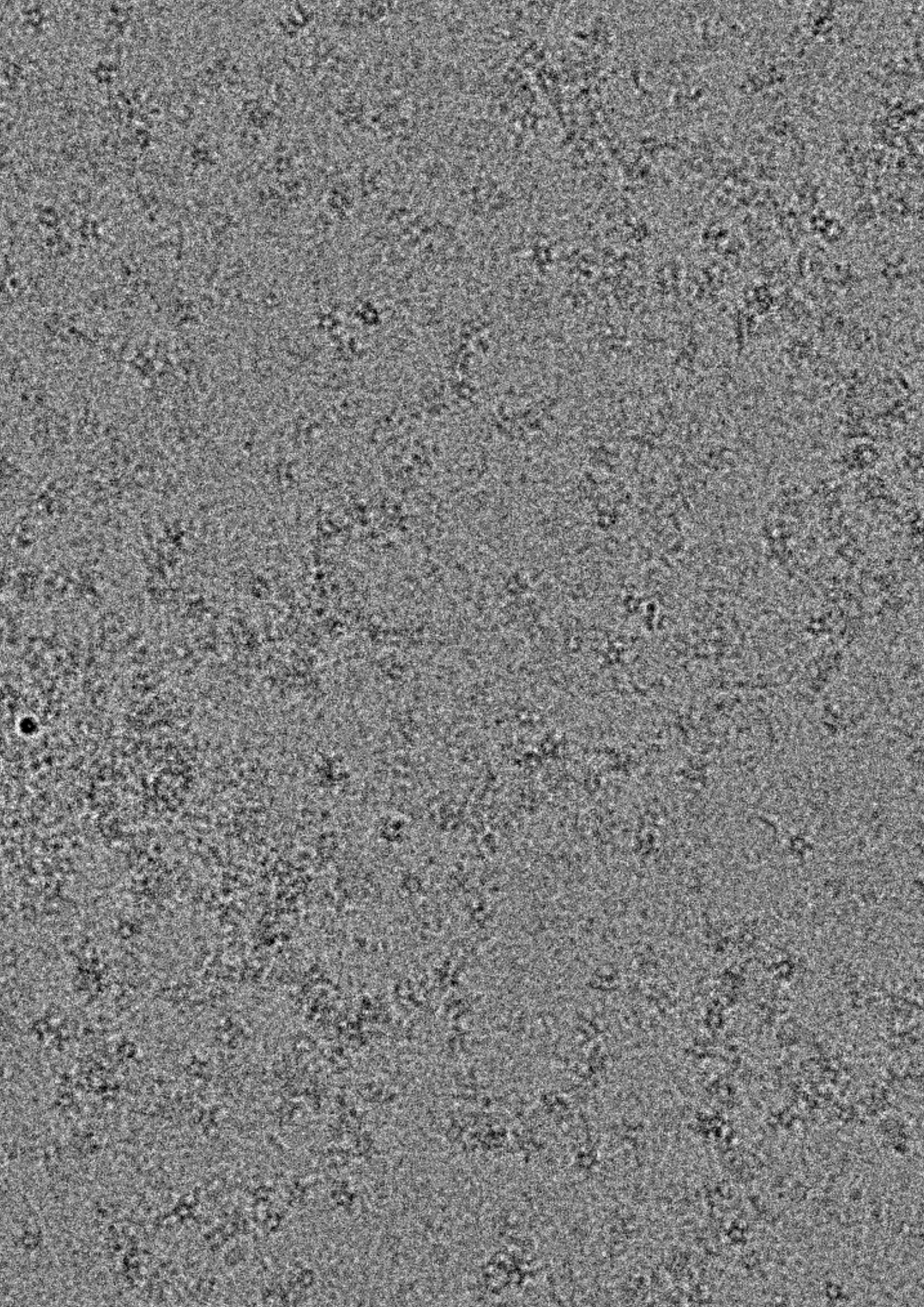


Structural Mechanisms of Multimeric BTB E3 Ligases and Interactors

Ashley E Ramdass

University of Oxford

Trinity Term 2025

Declaration

I declare that there are no parts of this thesis, or its research herein have been reproduced or accepted for another award, degree, or diploma at any other university, or learning institution. This thesis contains no other person's work except where stated in text.

Acknowledgments

I would like to express my deepest gratitude to the many people who have assisted me during my DPhil studies. From these people, I have learnt a great deal about scientific research, and about my own character:

- **Prof. Alex N Bullock** (Primary Supervisor). His insights and expertise have been instrumental in shaping this research project. I admire his patience, deep knowledge, and kindness.
- **Prof. Paul E Brennan** (Co-Supervisor). Although our chemical probe project was never realised, he always showed interest in my DPhil studies with his ever-present optimism.
- **Dr David B Sauer** (Internal Assessor) and **Prof. Stephen C Graham** (External Assessor). They both generously gave their time to read and assess this thesis, offering thoughtful and constructive feedback.
- **Dr Gamma Chi** and **Dr Ashley C W Pike**. They kindly shared their expertise and provided training in cryo-EM techniques.
- **Dr Zhuoyao Chen**. She shared substantial time and vast knowledge, particularly in X-ray crystallography and cell culture training.
- **Dr Xiangrong "Tina" Chen**. She assisted in my numerous efforts to establish biophysical assays and let me participate in her XChem experiments. She is hard working, humorous, and kind person.
- **Dr Julien Cros**. A friend who made the laboratory much more fun with his *joie de vivre*.

