3 substrate adaptor for the ubiquitin-proteasome system. However, no characterisation of this protein has been previously performed.

A recent analysis resampled a series of 1501 medulloblastomas with DNA-methylation profiling data, and proposed eight robust subtypes within Group 3/4 MBs [9]. Under this refined classification, KBTBD4 is regarded as the subtype cancer driver for subtype II (13%) and subtype VII (22%). Besides KBTBD4, a number of chromatin modifying genes are mutated in other subtypes, including PRDM6, KDM6A, ZMYM3 and KMT2C, implying a histone-related pathological mechanism [10].

This chapter aims to understand the tumorigenic mechanism of KBTBD4 in medulloblastoma and its potential function as an E3 substrate adaptor. I hypothesised that the recurrent insertions in KBTBD4 may promote recruitment of neosubstrates due to structural changes in the substrate recognition site. I therefore employed a mass spectrometry-based proteomics approach to compare the proteins associating with wild-type and mutant KBTBD4 forms. In agreement with the chromatin modifying genes described above [10], this work identified components of the CoREST-LSD1 complex as specific binders for mutant KBTBD4. Previous studies showed CoREST-LSD1 complexes as ontogenetic chromatin modifiers suppressing cell differentiation and promoting proliferation, consistent with a potential oncogenic role in KBTBD4-mutant medulloblastoma [11-19]. Thus, in this chapter, I performed further cellular experiments to validate the proteomic data suggesting CoREST-LSD1 complexes as neosubstrates for KBTBD4-dependent ubiquitination, which would confer an alternative mechanism for an altered epigenetic landscape in medulloblastoma.

6.2. Results

6.2.1. Mass Spectrometry Proteomics identified CoREST-LSD1 complex as specific binders for KBTBD4 MB mutants

6.2.1.1. Sample preparation for mass spectrometry

The study by Northcott et al in 2017 reported R313delinsPRR and P311delinsPP as the most recurrent KBTBD4 mutations in Medulloblastoma Group 3 and 4, respectively [20]. I therefore selected these alterations to represent MB Group 3 and Group 4 KBTBD4 mutants in a proteomics study (Figure 6.1A). Unfortunately, KBTBD4 mutants were not reported in any known MB patient derived cell lines that have been established in culture (*data not shown*). For proteomic study, KBTBD4 wild type (WT) and two mutants were therefore cloned into a 3xFlag pcDNA vector and then individually transfected into HEK293T cells alongside an empty vector (EV) control. After culturing for 40 hours, KBTBD4 and its binders were purified by Flag immunoprecipitation and then co-eluted with 3xFlag peptide, which recovered about 50% of total KBTBD4 expressed. (Figure 6.1B and C). I proceeded to perform a tryptic digest of the elution samples before providing the samples to Dr Iolanda Vendral (Oncology Department) who performed the MSMS experiments at the proteomics facility within the Target Discovery Institute.

3 and Group 4 mutant samples. After subtracting hits that appeared in the EV set, 237 hits were shared in all three KBTBD4 hit sets, including CUL3 and RBX1, the fundamental components of Cullin3 E3 complexes. In terms of WT and mutant comparison, 237 hits were identified only in KBTBD4 mutant sets, with 65 hits shared by both Group 3 and Group 4 sets, 108 in Group 3 only and 64 in Group 4 only. On the contrary, only 57 hits were unique to KBTBD4 WT set (Figure 6.2A). The distribution of the hits implies KBTBD4 mutants' gain of function whilst employing the similar Cullin3 E3 machinery as the WT. However most of the hits are of low peptide counts and exponentially modified protein abundance index (emPAI) values, suggesting low confidence in being direct binders. Therefore the numbers listed above need to be viewed with caution.

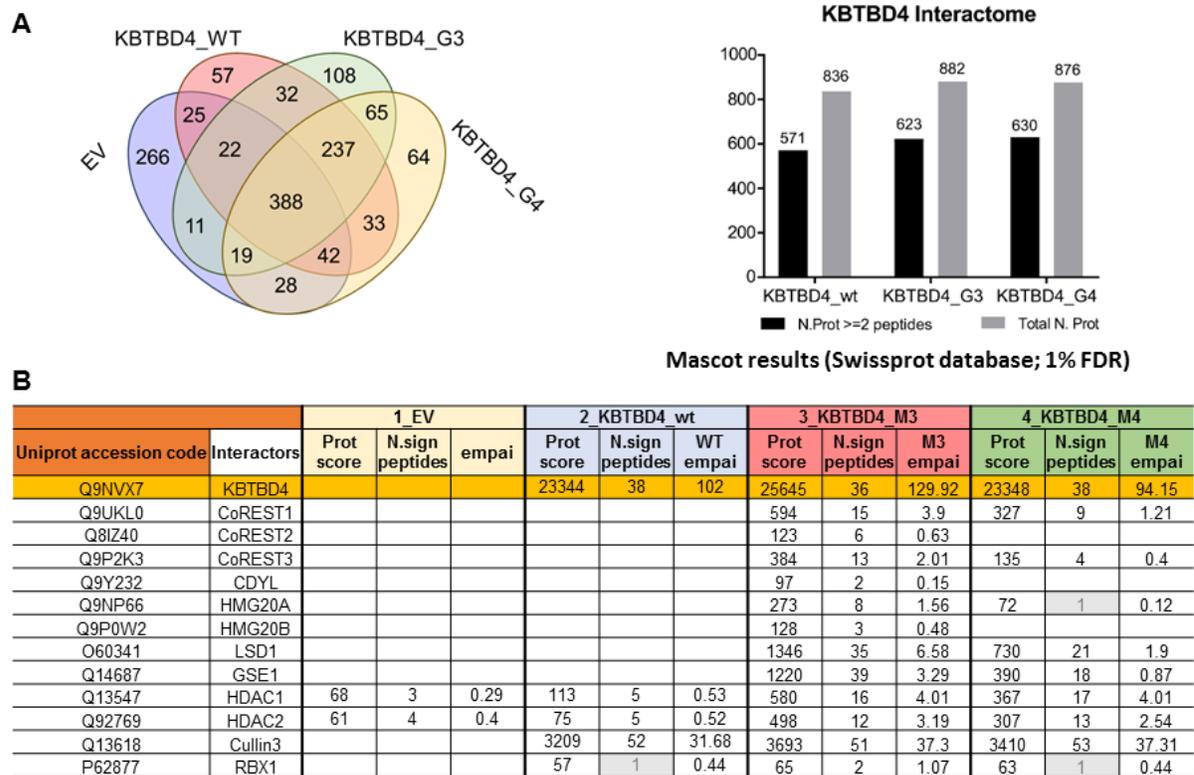


Figure 6.2 Mass Spectrometry Proteomics identified the CoREST-LSD1 complex as unique binders for KBTBD4 MB mutants (A) Venn diagram and bar chart figure panels showing an overview of the statistics of the proteomics data. (B) Table of the top high confidence interactors identified from mass spectrometry proteomics with respect to mutant KBTBD4. The protein scores suggest the confidence of the identification and emPAI values indicate the protein abundance.

Amongst the hits that were unique to KBTBD4 mutants, CoREST complexes were of outstanding significance due to their high abundance amongst the peptides captured by the KBTBD4 mutant baits (Figure 6.2B). In the CoREST complexes, a CoREST subunit (CoREST1-3, also known as RCOR1-3) functions as a scaffolding protein for enzymes that coordinate sequential deacetylation and demethylation events after being recruited to Histone H3 by zinc finger proteins. CoREST consists of an ELM2 and two SANT domains (Figure 6.3A). The ELM2-SANT1 region is required to recruit deacetylase HDAC1/2. The linker between the two SANT domains forms a long stalk to connect the catalytic domain of demethylase LSD1 and the CoREST SANT2 domain (Figure 6.3B). The SANT2 domain binds to DNA, which helps to determine the specificity of CoREST complex recruitment to nucleosomal substrates [21].

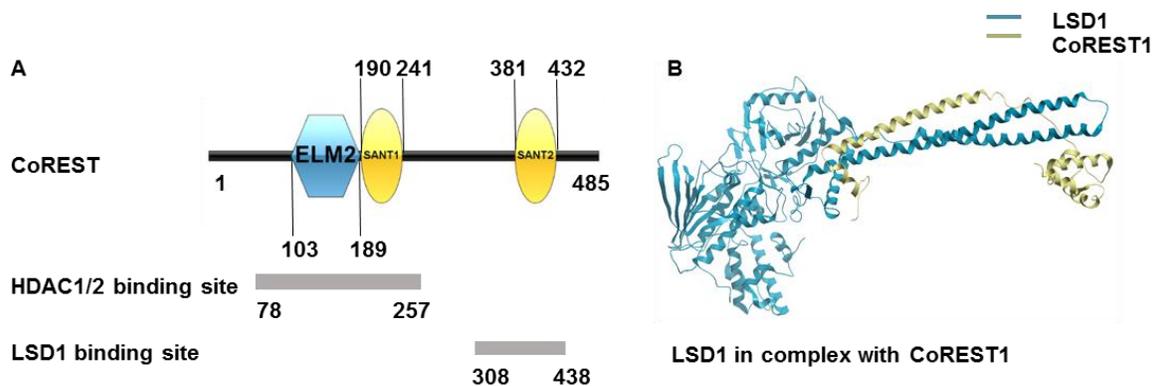


Figure 6.3 CoREST domain organisation and structure. (A) CoREST consists of ELM2 and two SANT domains. The N-terminal ELM2 and SANT1 domains are required for HDAC1/2 binding. The linker region and C-terminal SANT2 domain bind to LSD1. Residues are numbered for CoREST1 (B) Crystal structure of CoREST1 C terminal region (308-440 a.a.) in complex with LSD1. PDB: 4UV8.

The complete set of core components in the CoREST complexes were co-purified with the KBTBD4 Group 3 mutant, including CoREST1/2/3, HMG20B (also known as BRAF35, a zinc finger protein) and its paralog HMG20A, HDAC1/2 and LSD1 (Figure 6.2B). CoREST complex core components were also co-eluted with the KBTBD4 Group 4 mutant, but with slightly

reduced efficiency as evidenced by the lower abundance of recovered target peptides as well as reduced completeness of recovered targets. In addition, GSE1, a gene suppressor that is known to bind CoREST3, and CDYL, a Histone 3 methylation reader that binds to HDAC1/2, also appeared co-eluted with KBTBD4 mutants. Taken together, these results indicate that KBTBD4 mutants in medulloblastoma potentially promote tumorigenic chromatin remodelling by targeting CoREST complexes.