- Other members of the Bullock group: Dr Jesús Baltanás Copado, Dr Liz Brown, Will Burn, Dr David Cruz Walma, Dr Eva Dalietou, Dr Roxanne Lau, Charlotte Manning, Beatriz Montes Tan, Dr Will Richardson, Dr Mel Sweeney, Dr Lap “Carrie” Tse, and Dr Jong Fu Wong. This research group was a wonderful, diverse group of people that I came to see as a second family.

- Members of the Biotech SRF: Dr Souvika Bakshi, Laurent Brooke, Bindu, Humaira Islam, Dr Gavin McKinley, Dr Shubhashish Mukhopadhyay, and Dr Ellie Williams. For their support in protein production and cell culture training.

- Other members of the Centre for Medicines Discovery (CMD) including Dr Will Bradshaw, Dr Rod Chalk, Dr Jon Elkins, Dr Mike Fairhead, Dr Sarah Flannery, Prof. Benedikt Kessler, Dr Lizbe Koekemoer, Beth MacLean, Dr Brian Marsden, Elvis Ng, Dr Laura Ortega Varga, Chris Sluman, and Dr Iolanda Vendrell. It was a pleasure to be surrounded by so many talented researchers.

- Staff at the BBSRC Interdisciplinary bioscience doctoral training centre, who provided support and structure to my DPhil studies: Prof. Gail Preston (Academic supervisor), Prof. Esther Becker, Judith Bishop, and Jen Massey.

- The talented chemists at Servier (Budapest) with whom I completed my industrial placement: Prof. Andras Kotschy, Dr Andras Herner, Dr Zsolt Zsucs, Dr Timea Szabo, and Dr Timea Balo.

- **Debbie Ramdass**, **Adrian Ramdass**, **Marisa Garay**, and other friends and family. For their support both within and outside my studies.

- **Elliott Ramdass**. My twin brother, whose like-mindedness kept me laughing and feeling loved.

- **Jone McKeever Garay**. The kindest, sweetest person, with whom I have lived, loved, and grown together with over these past years. Thank you for everything.

COVID19 Statement

The UK went into national lockdown due to COVID-19 in March 2020, 7 months prior to the start of my DPhil studies. My initial training semester and first rotation project at the Centre for Medicines Discovery (CMD) was limited by strict social distancing rules aimed at curbing infections (3 days per week in the laboratory maximum). This also had minor effects at the start of my DPhil project, whilst everyone was reacclimatising to working back in the laboratory more regularly.

Additionally, my lab at the CMD underwent two moves: the first was going from the Old Road Campus Research Building to the Dorothy Crowfoot Hodgkin Building (DHCB) in September 2021, and the second going from DHCB to the current Nuffield Department of Medicine Research Building location in March 2023. Both moves resulted in disruptions (> 3 months) to my DPhil project, due to the amount of time spent packing and unpacking, the temporary loss of water supply, gas supply, functional equipment, and tissue culture capability. Despite these challenges, I was able to achieve the key objectives of my DPhil project and become proficient with molecular biology and structural biology techniques within the 4 years of funding.

Abbreviations Table

Abbreviation	Definition
AF	AlphaFold
AKT	Protein kinase B (involved in cell survival and metabolism)
ALG2	Apoptosis-linked gene 2 (a calcium-binding protein)
AMP	Adenosine monophosphate
APC/C	Anaphase-promoting complex
ASU	Asymmetric unit
ATP	Adenosine triphosphate
AUC	Analytical Ultra-centrifugation
BACK domain	BTB and C-terminal Kelch domain
BSA	Bovine serum albumin, or Buried Surface Area (Context)
BTB domain	Bric-a-brac, Tramtrack, Broad complex domain
CD1	Cyclin D1
CD3	Cyclin D3
CFTR	Cystic fibrosis transmembrane conductance regulator
CHB	Chronic hepatitis B
CMD	Centre for Medicines Discovery
COP9	COP9 signalosome
COPII	Coat Protein Complex II
COPS2	COP9 Signalosome Subunit 2
COPS3	COP9 Signalosome Subunit 3
Cryo-EM	Cryo-electron microscopy
CTD	C-terminal domain
CUL1, 2, 3, 4A, 4B, 5, 7, 9	Cullin-RING Ligase
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	Dithiothreitol
DUB	Deubiquitinating enzyme
DVL1-3	Dishevelled Protein 1-3
E1	Ubiquitin-activating enzyme
E2	Ubiquitin-conjugating enzyme
E3	Ubiquitin ligase
EGFR	Epidermal growth factor receptor
ER	Endoplasmic reticulum
ERAD	Endoplasmic-reticulum-associated protein degradation
FBS	Fetal bovine serum
FBXO31	F-box only 31
GFP	Green fluorescent protein
HBV-ACLF	Hepatitis B virus-induced acute-on-chronic liver failure
HECT	Homologous to the E6AP carboxyl terminus
HEK293	Human embryonic kidney 293 cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase
IL2	Interleukin-2
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry

KCTD	Potassium Channel Tetramerization Domain-containing protein
Kd	Dissociation constant
KEAP1	Kelch-like ECH-associated protein 1
KLHL	Kelch-like protein family
KLHL12	Kelch-like protein 12
MDM2	Mouse double minute 2 homolog
MS	Mass Spectrometry
NK cells	Natural killer cells
NTD	N-terminal domain
PBS	Phosphate-buffered saline
PDB	Protein Data Bank
PEF1	Peflin
PEI	Polyethylenimine
PLEKHA4	Pleckstrin Homology Domain Containing A4
POI	Protein of interest
PPR domain	Pentapeptide Repeat Domain
PROTAC	Proteolysis-targeting chimera
RBR E3 ligase	RING-between-RING E3 ligase
Rbx1	RING-box protein 1
Rbx2	RING-box protein 2
RING	Really Interesting New Gene
RMSD	Root mean square deviation
RNA	Ribonucleic acid
SCF complex	Skp1-Cullin-F-box protein complex
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
siRNA	Small interfering RNA
SKP1	S-phase kinase-associated protein 1
TCEP	Tris (2-carboxyethyl) phosphine
TPD	Targeted protein degradation
Ub	Ubiquitin
UBD	Ubiquitin-binding domain
UPS	Ubiquitin-proteasome system
USP	Ubiquitin-specific protease
Wnt	Portmanteau wingless and Int-1
WT	Wild-type
ZNT9	Zinc transporter 9
β-TrCP	Beta-transducin repeat-containing protein

Abstract

In recent years, E3 ligases and ubiquitination biology have garnered increased attention due to their critical role in emerging field targeted protein degradation field (TPD). However, many E3 ligases associate with multiple interactors and are often implicated in several diverse cellular processes. The ability to utilise these E3 ligases relies on complete characterisation of their structural mechanisms and interactor partners.

This thesis presents investigations into three distinct multimeric cullin-RING ligases. KCTD9 is a unique member of an unusual set of E3 ligases that typically assemble into higher-order assemblies. Newly generated structural data has provided insights into its function, while affinity mass spectrometry has identified a shortlist of potential interactors that may elucidate KCTD9's role in natural killer cell development.

Dimeric KLHL12 is a rare example of an E3 ligase that can engage multiple types of interactors through a shared recognition motif. Structural data confirm, that despite some variation in protein sequence among interactors, they adopt a common binding mode.

Finally, attempts were made to setup a reproducible crystal system and establish biophysical assays for the E3 ligase FBXO31, in order to screen compounds selected by virtual ligand screening. However, these experiments were waylaid by previous studies of FBXO31, in which its interactions with Cyclin D1 were poorly characterised.

Overall, these findings contribute to the growing understanding of E3 ligases, and this will aid in TPD developments.

Contents

Chapter 1: Introduction

1.0 Introduction

1.1 Ubiquitination

Ubiquitination describes the covalent linkage of ubiquitin (Ub), a small 76 amino acid protein, to the amino group of the N-terminal methionine residue or a surface lysine side chain of the substrate protein, through an ATP-dependent E1-E2-E3 enzyme cascade (Figure 1.1A,B).[1] The E1 ubiquitin-activating enzyme facilitates the ATP activation of a free Ub molecule and formation of thioester bond. A trans-thiolation reaction then then transfers the E1 bound Ub onto a cysteine residue of an E2 ubiquitin-conjugating enzyme. In the final stage of this process, the E3 ubiquitin ligase facilitates the formation of an isopeptide bond between the Ub molecule and the substrate protein; labelling the substrate with a Ub molecule.

After a single ubiquitination event (monoubiquitination), subsequent additions of Ub can be facilitated at either an alternative site on the substrate (multi-monoubiquitination), or onto the Ub molecule that previously was added; in order to form polymeric ubiquitin chains (polyubiquitination).[2,3] Given that Ub itself has eight potential sites (M1, K6, K11, K27, K29, K33, K48, and K63) for subsequent ubiquitination, there is a variety of homogenous, heterogenous, and branched polyubiquitin chains that can be created (Figure 1.1C).