6.2.2. Validation of interaction partners in HEK293T cells by co-immunoprecipitation and Western blot.

6.2.2.1. CoREST1 and LSD1 bind to KBTBD4 mutants but not WT.

To validate the interactions identified by mass spectrometry proteomics, HEK293T cells were treated as indicated in Figure 6.4 after transfection of wild type and mutant KBTBD4 variants. Flag immunoprecipitation was performed to recover the KBTBD4, and its interaction partners which were selected by available antibodies. Endogenous CoREST1 was co-immunoprecipitated with transfected KBTBD4 Group 3 mutant but not with WT. Similarly, endogenous LSD1 was pulled down by KBTBD4 Group3 and Group 4 mutants, but not by WT. Interestingly, MLN4924, a neddylation inhibitor which was expected to stabilise degraded substrates, was not able to stabilise either CoREST1 or LSD1

6.2.2.2. Two isoforms of CoREST3 bind to KBTBD4 mutants but not WT.

In agreement with the mass spectrometry data, endogenous CoREST3 was pulled down by KBTBD4 Group3 and 4 mutants but not by WT in the validation immunoprecipitation. However, two isoforms were recognized by the CoREST3 antibody and they exhibited different efficiency of recovery to KBTBD4 mutants (Figure 6.4C). To investigate the two isoforms, I cloned CoREST3 by Reverse Transcription PCR with RNA extracted from HEK293T cells. Sequencing of the clones suggested a longer isoform 3 and a shorter isoform 2 expressed in HEK293T

cells. CoREST3 isoform 3 (Uniprot Q9P2K3-3) follows the most recognized domain organisation pattern of CoRESTs, while the isoform 2 (Uniprot Q9P2K3-2) has a frameshift mutation compared to the canonical sequence, which results in loss of the SANT2 domain (Figure 6.4D).

As is shown in Figure 6.4C, CoREST3 isoform 3 exhibited a far more efficient recovery than isoform2, suggesting that the loss of SANT2 domain reduces its affinity to KBTBD4 mutants. Comparatively, the effect was more profound with KBTBD4 Group 3 than with the KBTBD4 Group 4 mutant.

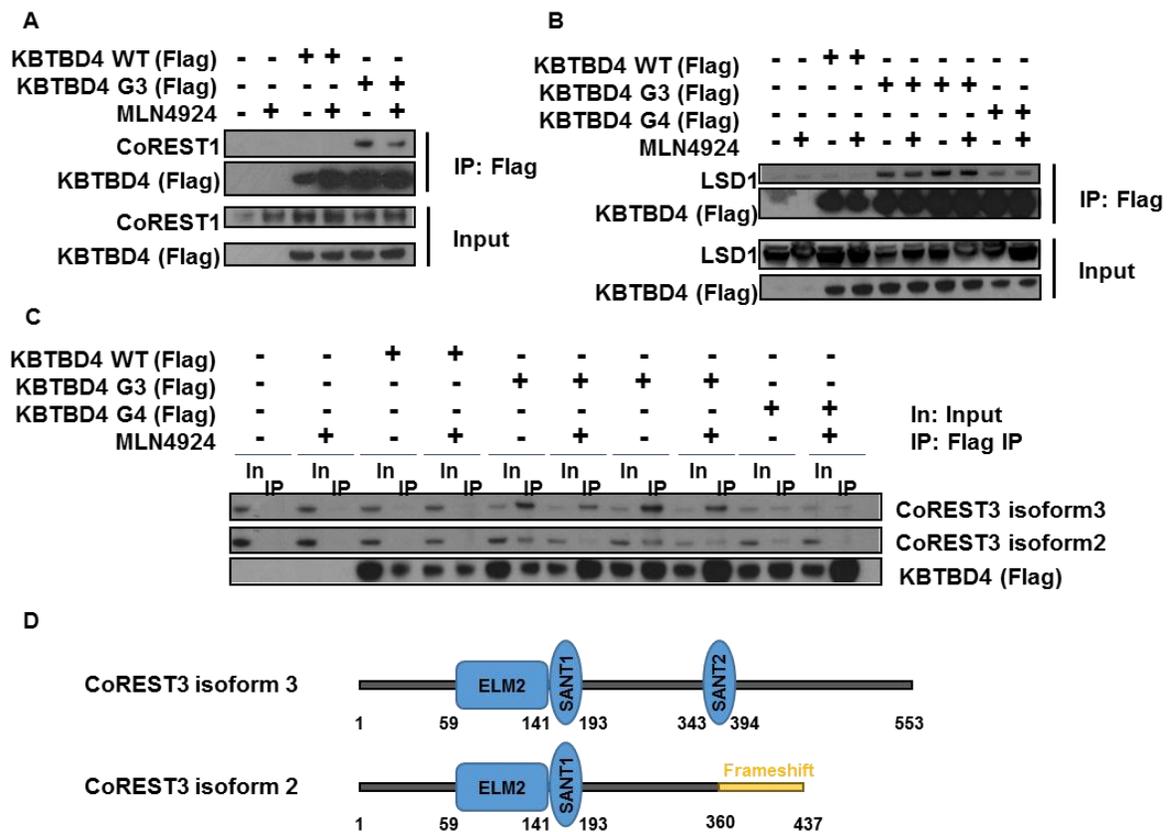


Figure 6.4 Interaction validation via co-immunoprecipitation in HEK293T cells. (A) CoREST1 was co-precipitated with KBTBD4 Group 3 mutant but not with WT. (B) LSD1 was co-precipitated with KBTBD4 Group 3 and Group 4 mutants, while only recovered to background level with KBTBD4 WT. For A-C, MLN4924 treatment didn't stabilise CoREST-LSD complex. (C) Two isoforms of CoREST3 were co-precipitated with KBTBD4 Group 3 and 4 mutants but not with WT. (D) Schematic diagrams of CoREST3 isoform 3 and 2.

6.2.3. CoREST3 is ubiquitinated and destabilised by KBTBD4 Group 3 and 4 mutants.

6.2.3.1. Endogenous CoREST3 are destabilised by KBTBD4 Group 3 and 4 mutants.

To understand the effect of KBTBD4 mutants on the CoREST complex, I examined the changes of endogenous CoREST1 and 3 protein levels upon KBTBD4 transfection in the cytoplasm and nucleus separately in HEK293T cells (Figure 6.5). In the cytosolic fraction, CoREST3 isoform 3 was drastically reduced when KBTBD4 mutants were expressed, while CoREST3 isoform 2 level remained unchanged. In the nucleus, however, both CoREST3 isoforms were destabilised by KBTBD4 mutants. Consistent with the findings in section 6.2.2, MLN4924 failed to stabilise CoREST3.

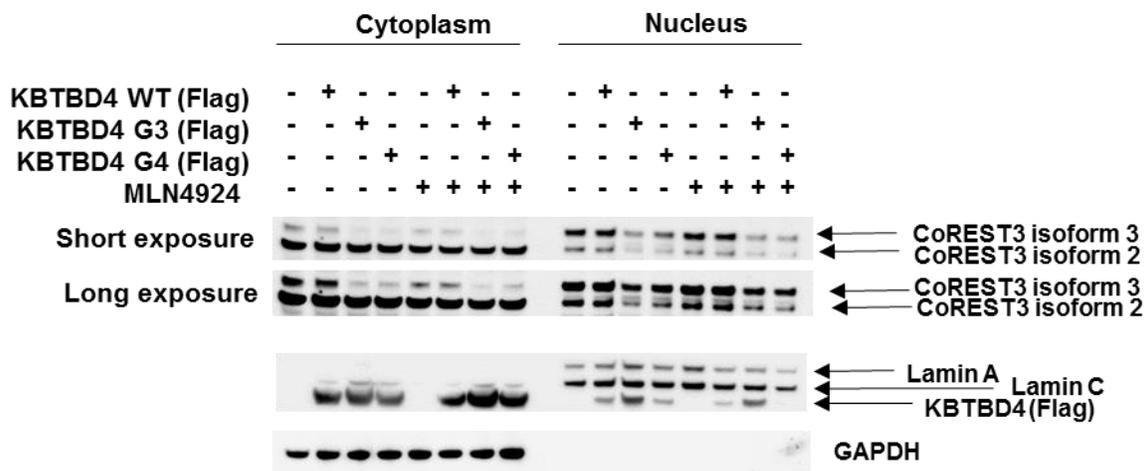


Figure 6.5 CoREST3 is destabilised by KBTBD4 mutants. KBTBD4 variants were transfected into HEK293T cells. Cells were treated with MLN4924 as indicated. Endogenous CoREST3 levels in the cytoplasm and nucleus were analysed by Western blot. GAPDH and LaminA/C were also blotted as loading control for cytoplasm and nucleus correspondingly. The experiment was performed once.

6.2.3.2. In vitro ubiquitination assay for CoREST3 as a substrate.

To investigate whether CoREST3 is a neo-substrate of KBTBD4 mutants for ubiquitination, I performed an in vitro ubiquitination assay. CoREST3 isoform 3 was constructed into pcDNA3.1(-) vector with an N-term HA tag. In the in vitro ubiquitination system, Myc-ubiquitin, Flag-KBTBD4 and HA-CoREST3 were co-expressed in HEK293T cells. Cells were lysed under denaturing condition (see section 2.12) to preserve the attached ubiquitin chains. CoREST3 was immunoprecipitated using HA-agarose beads and the ubiquitinated CoREST3 was detected by immunoblotting with anti-Myc antibody. As shown in Figure 6.6, the ubiquitination levels of CoREST3 remained unchanged in the absence or presence of KBTBD4 WT. In contrast, the ubiquitination levels were significantly increased in the presence of KBTBD4 mutants. A lower level of CoREST3 ubiquitination was observed with KBTBD4 Group 4 mutant than with Group 3 mutant, which is consistent with the reduced binding of CoREST3 to KBTBD4 Group 4 mutant observed in Figure 6.4C. Of note, treatment of the cells with the proteasome inhibitor MG132 did not enrich the ubiquitination on CoREST3 unexpectedly.

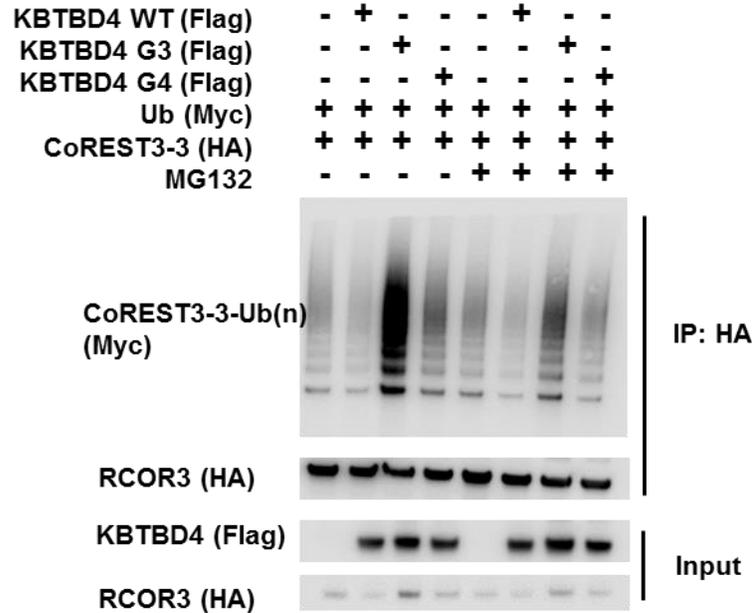


Figure 6.6 KBTBD4 mutants ubiquitinate CoREST3 isoform 3. Flag-KBTBD4, HA-CoREST3 isoform 3 and Myc-Ubiquitin were transfected into HEK293T cells as indicated. Cells were treated with 5 μ M MG132 as indicated. HA immunoprecipitation was performed. The ubiquitinated CoREST3 was detected by immunoblotting with anti-Myc antibody. Compared to the WT, KBTBD4 mutants, especially Group 3 mutant, exhibited a drastically stronger ubiquitination activity on CoREST3. The experiment was performed once.

6.3. Discussion

KBTBD4 is one of the least studied members of the BTB-Kelch family. In this chapter, the mass spectrometry proteomics study identified a large number of putative interactors of WT and mutant KBTBD4, which provides a comprehensive dataset to help further define the physiological roles of KBTBD4 in future. Excitingly, CoREST complexes were shown to exclusively bind to KBTBD4 medulloblastoma Group 3 and 4 mutants. Follow-up assays in HEK293T cells confirmed that KBTBD4 mutants recruit CoREST3 for ubiquitination and degradation. Of note, MLN4924 and MG132 treatment did not enrich the protein level and the ubiquitination level of CoREST3, respectively. These unexpected phenomena may be explained by the dominant negative effect induced by KBTBD4 overexpression. For example,

when overexpressed, KBTBD4 may fail to form a ternary complex with Cullin3 and the substrate, but instead, form a binary complex with Cullin3 or the substrate. Therefore, in this system, the interaction of KBTBD4 with CoREST3 can be detected whereas the E3 enzymatic activity of KBTBD4 is impeded, causing an unexpected effect of inhibitors as I observed. Similar issue of this dominant negative effect was observed with a related Cullin5 SOCS2 E3 ligase [22]. Overexpression or knockout of SOCS2 causes a similar gigantism phenotype [23, 24].

These initial results support our hypothesis that the mutations in KBTBD4 Kelch domain cause neo-substrate recruitment in tumorigenic events. To further test this, a disease model has to be established in medulloblastoma cell lines. Medulloblastoma is well known for its diverse genetic background. WNT-MBs are characterised by accumulation of stabilised β -catenin in the nucleus, which constitutively activates the WNT pathway. SHH-MBs are the best characterised genetically, with the majority of mutations occurring in the critical genes of the SHH signalling pathway including *PTCH1*, *SUFU*, *SMO*, *GLI1*, *GLI2* and *MYCN*. In Group 3 and 4 MBs where KBTBD4 mutations were identified, somatic gene-level mutations are less common. Nevertheless, structural variants associated with enhancer hijacking lead to profound upregulation of *GFI1* or *GFI1B* [25]. In addition, Group 3 MBs usually exhibit a high level of *MYC* amplification and genomic instability. In order to understand the disease mechanism in medulloblastoma Group 3 and 4 specific contexts, stable cell lines expressing KBTBD4 mutants are being generated in D283 and DAOY cell lines

My data suggest that KBTBD4 mutants may be involved in chromatin remodelling by targeting CoREST-LSD1 complexes. LSD1 complexes are recruited to histones by directly binding to zinc finger proteins including *GFI1/1B*, *REST*, *ZNF217*, *INSM1*, *BRAF35*, *ZNF750*, *ZNF516* etc [26-32]. Although LSD1 and CoRESTs are rarely mutated in medulloblastoma, the associated zinc finger proteins are frequently observed to be altered. Upregulation of *GFI1/1B*, for example, is observed in 15-20% of Group3 MBs and 5-10% of Group4 MBs. *REST* is

upregulated in SHH medulloblastoma tumours and is regarded as a prognostic factor [33, 34]. GSE1, a co-factor of the CoREST-LSD1 complex, is also frequently frameshift-mutated in SHH MBs. Therefore, the catalytic domains of LSD1 and HDAC1/2 have been considered potential therapeutic targets for medulloblastoma. Several studies have reported that targeting LSD1 and HDAC1/2 with small molecule inhibitors could block MB tumour migration or cell proliferation [13, 35]. However, LSD1 modifies the chromatin at different sites through distinct complexes in various cell types, suggesting complexity in regulation of LSD1 activity. Furthermore, the effects of inhibiting LSD1 remain context dependent [36]. This needs to be taken into consideration in the future work on KBTBD4 and CoREST-LSD1 complex.

This chapter mostly studied CoREST3 in light of reagent accessibility. More reagents have since been generated for CoREST1 and 2 to extensively investigate all three CoREST paralogs in parallel. Although the three CoRESTs possess ~80% sequence identity, they show differential properties in the HDAC-CoREST-LSD1 complex. CoREST1 was first identified as a co-repressor for REST/NRSF, which represses neuron-specific gene expression in non-neuronal cells [37]. Nevertheless, mice lacking CoREST1/2 die perinatally mainly due to abnormally high numbers of neural progenitors at the expense of differentiated neurons and oligodendrocyte precursor cells [28]. In an independent mouse model study, depletion of CoREST2 results in reduced Neural Progenitor Cell (NPC) proliferation, neuron population, neocortex thickness and brain size [38]. Unlike CoREST1 and CoREST2, CoREST3 leads to a reduced LSD1/KDM1A catalytic efficiency [39]. Consistent with the molecular properties, endogenous CoREST1/2 promotes cell differentiation via LSD1 demethylation activity, whereas CoREST3 serves as a natural inhibitor of LSD1 and opposes these processes.[40] In summary, the three CoRESTs facilitate a dual mechanism in regulating LSD1 enzymatic activity and maintaining the homeostasis during cell differentiation. In my future work I aim to reveal how KBTBD4 is involved in this process through regulating CoREST paralogs.

Overall the somatic mutations in KBTBD4 seem to alter the substrate binding pocket and cause tumorigenesis by a mechanism of neo-substrate recruitment. Such features are likely beneficial for drug discovery and personalised therapy. A small molecule that selectively binds a KBTBD4 mutant will specifically target MB cancer harbouring the corresponding mutations. The off-target effect will be minimal as the KBTBD4 disease mutants are unique to cancer cells over normal cells.

6.4. References

1. Northcott, P.A., et al., *The whole-genome landscape of medulloblastoma subtypes*. Nature, 2017. **547**(7663): p. 311-317.
2. Northcott, P.A., et al., *Medulloblastoma*. Nat Rev Dis Primers, 2019. **5**(1): p. 11.
3. Northcott, P.A., et al., *Medulloblastoma comprises four distinct molecular variants*. J Clin Oncol, 2011. **29**(11): p. 1408-14.
4. Jones, D.T., et al., *Dissecting the genomic complexity underlying medulloblastoma*. Nature, 2012. **488**(7409): p. 100-5.
5. Northcott, P.A., et al., *Subgroup-specific structural variation across 1,000 medulloblastoma genomes*. Nature, 2012. **488**(7409): p. 49-56.
6. Pugh, T.J., et al., *Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations*. Nature, 2012. **488**(7409): p. 106-10.
7. Robinson, G., et al., *Novel mutations target distinct subgroups of medulloblastoma*. Nature, 2012. **488**(7409): p. 43-8.
8. Lee, J.C., et al., *Recurrent KBTBD4 small in-frame insertions and absence of DROSHA deletion or DICER1 mutation differentiate pineal parenchymal tumor of intermediate differentiation (PPTID) from pineoblastoma*. Acta Neuropathologica, 2019. **137**(5): p. 851-854.
9. Sharma, T., et al., *Second-generation molecular subgrouping of medulloblastoma: an international meta-analysis of Group 3 and Group 4 subtypes*. Acta Neuropathol, 2019.
10. Northcott, P.A., et al., *Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma*. Nat Genet, 2009. **41**(4): p. 465-72.
11. Laugesen, A. and K. Helin, *Chromatin repressive complexes in stem cells, development, and cancer*. Cell Stem Cell, 2014. **14**(6): p. 735-51.
12. Pajtler, K.W., et al., *The KDM1A histone demethylase is a promising new target for the epigenetic therapy of medulloblastoma*. Acta Neuropathologica Communications, 2013(2051-5960 (Electronic)).
13. Callegari, K., et al., *Pharmacological inhibition of LSD1 activity blocks REST-dependent medulloblastoma cell migration*. Cell Commun Signal, 2018. **16**(1): p. 60.
14. Kalin, J.H., et al., *Targeting the CoREST complex with dual histone deacetylase and demethylase inhibitors*. Nat Commun, 2018. **9**(1): p. 53.
15. Maiques-Diaz, A., et al., *Enhancer Activation by Pharmacologic Displacement of LSD1 from GF1 Induces Differentiation in Acute Myeloid Leukemia*. Cell Rep, 2018. **22**(13): p. 3641-3659.