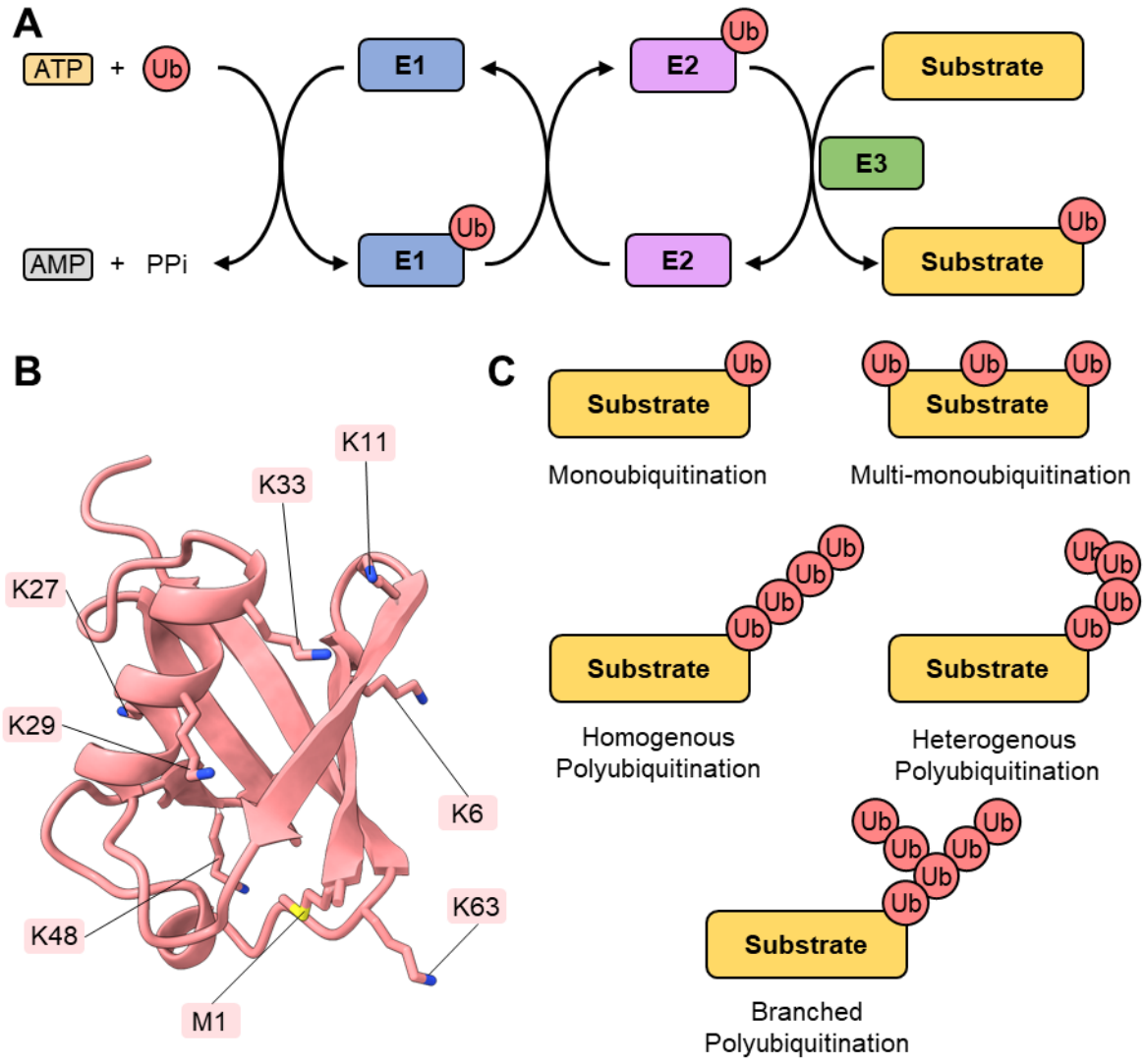


Figure 1.1: A) E1-E2-E3 ubiquitination enzyme cascade. B) Ubiquitin molecule (PDB: 1UBQ) with ubiquitination sites shown.[4] C) Ubiquitination pattern types.

The information stored by these distinct Ub patterns is often referred to as 'the ubiquitin code' and their role is implicated in almost every cellular mechanism.[2] The arrangement of these Ub molecules identifies the substrate for specific cellular processes and it is recognised, or 'read', by a wide range of interactors containing some form of a ubiquitin binding domain (UBD). While no single consensus motif defines all UBDs, there are over 20 different subtypes which have been identified (Table 1.1).[5]

UBD Class	Example Protein	Ubiquitin Signalling Role
Alpha-helical		
UIM	Rap80, Vps27, S5a/Rpn10, STAM, etc.	DNA repair, Endocytosis, Proteasomal degradation, etc.
IUIM/MIU	Rabex-5, RNF168	Endocytosis, DNA repair
DUIM	Hrs	Endocytosis
UBM	Polymerase iota	DNA repair
UBAN	NEMO, ABIN1/2/3	NF-kappaB signalling
UBA	hPLIC/Dsk2, NBR1, p62	Proteasomal degradation, Autophagy
CUE	Vps9	Endocytosis
GAT	TOM1, GGA3	Endocytosis
VHS	STAM, GGA3	Endocytosis
UMI	RNF168	DNA repair
Zinc-finger (ZnF)		
UBZ	Polymerase eta, Polymerase iota, TAX1BP1	DNA repair, NF-kappaB signalling
NZF	TAB2/3, Vps36, Npl4	NF-kappaB signalling, Endocytosis, ERAD
ZnF A20	Rabex-5, A20	Endocytosis, NF-kappaB signalling
ZnF UBP	IsoT/USP5, HDAC6	Deubiquitination, Autophagy
Pleckstrin-homology (PH)		
PRU	Rpn13	Proteasomal degradation
GLUE	Eap45	Endocytosis
Ubiquitin conjugating (UBC)-related		
UEV	TSG101, Mms2	Viral budding and VPS, DNA repair
UBC	UbcH5c	Ubiquitin conjugation/chain assembly
Src homology (SH)		
SH3	Sla1, CIN85	Endocytosis
WD40 beta-propeller	PLAA/Doa1, Cdc4	ERAD, Ubiquitin chain assembly
Additional UBDs		
PFU	PLAA/Doa1	ERAD
JAB1/MPN	Prp8	Pre-mRNA processing
DC-UbP_N	DC-UbP	Ubiquitinated substrate delivery
MDA-9/ syntenin UBD	MDA-9/syntenin	Endocytosis

Table 1.1: Ubiquitin binding domain structural folds summary table. Table curated from a review by L. Randles and K. J. Walters (table 2). Individual citations for each UBD subgroup can be found in this review.[5]

Monoubiquitination can trigger a number of cellular processes: DNA repair, transcription, endocytosis, or protein trafficking.[6,7] One specific example is the multi-monoubiquitination of receptor tyrosine kinases (RTKs) to promote their endocytosis.[8] The E3 ligase Cbl ubiquitinates transmembrane epidermal growth factor receptor (EGFR) and endocytic adaptor proteins recognise this modification through a UBD. These interactions promote endocytosis of the receptor and promote degradation via the lysosome, ultimately down regulating EGFR signal transduction.

K11 and K48 linked polyubiquitin chains are markers for protein degradation by the 26S proteasome; a 2.5 MDa multimeric complex that can be separated into two subcomplexes (Figure 1.2).[9,10] The regulatory particle recognises polyubiquitinated substrates through the Rpn1, Rpn10, and Rpn13 ubiquitin receptors. Once degradation of the substrate is initiated, the substrate is drawn into the core particle and Rpn11, a deubiquitinating enzyme (DUB), removes the polyubiquitin chain.[11] The core particle has a barrel-shaped structure containing two rings of β subunits that are sandwiched between two rings of α subunits. At the centre of this sub complex, there are a variety to proteases (capases, trypsin, etc.) which facilitate the degradation of the substrate into short (7 and 8-mer) peptides.

The ubiquitin-proteasome system (UPS) facilitates the degradation of a wide range of proteins, allowing it to serve as key regulator of protein levels in order to maintain homeostasis across a number of cellular processes.[12] One example is the tumour suppressor p53, which regulates cell cycle arrest and apoptosis, and whose protein levels are tightly controlled by the E3 ligase MDM2.[13] Secondly, Cyclin B1, a regulator of mitosis, is ubiquitinated by the multimeric anaphase-promoting complex/cyclosome (APC/C).[14]

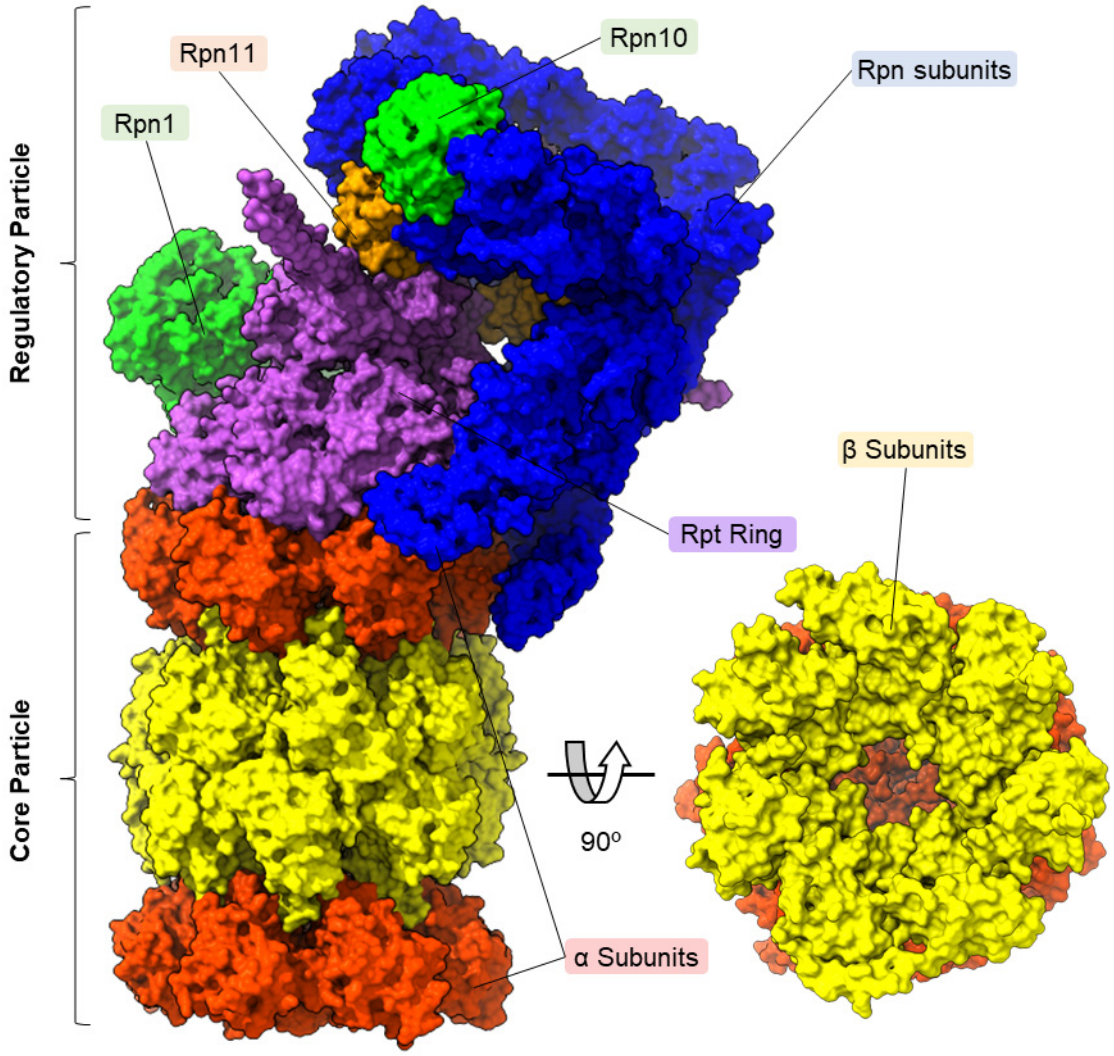


Figure 1.2: Human 26 proteasome (PDB: 5GJR).[10] α subunits are coloured red-orange, β subunits are coloured yellow, the Rpt ring is coloured violet, Rpn subunits are coloured blue. Ubiquitin receptors, Rpn1 and Rpn10, are coloured lime. DUB Rpn11 is coloured orange. Rpn13 was not resolved in cryo-EM dataset.