16. Roussel, M.F. and J.L. Stripay, *Epigenetic Drivers in Pediatric Medulloblastoma*. Cerebellum, 2018. **17**(1): p. 28-36.
17. Sheng, W., et al., *LSD1 Ablation Stimulates Anti-tumor Immunity and Enables Checkpoint Blockade*. Cell, 2018. **174**(3): p. 549-563 e19.
18. Yamamoto, R., et al., *Selective dissociation between LSD1 and GFI1B by a LSD1 inhibitor NCD38 induces the activation of ERG super-enhancer in erythroleukemia cells*. Oncotarget, 2018(1949-2553 (Electronic)).
19. Yi, J. and J. Wu, *Epigenetic regulation in medulloblastoma*. Mol Cell Neurosci, 2018. **87**: p. 65-76.
20. Northcott, P.A., et al., *The whole-genome landscape of medulloblastoma subtypes*. Nature, 2017. **547**(7663): p. 311-317.
21. Yang, M., et al., *Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase*. Mol Cell, 2006. **23**(3): p. 377-87.
22. Turnley, A.M., *Role of SOCS2 in growth hormone actions*. Trends in Endocrinology & Metabolism, 2005. **16**(2): p. 53-58.
23. Metcalf, D., et al., *Gigantism in mice lacking suppressor of cytokine signalling-2*. Nature, 2000. **405**(6790): p. 1069-1073.
24. Greenhalgh, C.J., et al., *Biological Evidence That SOCS-2 Can Act Either as an Enhancer or Suppressor of Growth Hormone Signaling*. Journal of Biological Chemistry, 2002. **277**(43): p. 40181-40184.
25. Northcott, P.A., et al., *Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma*. Nature, 2014. **511**(7510): p. 428-34.
26. Ballas, N., et al., *REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis*. Cell, 2005. **121**(4): p. 645-657.
27. Li, L., et al., *ZNF516 suppresses EGFR by targeting the CtBP/LSD1/CoREST complex to chromatin*. Nat Commun, 2017. **8**(1): p. 691.
28. Monaghan, C.E., et al., *REST corepressors RCOR1 and RCOR2 and the repressor INSM1 regulate the proliferation-differentiation balance in the developing brain*. Proc Natl Acad Sci U S A, 2017. **114**(3): p. E406-E415.
29. You, A., et al., *CoREST is an integral component of the CoREST- human histone deacetylase complex*. Proc. Natl. Acad. Sci. USA, 2000(0027-8424 (Print)).
30. Hakimi, M.A., et al., *A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes*. Proc. Natl. Acad. Sci. USA, 2002. **99**(11): p. 7420-5.
31. Laurent, B., et al., *A short Gfi-1B isoform controls erythroid differentiation by recruiting the LSD1-CoREST complex through the dimethylation of its SNAG domain*. J Cell Sci, 2012. **125**(Pt 4): p. 993-1002.
32. Boxer, L.D., et al., *ZNF750 interacts with KLF4 and RCOR1, KDM1A, and CTBP1/2 chromatin regulators to repress epidermal progenitor genes and induce differentiation genes*. Genes Dev, 2014. **28**(18): p. 2013-26.
33. Dobson, T.H.W., et al., *Transcriptional repressor REST drives lineage stage-specific chromatin compaction at *Ptch1* and increases AKT activation in a mouse model of medulloblastoma*. Science Signaling, 2019. **12**(565): p. eaan8680.
34. Taylor, P., et al., *REST is a novel prognostic factor and therapeutic target for medulloblastoma*. (1538-8514 (Electronic)).
35. Becher, O.J., *HDAC inhibitors to the rescue in Sonic Hedgehog Medulloblastoma*. Neuro Oncol, 2019.
36. Marabelli, C., B. Marrocco, and A. Mattevi, *The growing structural and functional complexity of the LSD1/KDM1A histone demethylase*. Curr Opin Struct Biol, 2016. **41**: p. 135-144.

37. Andres, M.E., et al., *CoREST: a functional corepressor required for regulation of neural-specific gene expression*. Proc. Natl. Acad. Sci. USA, 1999(0027-8424 (Print)).
38. Wang, Y., et al., *LSD1 co-repressor Rcor2 orchestrates neurogenesis in the developing mouse brain*. Nat Commun, 2016. **7**: p. 10481.
39. Barrios Ap Fau - Gomez, A.V., et al., *Differential properties of transcriptional complexes formed by the CoREST family*. Molecular and Cellular Biology, 2014(1098-5549 (Electronic)).
40. Upadhyay, G., et al., *Antagonistic actions of Rcor proteins regulate LSD1 activity and cellular differentiation*. Proc. Natl. Acad. Sci. USA, 2014. **111**(22): p. 8071-6.

Chapter 7

A selectivity panel of Kelch proteins

7.1. Introduction

As described in section 1.3.2., many BTB-Kelch proteins play important physiological roles in cells and their dysfunction can lead to various diseases. For example, KEAP1 inhibits the anti-oxidant defence process by degrading NRF2 through the ubiquitin-proteasome system [1]. Targeting KEAP1 with small molecule inhibitors is regarded as an attractive strategy in chemoprevention for cancer or treatment for diseases involving inflammation [2, 3]. Extensive efforts have been made to discover drug candidates, with the most advanced non-covalent compound GSK-7 binding to the KEAP1 Kelch domain with a K_D of 1.3 nM [4-10]. Most relevant to my work, the Growth Factor Signalling and Ubiquitination Group in the SGC has collaborated with the CHDI Foundation to characterize 8 literature and proprietary non-covalent inhibitors of the KEAP1 Kelch domain, including GSK-7.

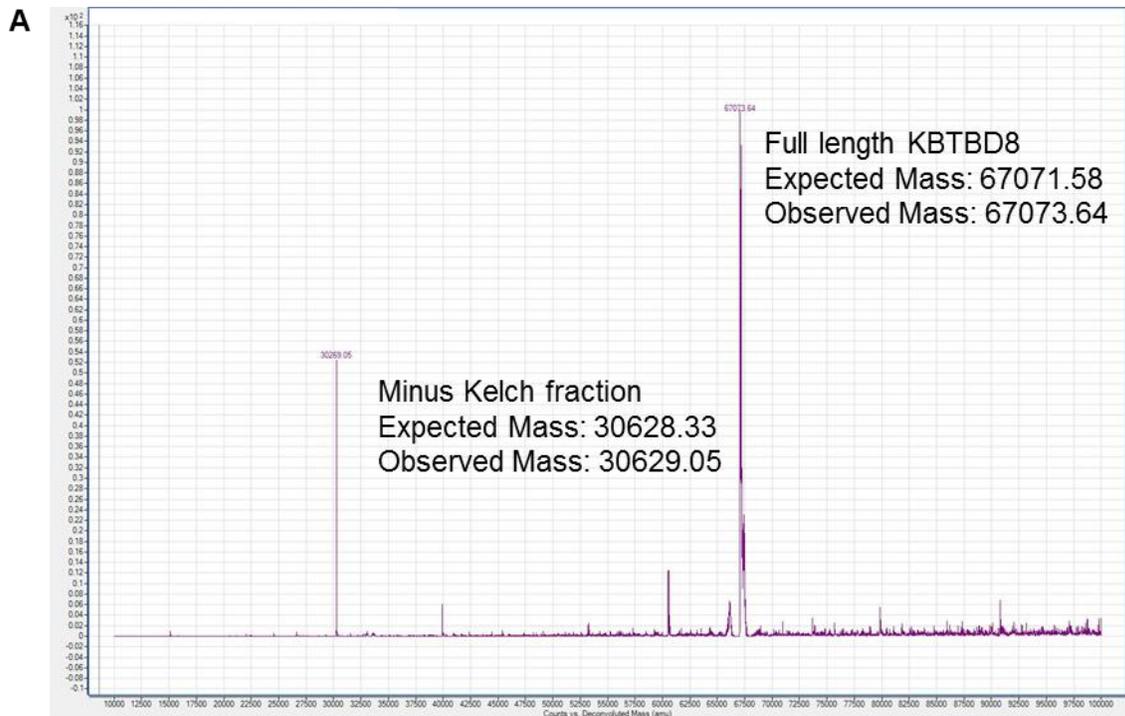
In this chapter, I expressed and purified several other BTB-Kelch proteins for screening against these 8 KEAP1 inhibitors as well as for further crystallisation trials. These studies were designed to develop a selectivity panel of human Kelch domain proteins to understand how selectively the Kelch domains could be targeted by small molecule inhibitors and to further define the structural diversity of the substrate pockets. While selectivity is desirable for therapeutic development, some cross-reactivity, if observed, would kick start inhibitor development for other Kelch family members. From these studies, I was able to determine the structure of the KLHL17 Kelch domain. I also observed that the non-covalent inhibitors are highly selective for KEAP1 over other family homologues.

7.2. Results

7.2.1. Protein expression and purification

I cloned constructs of human KBTBD10^{289-606a.a.} and KBTBD10^{300-606a.a.} into the pNIC28-Bsa4 vector. The domain boundaries were designed based on the homology alignment with the Kelch domain of rat KBTBD10, for which the crystal structure has been solved (PDB: 2WOZ). The construct of KBTBD10^{289-606a.a.} yielded far more protein than the other in a 50 mL scale expression test and was therefore chosen for large scale (2 L) expression. Constructs of human Kelch domains of KLHL2, KLHL3, KLHL7, KLHL12, KLHL17, KLHL20 and KBTBD5 were cloned into the pNIC28-Bsa4 vector by SGC staff before my DPhil project started. I expressed these Kelch domains in the BL21(DE3)-R3-pRARE2 strain. All constructs were His-tagged and therefore purified using Nickel sepharose chromatography followed by TEV cleavage of the tag, gel filtration and ion exchange. Detailed expression and purification methods are provided in sections 2.2 and 2.3.