1.2 E3 Ligases and DUBs

Humans possess approximately two E1, 40 E2, and >600 E3 enzymes.[15] The specific Ub chain linkage is largely attributed to the E2 enzyme utilised in ubiquitination cascade, whereas the target and site selectivity is conferred by E3 ligase (as it facilitates substrate interactions). E3 ligases can be broadly divided into one of three families, defined by their catalytic domain and mechanism of Ub transfer: RING-type, HECT-type, and RBR-type.[16]

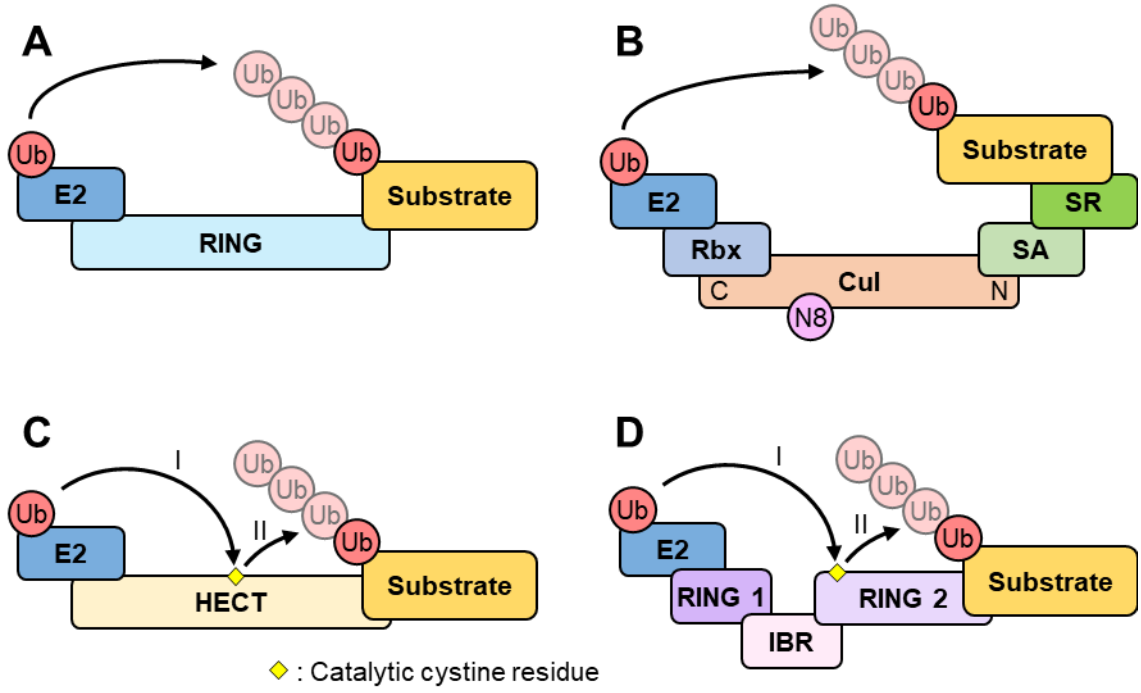


Figure 1.3: A) Monomeric RING finger E3 ligase. B) Multimeric cullin-RING E3 ligase assembly. SA: Substrate adaptor. SR: Substrate receptor. C) HECT E3 ligase. D) RBR E3 ligase. INR: In-between RING.

RING-type (really interesting new gene) E3 ligases are a large family of E3 ligases that contain a catalytic RING domain and facilitate the direct transfer of Ub to the substrate (Figure 1.3A,B).[17] This family can be divided further into two subgroups: monomeric RING finger proteins, and multimeric cullin-RING ligases (CRLs).[18]

In human CRLs, the cullin component (CUL1, 2, 3, 4a, 4b, 5, 7, or 9) serves as a scaffold protein between the E2 ligase-RING proteins (Rbx1 or Rbx2) and the adaptor proteins (such as F-Box/Skp1, KCTD9, etc.). Many subgroups of these components are interchangeable with one another and this combinatorial assembly allows CRLs to facilitate a wide range of selective ubiquitination activity.[18]

HECT (homologous to the E6AP carboxyl terminus) E3 ligases have a HECT domain which contains an active cystine residue.[19] In a two-step process, Ub is temporarily transferred from the E2 ligase to this cysteine residue, and then onto the substrate (Figure 1.3C). RBR (RING-between-RING) E3 ligases contain two distinct zinc binding domains, RING1 and RING2, separated by an in-between RING domain. The RING1 domain associates with a E2-Ub complex, and Ub is transferred to the substrate via a catalytic cystine residue found within the RING2 domain (Figure 1.3D).[20]

The cell can reverse the addition of Ub molecules through the use of deubiquitinating enzymes (DUBs).[21] These enzymes recognise a specific ubiquitin linkage and cleave the isopeptide to remove single Ub molecules or polyubiquitin chains from the substrate. In humans, there are more than 102 DUB genes which can be divided into two classes: cystine proteases, and metalloproteases.

The first class encompasses most of the DUBs, with over 80 examples in humans, and 58 members belong to the ubiquitin-specific proteases (USP) family. In these enzymes, the catalytic cystine region is situated between a palm region, and a thumb region, while a finger region mediates interactions with the ubiquitin moiety (Figure 1.4A).[22] Other subgroups that contain an active cystine catalytic site include ubiquitin C-terminal hydrolases (UCH), the ovarian tumour (OUT) family, and the Machado–Josephin domain (MJD) family. Metalloproteases are a smaller family of DUBs that are characterised by their utilisation of a zinc ion within the catalytic site to hydrolyse the ubiquitin isopeptide bond. One member includes Rpn11, which is integrated into the regulatory particle of the 26S proteasome and removes ubiquitin chains once degradation is initiated.[11]

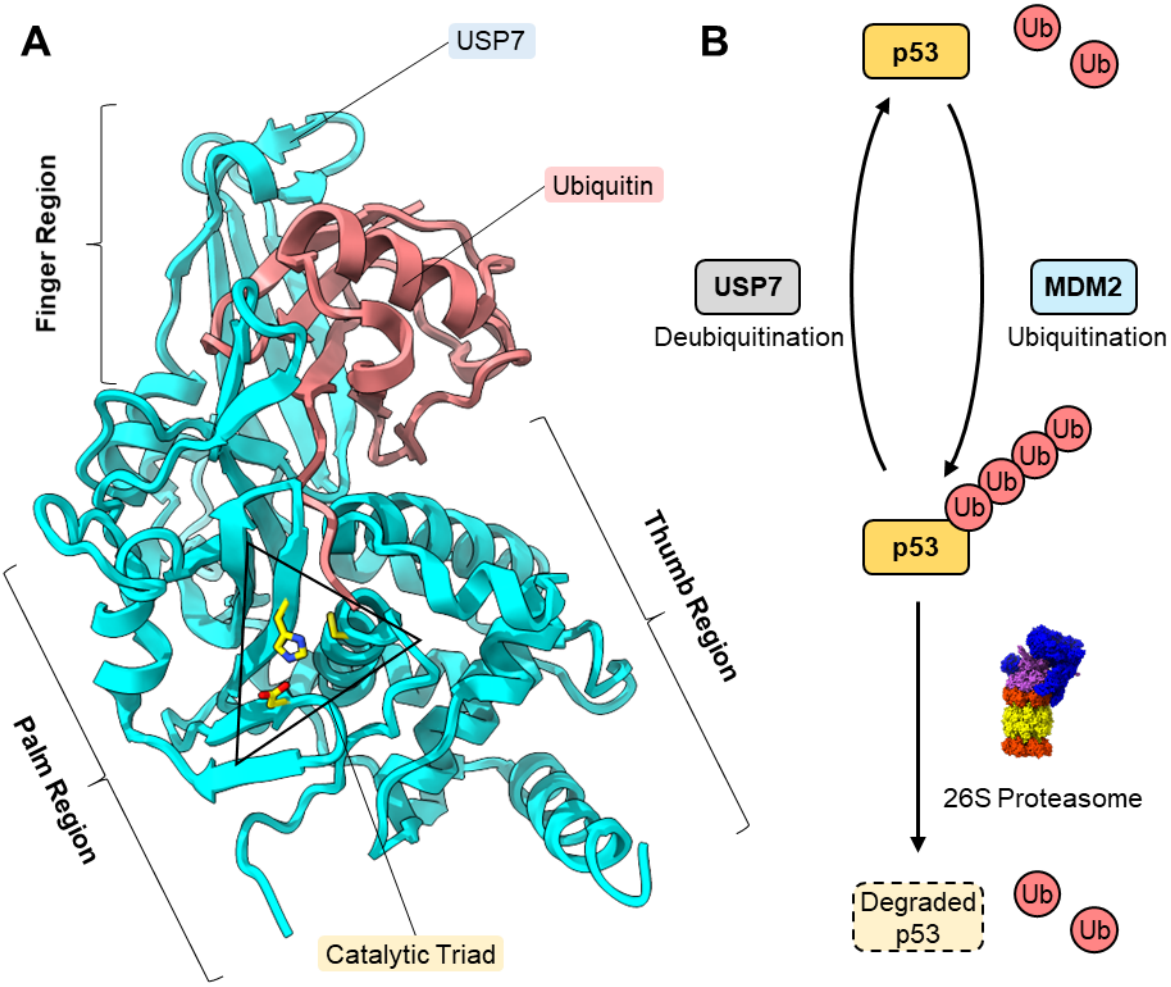


Figure 1.4: A) USP7 (cyan) in complex with Ubiquitin (salmon). Residues of in catalytic (C223, H464, and D481) are coloured yellow (PDB: 1NBF).[22] B) Regulation of p53 ubiquitination by E3 ligase MDM2 and DUB USP7.[23]

Like E3 ligases, DUBs are implicated in a broad range of cellular processes by limiting ubiquitination through the removal of ubiquitin from substrate proteins, effectively counteracting E3 ligase activity. In one part of a complicated regulation system of p53 activity, DUB USP7 is able to reverse the ubiquitination activity of the E3 ligase MDM2, and prevent degradation the tumour suppressor (Figure 1.4B).[24] Furthermore, DUBs ensure that there are sufficient levels of free Ub molecules within the cell by either liberating Ub from ubiquitinated substrates, or by breaking down newly synthesised ubiquitin polymers.[21,25]

1.3 Targeting E3 Ligases for Drug Development

Dysregulation of ubiquitination activity is implicated in a number of diseases and so E3 ligases are attractive targets for drug discovery. For example, an antagonist for the MDM2 E3 ligase (RG7712) was developed by Roche (Figure 1.5A).[23] As mentioned, this E3 ligase tightly regulates the levels of tumour suppressor p53 and is found to be overly expressed in leukaemia cells. Inhibition of MDM2 allows increased levels of p53 to be achieved and promote cell apoptosis. The small molecule inhibitor entered phase I clinical studies but the project was limited by gastrointestinal toxicity.[26] Given that most E3 ligases have several interaction partners, it might be difficult to target a specific E3 ligase/interactor complex without prompting off-target effects.