KBTBD8 and KLHL21 were expressed utilising a pFB-LIC-Bse vector for baculoviral expression in Sf9 cell cultures. The constructs of KBTBD8 and KLHL21 were also cloned by SGC staff. I performed transposition in DH10Bac, transfection and virus amplification in Sf9 for the constructs of the full length KBTBD8 and Kelch domains of KBTBD8 and KLHL21. The expression was scaled up to 2 L for each construct. However the full length KBTBD8 protein got cleaved at a RLG[^]MTA sequence site (Figure 7.1) during the purification steps, causing the partial loss of the Kelch domain in the final product. As a result, following experiments were performed with the construct of the KBTBD8 Kelch domain only.



B

SMPSSDPASDAMPFHACSIKQLKTM YDEGQLTDIVVEVDHGKTFSCHRNVLA AISPYFR
 SMFTSGLTESTQKEVRIVGVEAESMDLVLN YAYTSRVILTEANVQALFTAASIFQIPSIQD
 QCAKYMISHLDPQNSIGVFI FADHYGHQELGDRSKEYIRKKFLCVTKEQEF LQLTKDQLIS
 ILDSDDLNV DREEHVYESIIRWFEHEQNEREVHLPEIFAKCIRFPLMEDTFIEKI PPQFAQ
 AIAKSCVEKGPSNTNGCTQRLG^{MTA}SEMIICFDAAHKHS GKKQTVPCLDIVTGRVFKLCKP
 PNDLREVGILVSPDNDIYIAGGYRPSSEV SIDHKAENDFWMYDHSTNRWLSKPSLLRARI
 GCKLVYCCGKMYAIGGRVYEGDGRNSLKSVECYDSRENCWTTVCAMPVAMEFHNAVEYKEK
 IYVLQGEFFLFYEPQKDYWGFLTPMTVPRIQGLAAVYKDSIYYIAGTCGNHQRMFMTVEAYD
 IELNKWTRKKDFPCDQSINPYLKLVLVFNKHLH L FVRATQVTVEEHVFRTSRKNSLYQYDDI
 ADQWMKVYETPDRLWDLGRHFEC AVAKLYPQCLQ

Figure 7.1 Full length KBTBD8 was cleaved at RLG^{MTA}. (A) Intact mass spectrometry analysis for purified full length KBTBD8. Two species were de-convoluted at 30629.05 Da and at 67073.64 Da, corresponding to a truncated fraction lacking the Kelch and a full length fraction in the KBTBD8 sample respectively, as annotated in the figure. (B) Corresponding protein sequence of the KBTBD8 full length construct. The theoretical mass of the sequence in black matches the observed mass of the smaller fraction in (A), indicating the protein was cleaved at a RLG^{MTA} site. The sequence in grey was lost during Ni affinity purification, as the 6xHis was N-terminally tagged to KBTBD8.

In addition to the 10 Kelch domain proteins that I prepared, a further 7 Kelch domain proteins, including KEAP1, were prepared at SGC by Dr Roslin Adamson, who was funded by the CHDI

Foundation. Together, these helped to establish a selectivity panel of 17 Kelch domain proteins (Figure 7.2A). Dr Roslin Adamson collaborated with me to perform a differential scanning fluorimetry assay to determine the thermostability of the Kelch domains and binding potency of the 8 KEAP1 inhibitors. This assay was chosen as it can be performed at scale and without knowledge of the relevant substrate peptides which is lacking for most of the Kelch domain proteins in the panel. Interestingly, the Kelch domain panel exhibited a variety of protein melting temperatures (Figure 7.2B). In the differential scanning fluorimetry assay in the presence of non-covalent small molecule KEAP1 inhibitors, significant T_m shifts of up to 22.3°C (compound 8, GSK-7) were observed with the KEAP1 Kelch domain, whereas all the other Kelch domains exhibited only marginal T_m shifts equivalent to the buffer only controls (Figure 7.2C). These data indicate the high selectivity of the KEAP1 inhibitors in this panel.

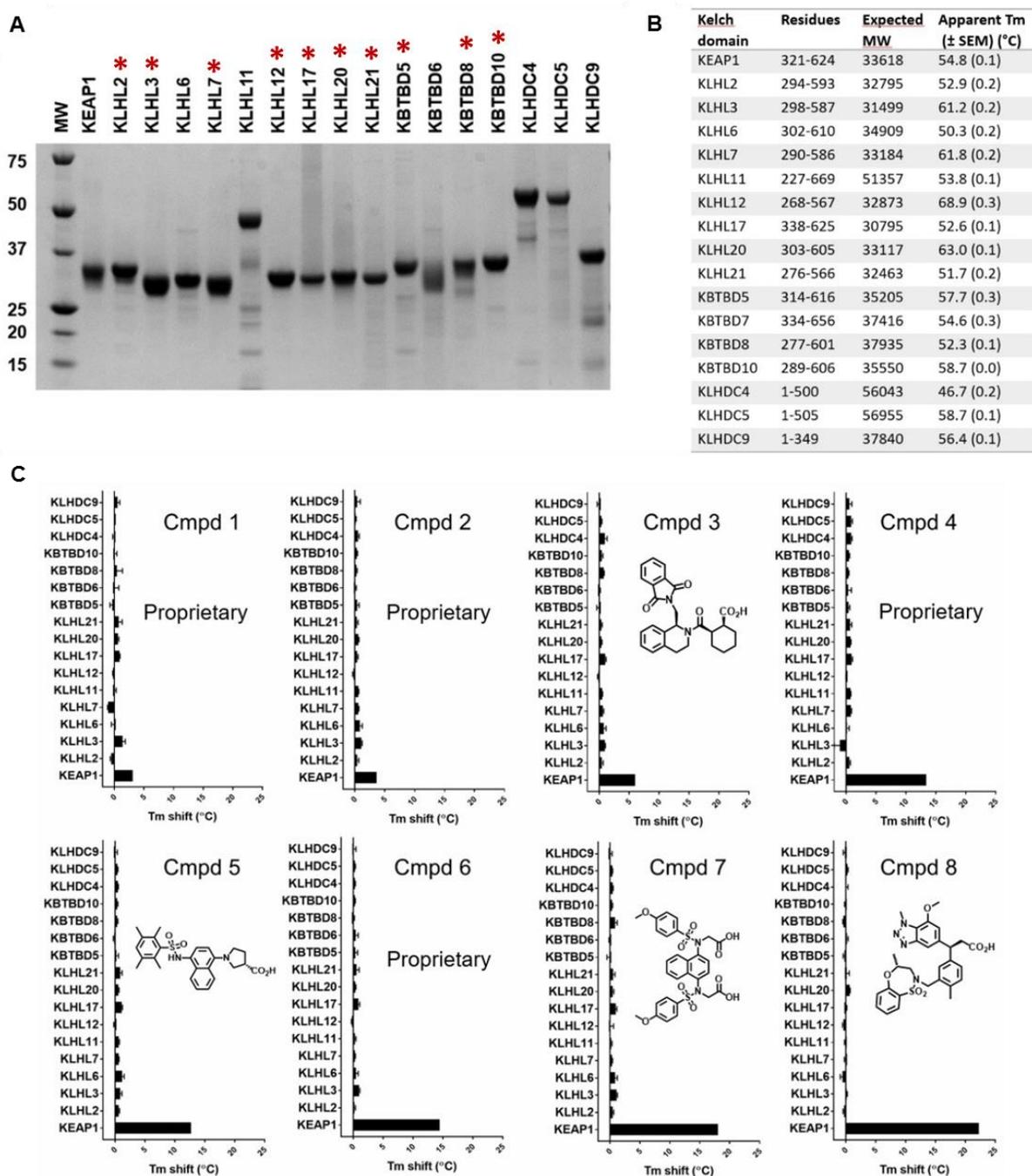


Figure 7.2 Developing a selectivity panel of human Kelch domains for non-covalent small molecule screening. (A) SDS PAGE of purified recombinant Kelch proteins. Proteins purified by myself are marked with asterisks. (B) Construct details and apparent T_m (in absence of compound or DMSO). (C) Inhibitor selectivity profiles. T_m shift values are plotted for different Kelch proteins in the presence of indicated non-covalent KEAP1 inhibitors. Each datapoint represents the mean value of at least three technical replicates. Error bars represent standard deviation. Assays in (B) and (C) were performed by Dr Roslin Adamson.

7.2.2. Crystallisation trials

To date, structures of human KLHL17, KLHL21, KBTBD8 and KBTBD10 Kelch domains have not been determined. Therefore I set up crystallisation trials for these proteins. Crystals of KLHL17 Kelch domain were obtained with good diffraction quality when mixing 150 nL of 10 mg/mL protein with 50 nL reservoir of 1.26 M ammonium sulfate, 0.2 M lithium sulfate, 0.1 M Tris pH 8.5 and grown at 20°C (Figure 7.3). Crystals were screened for diffraction on beamlines I04 and I24 at the Diamond Light Source, Didcot, U.K. A diffraction dataset leading to the structure was collected on beamline I04. Other trials yielded crystalline precipitate or microcrystals that have poor diffraction.

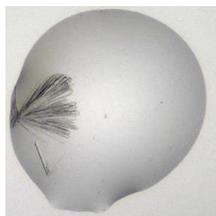


Figure 7.3 Crystals of KLHL17 Kelch domain. Crystals were obtained using a precipitant containing 1.26 M ammonium sulfate, 0.2 M lithium sulfate, 0.1 M tris pH 8.5.

7.2.3. Structure determination of KLHL17

The structure of KLHL17 Kelch domain was determined at 2.6 Å (Table 7.1) with 2 macromolecule chains in the asymmetric unit. The molecular replacement phasing was performed using 5NKP chain A (KLHL3) as a search model. Geometry factors and real space fitting were restrained in the first few rounds refinement. TLS and NCS were employed at the later stage in the refinement. The structure of KLHL17 Kelch domain was validated and deposited in PDB as 6HRL.

Table 7.1 Data collection and refinement statistics

Structure of human KLHL17 Kelch domain, PDB 6HRL	
Data collection	
Beamline	Diamond Light Source, I04
Wavelength (Å)	0.980
Resolution range (Å)	62.49 - 2.60 (2.69 - 2.60)
Space group	P 2 ₁ 2 ₁ 2 ₁
Unit cell dimensions	
<i>a, b, c</i> (Å)	67.4 118.6 125.0
α, β, γ (°)	90 90 90
Total reflections	194041
Unique reflections	31515 (3068)
Completeness (%)	97.2 (90.5)
Mean <i>I</i> / σ (<i>I</i>)	6.4 (1.6)
CC1/2	0.986 (0.343)
R-merge	0.188(0.809)
Refinement	
Reflections used in refinement	30691 (2817)
Reflections used for R-free	1448 (128)
R-work	0.229 (0.428)
R-free	0.269 (0.475)
Number of non-hydrogen atoms	4509
RMS deviation (bonds, Å)	0.007
RMS deviation (angles, °)	1.08
Ramachandran favored (%)	96.5
Ramachandran allowed (%)	3.5
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.14
Average B-factor (Å ²)	46.95

*Values in brackets show the statistics for the highest resolution shells. RMS indicates root-mean-square.

The structure traced the full KLHL17 Kelch domain from residues 337 to 623 (Figure 7.4A). The KLHL17 Kelch domain adopts the canonical β -propeller fold as described in section 1.3.1. The six Kelch repeats form the six blades (I to VI) of the propeller arranged radially around a central axis. Each repeat consists of four antiparallel β strands (A to D, Figure 7.4B). Conserved hydrophobic residues as well as several buried charged residues that recur within each Kelch

repeat mediates the packing within and between each blade (Figure 7.4B). Notably, the long BC loops which shape the substrate binding pocket of KLHL17, are all of equal length consisting of 11 residues, and appear conserved between blades II-VI (Figure 7.4B). Blade I, however, differs from the rest blades in β B strand length and BC loop sequence. These features likely reflect the substrate binding mode of KLHL17 which remains to be investigated.

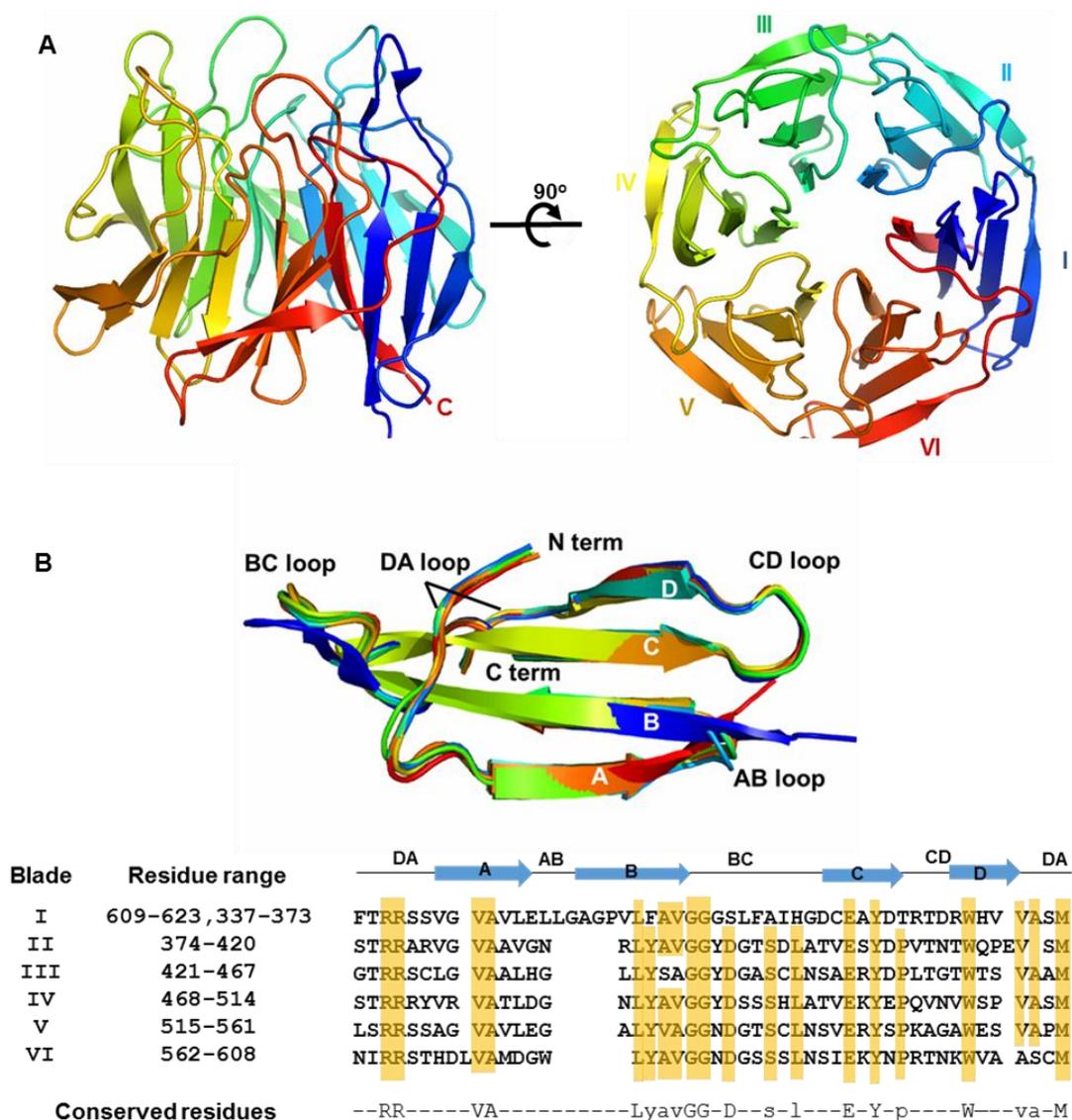


Figure 7.4 Structure of KLHL17 Kelch domain. (A) Overview of the KLHL17 Kelch domain structure in rainbow ribbon representation. The N and C termini and blades I to VI are labelled. **(B)** Superposition of Kelch repeats. Blades I to VI are coloured from blue to red. Each blade is composed of four β strands labelled A–D and connecting loops. The lower panel shows the

sequence alignment of six Kelch repeats in KLHL17. Conserved residues in each blade are highlighted in orange.

7.3. Discussion

In addition to KLHL20, KLHL12 and KLHL3 which were described in previous chapters, I purified another 7 human Kelch domains utilising either E.coli or Baculo-virus systems. This work contributes to generating a selectivity panel of 17 Kelch proteins, which is valuable for screening inhibitors against BTB-Kelch family E3 ligases. Indeed, in collaboration with Dr Roslin Adamson this was employed to show a high level of selectivity of KEAP1 inhibitors using thermoshift assays. Interestingly, variable apparent melting temperatures were observed with different Kelch domains, reflecting their diverse sequences, substrate pocket sizes and surface charge.

KLHL17, also known as actinfilin, binds to actin in the postsynaptic density [11] and controls the synapse localisation [12]. As an E3 ligase, KLHL17 recruits and targets the Kainate glutamate receptor subunit 6 (GluR6) for ubiquitin-dependent proteasomal degradation [12]. GluR6 has been implicated in epilepsy-induced excitotoxic neuronal death [13]. Consistent with this finding, clinical data also discovered that deletion of *KLHL17* is associated with infantile epilepsy disorder [14]. The structure of the KLHL17 Kelch domain solved in this chapter provides a starting point to understand the molecular basis of the functions of KLHL17 and further investigations are remaining to be carried out. For example, mapping of the degron in GluR6 and the nature of its substrate recognition by KLHL17.

The further characterisation of the other BTB-Kelch proteins reported in this chapter is underway. The recombinant proteins provide reagents for in vitro exploration. Some of them may become new targets for compound screening, potentially allowing development of a wider repertoire of E3 ligase handles for PROTAC development (see Chapter 8).

7.4. Reference

1. Bellezza, I., et al., *Nrf2-Keap1 signaling in oxidative and reductive stress*. *Biochim Biophys Acta Mol Cell Res*, 2018. **1865**(5): p. 721-733.
2. Cuadrado, A., et al., *Therapeutic targeting of the NRF2 and KEAP1 partnership in chronic diseases*. *Nature Reviews Drug Discovery*, 2019. **18**(4): p. 295-317.
3. Suzuki, T., H. Motohashi, and M. Yamamoto, *Toward clinical application of the Keap1–Nrf2 pathway*. *Trends in Pharmacological Sciences*, 2013. **34**(6): p. 340-346.
4. Hu, L., et al., *Discovery of a small-molecule inhibitor and cellular probe of Keap1–Nrf2 protein–protein interaction*. *Bioorganic & Medicinal Chemistry Letters*, 2013. **23**(10): p. 3039-3043.
5. Jiang, Z.-Y., et al., *Discovery of Potent Keap1–Nrf2 Protein–Protein Interaction Inhibitor Based on Molecular Binding Determinants Analysis*. *Journal of Medicinal Chemistry*, 2014. **57**(6): p. 2736-2745.
6. Hancock, R., et al., *Peptide inhibitors of the Keap1–Nrf2 protein–protein interaction*. *Free Radical Biology and Medicine*, 2012. **52**(2): p. 444-451.
7. Abed, D.A., et al., *Discovery of direct inhibitors of Keap1–Nrf2 protein–protein interaction as potential therapeutic and preventive agents*. *Acta Pharmaceutica Sinica B*, 2015. **5**(4): p. 285-299.
8. Marcotte, D., et al., *Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism*. *Bioorganic & Medicinal Chemistry*, 2013. **21**(14): p. 4011-4019.
9. Davies, T.G., et al., *Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery*. *J Med Chem*, 2016. **59**(8): p. 3991-4006.
10. Heightman, T.D., et al., *Structure-Activity and Structure-Conformation Relationships of Aryl Propionic Acid Inhibitors of the Kelch-like ECH-Associated Protein 1/Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1/NRF2) Protein-Protein Interaction*. *J Med Chem*, 2019. **62**(9): p. 4683-4702.
11. Chen, Y., et al., *Actinfilin, a Brain-specific Actin-binding Protein in Postsynaptic Density*. *Journal of Biological Chemistry*, 2002. **277**(34): p. 30495-30501.
12. Salinas, G.D., et al., *Actinfilin Is a Cul3 Substrate Adaptor, Linking GluR6 Kainate Receptor Subunits to the Ubiquitin-Proteasome Pathway*. *Journal of Biological Chemistry*, 2006. **281**(52): p. 40164-40173.
13. Vincent, P. and C. Mulle, *Kainate receptors in epilepsy and excitotoxicity*. *Neuroscience*, 2009. **158**(1): p. 309-323.
14. Paciorkowski, A.R., et al., *Copy number variants and infantile spasms: evidence for abnormalities in ventral forebrain development and pathways of synaptic function*. *European journal of human genetics : EJHG*, 2011. **19**(12): p. 1238-1245.

Chapter 8

Conclusions and Future Directions

The BTB-Kelch family plays critical roles in various cellular processes by targeting a wide range of substrates. The aims of this thesis were to understand the substrate recruitment mechanisms for BTB-Kelch proteins, with research focused on four representative members – KLHL20, KLHL12, KLHL3 and KBTBD4. Multiple techniques were employed in this thesis to provide comprehensive data. Novel substrates of KBTBD4 in tumour driving events were identified by affinity-purified mass spectrometry proteomics and validated by co-immunoprecipitation and ubiquitination assay in HEK293T cells. For reported substrates of KLHL20 and KLHL12, I successfully mapped the key interacting motifs (degrons) in substrates using SPOTs peptide arrays. The putative degrons were subsequently confirmed with both biophysical binding assays and cellular assays. Degron peptides were co-crystallised with corresponding Kelch domains to unravel the structural basis of the interactions of KLHL20-DAPK1, KLHL12-DVL1 and KLHL3-WNK3. Furthermore, the degron peptides provided tools for peptide-displacement based inhibitor development for BTB-Kelch proteins, with KLHL20 inhibitor screening as a proof-of-concept example. Expression and purification of other recombinant Kelch domains generated reagents for future BTB-Kelch study as well as a selectivity panel for drug discovery and characterisation. This work enabled a further structure of the KLHL17 Kelch domain and, in collaboration with Dr Roslin Adamson, revealed that existing KEAP1 inhibitors are highly selective for KEAP1 over other BTB-Kelch family members.

Despite adopting a conserved fold, the Kelch family proteins are relatively diverse in their primary sequences (25~50% identity), especially in the large loops that define the critical functional sites. The substrate binding pocket is shaped by the six long loops connecting the β 2- β 3 strands and the buried β 4- β 1 loops connecting adjacent blades. The low sequence similarity across the Kelch domains is reflected in the diversity of their electrostatic surface potentials and substrate binding pocket shapes (Figure 8.1). The substrate binding pockets of KLHL12 and KBTBD5 both exhibit fairly neutral potential on the surface. In contrast, the surface of the substrate binding pockets of KLHL17, KLHL20, KLHL3 and KEAP1 shows positive potential. In terms of pocket shapes, the pockets of KLHL20 and KEAP1 are of similar size, whereas other Kelch proteins shown in Figure 8.1 possess either shallower or narrower pockets.

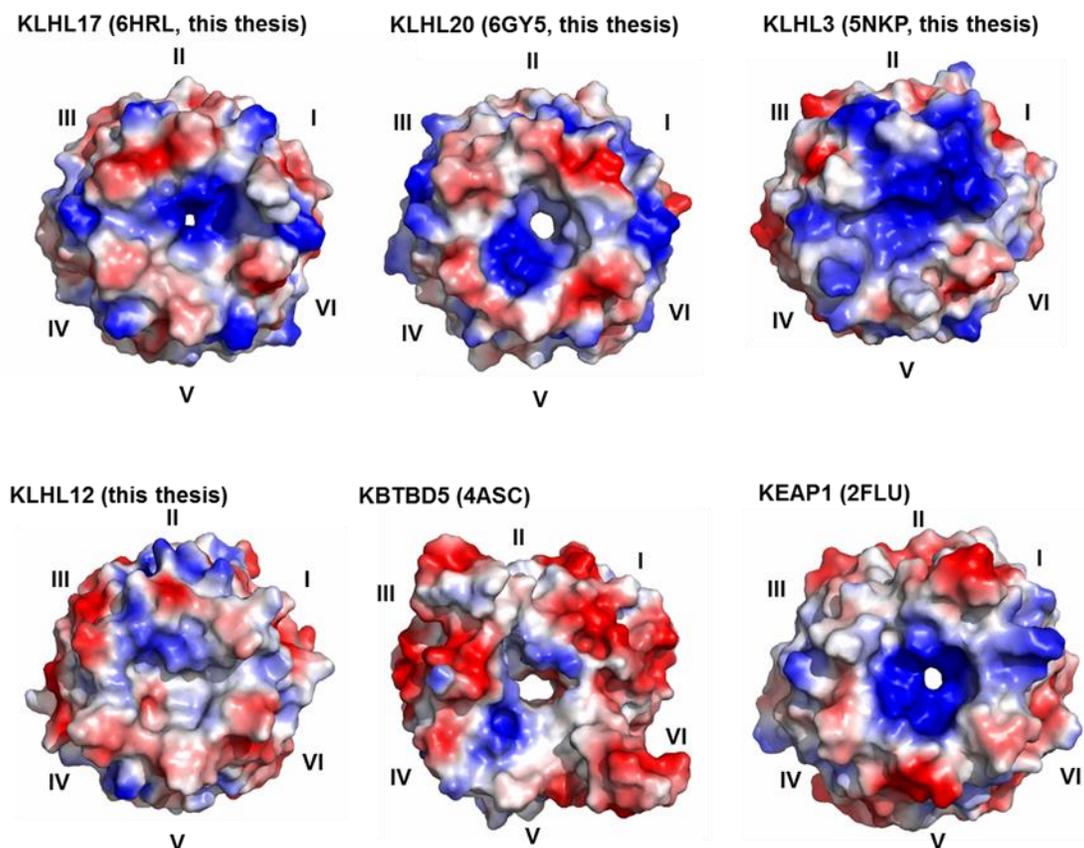


Figure 8.1 Structural diversity across different Kelch domains. The substrate binding pockets across the BTB-Kelch family show different pocket shapes and electrostatic surface potential. As shown in the colour bar, the red colour indicates negative potential while the blue colour indicates the positive potential near the surface. The white colour corresponds to fairly neutral potential. Blades are labelled I-VI for each Kelch domain and the PDB codes are indicated in parentheses.

The structural divergence across BTB-Kelch family members provides the molecular basis for the substrate specificity. To date, few Kelch-substrate complexes have been structurally characterized, with the major examples being the KEAP1-NRF2 [1-3] and KLHL3-WNK4 systems [4]. Superposition of KEAP1-NRF2 and the complex structures solved in this thesis shows that the substrate peptides are bound to their respective Kelch domains at different positions within the central pocket of the β -propeller (Figure 8.2A). For example, the NRF2 peptide inserts towards KEAP1 blades I and VI, whereas the WNK3 peptide packs more

towards KLHL3 blades III and IV (Figure 8.2A). Both DVL1 and DAPK1 pack at the centre of the pockets of KLHL12 and KLHL20 respectively (Figure 8.2A). These differences are supported by the variable $\beta 2$ - $\beta 3$ loop lengths observed across the BTB-Kelch family [5]. The patterning of hydrophobic and charged residues also differs across the different structures (Figures 8B-D).

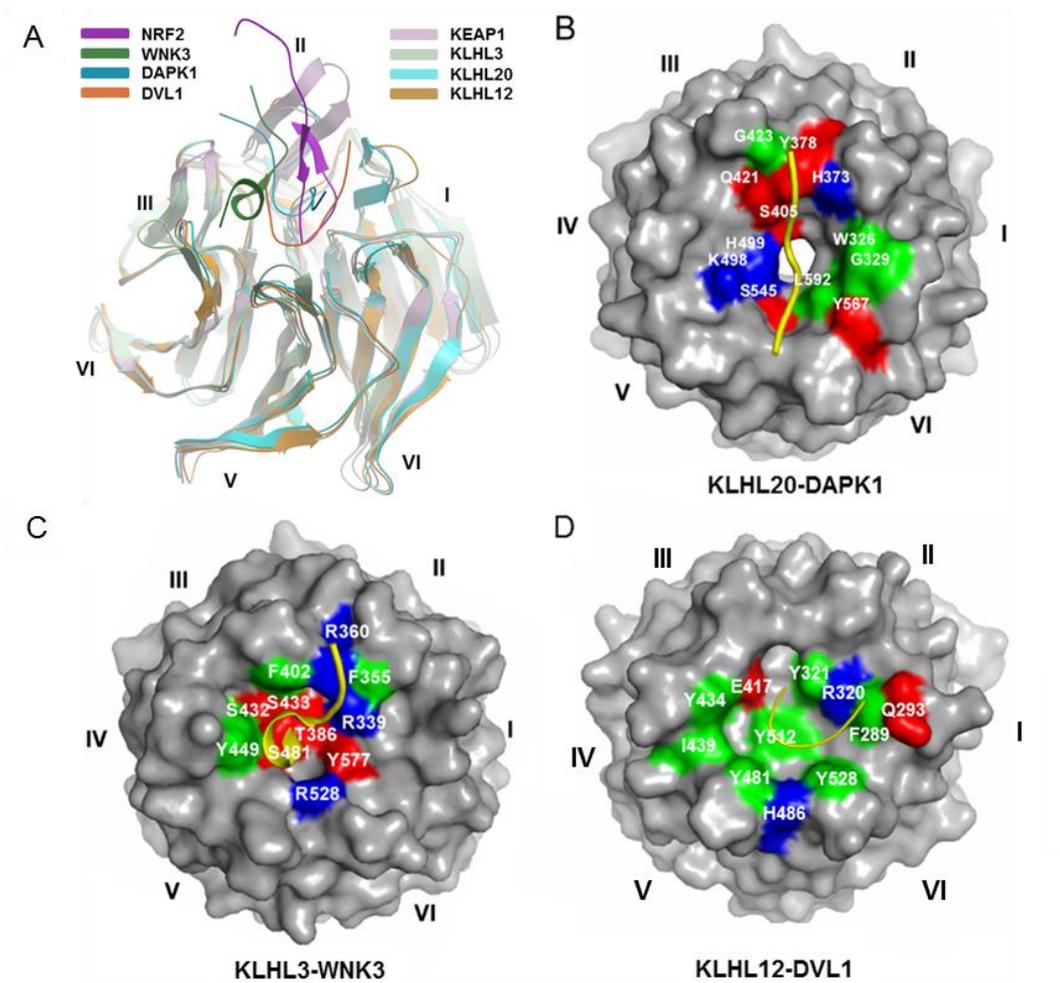


Figure 8.2 Kelch-substrate complexes reveal distinct binding modes. (A) Superposition of the complex structures of KEAP1-NRF2 (PDB: 2FLU, light and dark purple), KLHL3-WNK3 (PDB: 5NKP, light and dark green), KLHL20-DAPK1 (PDB: 6GY5, light and dark blue), and KLHL12-DVL1 (light and dark orange). (B-D) Surface representations of the Kelch domains in the KLHL20-DAPK1 (B), KLHL3-WNK3 (C), and KLHL12-DVL1 (D) complexes with key contact residues highlighted by their binding characteristics (blue, basic residue; red, other polar; green, hydrophobic). The distinct surfaces and bound peptide conformations (yellow ribbons) highlight the rich variety of binding modes that can be established by the Kelch domain substrate pockets.

It has been previously reported that a related Kelch-SOCS box family E3 ligase CUL2^{KLHDC2} recognises and binds di-glycine C-end degrons [6]. By similarity, it is likely that the multiple substrates of a BTB-Kelch protein possess conserved recognition motifs. Indeed, a 'PGXPP' motif identified in DVL paralogs is observed in other binders of KLHL12. For KLHL20, the 'LPDLV' motif identified in DAPK1 and the 'GPDVL' motif in EPAS1 are also of some resemblance. More investigation is needed to validate these conserved patterns. However, the available data may be helpful in future to map putative binding sites in other binders or substrates for KLHL12 and KLHL20.

As described above and in section 1.5, the workflow of characterising a known substrate with the respective BTB-Kelch protein has been well-established. In addition to characterising the known substrates, I attempted to unravel the functions of KBTBD4 by identifying the novel substrates. The promising yet preliminary data suggest that KBTBD4 medulloblastoma mutants drive tumorigenesis by recruiting CoRESTs as neo-substrates. Future work will be guided by the established substrate characterisation workflow and will attempt to elucidate the disease mechanism at the molecular level.

In summary, my studies reveal the distinct mechanisms of substrate recruitment by KLHL20, KLHL12 and KLHL3, and provide starting points to further investigate KBTBD4 and KLHL17. The Kelch domain β -propeller offers a central pocket that can accommodate peptides with unexpected turns, twists and helices. Overall, this feature is likely to increase the diversity of substrate interaction modes and allow for selective drug design. The initial inhibitor screening for KLHL20 has been carried out and the best hits have been validated with low micromolar IC₅₀s. Future work will aim to solve co-crystal structures and order compound analogues to establish a route to improve potency. Recently, proteolysis targeting chimera (PROTAC) has emerged as a novel drug discovery strategy that attracts considerable attention [7, 8]. The PROTAC exploits bifunctional molecules which recruit target proteins with one end while hijack

E3 ligases with the other to induce ubiquitination and degradation of the target proteins. The advantages of this 'event-driven' technique include reduced drug exposure and the ability to target proteins that are intractable with traditional drug discovery strategies due to lack 'druggable pockets' [7]. Thus, the chemical binders at the BTB-Kelch substrate binding pockets, can not only be used to develop inhibitors, but also to serve as the 'E3 handles' for PROTAC development.

References

1. Lo, S.C., et al., *Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling*. EMBO J, 2006. **25**(15): p. 3605-17.
2. Fukutomi, T., et al., *Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degnon and Keap1*. Mol Cell Biol, 2014. **34**(5): p. 832-46.
3. Padmanabhan, B., et al., *Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer*. Mol Cell, 2006. **21**(5): p. 689-700.
4. Schumacher, F.R., et al., *Structural and biochemical characterization of the KLHL3-WNK kinase interaction important in blood pressure regulation*. Biochem J, 2014. **460**(2): p. 237-46.
5. Canning, P., et al., *Structural basis for Cul3 protein assembly with the BTB-Kelch family of E3 ubiquitin ligases*. J Biol Chem, 2013. **288**(11): p. 7803-14.
6. Rusnac, D.-V., et al., *Recognition of the Diglycine C-End Degron by CRL2KLHDC2 Ubiquitin Ligase*. Molecular Cell, 2018. **72**(5): p. 813-822.e4.
7. Lai, A.C. and C.M. Crews, *Induced protein degradation: an emerging drug discovery paradigm*. Nature reviews. Drug discovery, 2017. **16**(2): p. 101-114.
8. Sakamoto, K.M., et al., *Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(15): p. 8554-8559.

Appendix - List of constructs and cloning primers

Targets	Vector	Forward primers	Reverse primer	Cloning method	Comments
KBTBD10 Kelch domain	pNIC28-Bsa4	TACTTCCAATCCATGCCTGGTTACCTGAATGAC ATTCCC	TATCCACCTTTACTGTCAAGTTTGA CAGTTTGAAGAGATTTAAACG	LIC	N-terminal 6xHis tag encoded within vector backbone. M288-L606
KLHL20 Kelch domain	pcDNA3-N-Flag	TACTTCCAATCCATGCAAGGACCAAGGACGA GAC	TATCCACCTTTACTGTcaTGTCATTTTA ATAACTCTACGCCA	LIC	N-terminal Flag tag encoded within vector backbone. M303-T602
KLHL20 Full length	pcDNA3-N-Flag	TACTTCCAATCCATGGAAGGAAAAGCCAATGCG	TATCCACCTTTACTGTcaTGTCATTTTA ATAACTCTACGCCA	LIC	N-terminal Flag tag encoded within vector backbone
KBTBD4 WT Full length	pcDNA3-N-3xFlag	CCGGCCGGATCCATGGAATCACCAGAGGAGC CTGG	GCATACGTCGACTTAGGCCAACACA AACTGCAAATTG	REC	N-terminal 3xFlag tag encoded within vector backbone
KBTBD4 G3 mutant	pcDNA3-N-3xFlag	CCGGCCGGATCCATGGAATCACCAGAGGAGC CTGG; Mutagenesis forward - TTGTATGTGGTGGGAGGGTCCATccacggCCA CGGCGCATGTGGAAGT	GCATACGTCGACTTAGGCCAACACA AACTGCAAATTG; Mutagenesis reverse - ACTTCCATGCGCCGTGGcctggGA TGGACCTCCACCACATACAA	REC	295PRR insertion/deletion 295PRRR.
KBTBD4 G4 mutant	pcDNA3-3xFlag	CCGGCCGGATCCATGGAATCACCAGAGGAGC CTGG; Mutagenesis forward - TTGTATGTGGTGGGAGGGTCCcaCCACGG CGCATGTGGAAGT	GCATACGTCGACTTAGGCCAACACA AACTGCAAATTG; Mutagenesis reverse - ACTTCCATGCGCCGTGGtggGATG GACCTCCACCACATACAA	REC	295P insertion 295PP
KLHL12 Full length	pcDNA3.1(-)	GCATACGTCGACATGGGAGGCATTATGGCCCC	CCGGCCGAATTCTCACTTATCGTCGT CATCTTGTAACTCTCTCGCGGAGA ACAAACACC	REC	C-terminal Flag tag encoded in the reverse primer
DAPK1 WT Full length	pcDNA3.1(-)	GCATACGTCGACATGACCGTGTTCAGGCAGG AAAAAC	CCGGCCGAATTCTCAAGCGTAATCTG GAAATCTGTATGGGTA CCGGGATACACAGAGCTAATGG	REC	C-terminal HA tag encoded in the reverse primer
DAPK1 Missense mutant	pcDNA3.1(-)	GCATACGTCGACATGACCGTGTTCAGGCAGG AAAAAC; Mutagenesis forward primer - CGCCATGAACCTAGCCCTCCTgcagcaGTGGC AAAGTACAAACACAGTAACGG	CCGGCCGAATTCTCAAGCGTAATCTG GAAATCTGTATGGGTA CCGGGATACACAGAGCTAATGG; Mutagenesis forward primer - CCGTTACTGGTGTGTACTTTGCCA ctgctg AGGGAGGCTAAGTTCATGGC G	REC	1338DL-1338AA
DVL1 WT Full length	pcDNA3.1(-)	GCATACGTCGACATGTACCCATACGATGTCC AGATTACGCTATGGCGGAGACCAAGATTATCT ACCAC	CCGGCCGAATTCTCACATGATGTCCA CGAAGAACTCG	REC	N-terminal HA tag encoded in the forward primer
DVL1 Missense mutant	pcDNA3.1(-)	GCATACGTCGACATGTACCCATACGATGTCC AGATTACGCTATGGCGGAGACCAAGATTATCT ACCAC; Mutagenesis forward primer - GGCCTATACAGTGGTGGGGGGctgctgctgctg ctgct GTCCGGGAGCTGGCTGCC	CCGGCCGAATTCTCACATGATGTCCA CGAAGAACTCG; Mutagenesis forward primer - GGCAGCCAGCTCCCGGACagcagcagc agcagcagc CCCCCCCCCACTGTATAG GCC	REC	657PPGGPP-657AAAAAA
DVL1 Deletion mutant	pcDNA3.1(-)	GCATACGTCGACATGTACCCATACGATGTCC AGATTACGCTATGGCGGAGACCAAGATTATCT ACCAC; Mutagenesis forward primer - GGCCTATACAGTGGTGGGGGGTCCGGGAG CTGGCTGCC	CCGGCCGAATTCTCACATGATGTCCA CGAAGAACTCG; Mutagenesis forward primer - GGCAGCCAGCTCCCGGACCCCCCCC ACCACTGTATAGGCC	REC	657PPGGPP deletion
CoREST3 Full length isoform 2	pcDNA3.1(+)	CGGAGATCTATGTACCCATACGATGTCCAGA TTACGCTATGCCCGCATGATGGAGAAAGGG CCC	CGGCTCGAGTTAGTGCAGTGAGGAC TGTGAATCTGTCTGAATTCC	REC	N-terminal HA tag encoded in the forward primer
CoREST3 Full length isoform 3	pcDNA3.1(+)	CGGAGATCTATGTACCCATACGATGTCCAGA TTACGCTATGCCCGCATGATGGAGAAAGGG CCC	CGGCTCGAGTTAGTGCAGTGAGGAC TGTGAATCTGTCTGAATTCC	REC	N-terminal HA tag encoded in the forward primer