In recent years, a whole field of research has emerged where E3 ligases are exploited to facilitate targeted protein degradation (TPD).[27] This event-driven pharmacologically seeks to achieve therapeutic effects by depleting levels of the target protein, rather than inhibiting its function through a classical, occupancy-driven mode of action. TPD can minimises off-target effects, as the catalytic nature of this approach allows much lower (substoichiometric) treatment of the drug, and the precise ternary complex infers an additional degree of selectivity.[28,29] However, there still needs to be consideration of potential side effects caused by recruiting an E3 ligase for non-natural ubiquitination events.

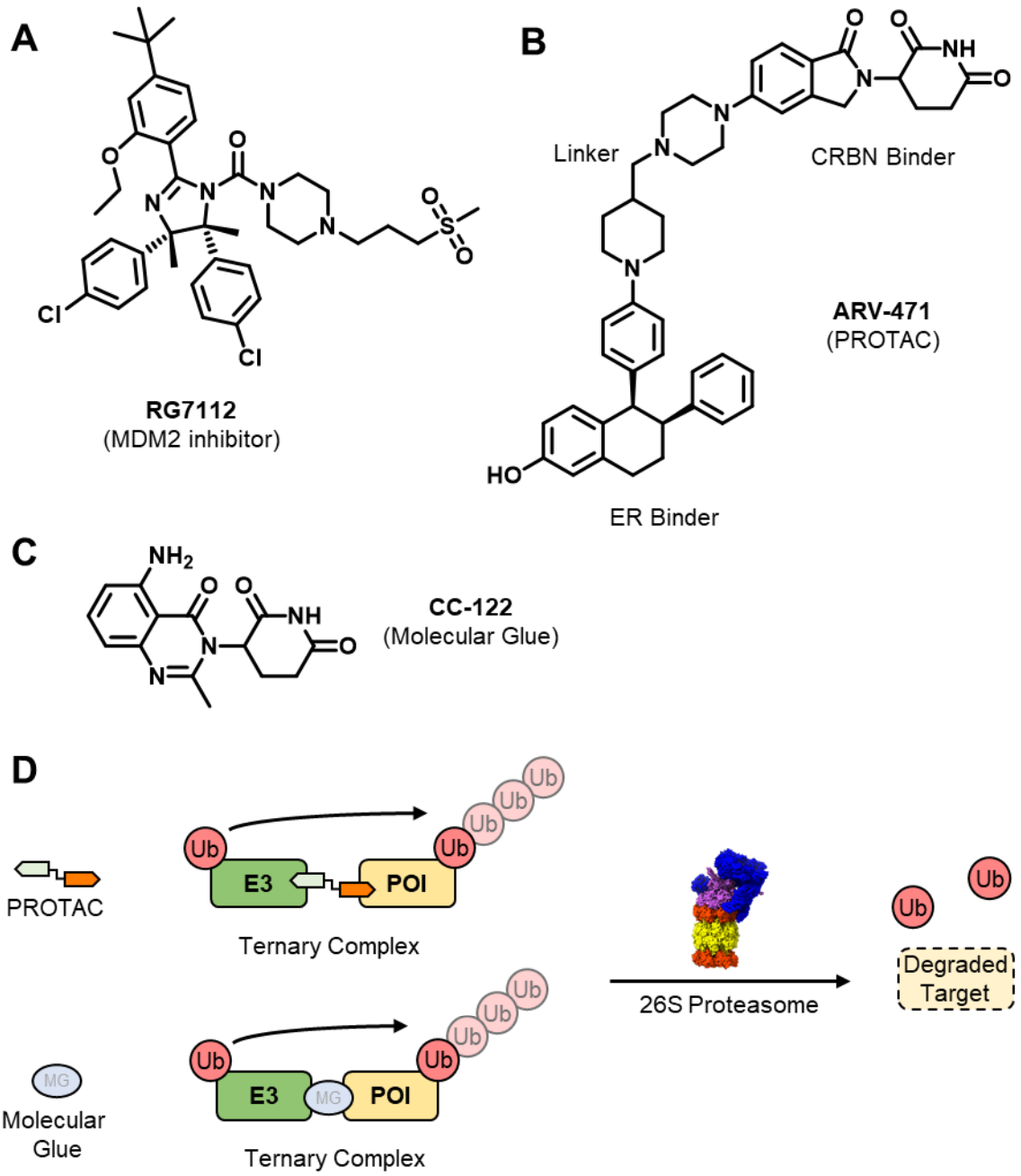


Figure 1.5: A) MDM2 antagonist RG7112. B) Estrogen receptor degrader ARV-471. C) PROTAC mediated targeted protein degradation.

Furthermore, if the degradation of the POI (protein of interest) occurs much more rapidly than its resynthesis, a disconnect between pharmacokinetics and pharmacodynamics can occur, which expands the therapeutic window and improves treatment efficacy.[30] Most excitingly, TPD has the potential to target proteins which may have been previously termed as 'undruggable'.[31,32] The binding site of the POI does not necessarily need to be the active site; it only requires formation of a suitable ternary complex with the E3 ligase. This creates the possibility of degrading targets with either shallow active sites, high-affinity interactors, or 'unblockable' scaffolding functions.

Proteolysis Targeting Chimeras (PROTAC) are hetero-bifunctional molecules that hijack the UPS by simultaneously binding an E3 ligase and a protein of interest (POI).[31] Ternary complex formation induces targeted ubiquitination of the POI, and its eventual degradation by the 26S proteasome (Figure 1.5D). The linkage of two protein binders results in large, complicated compounds with often poor physicochemical properties, which require large optimisation efforts in order to generate compounds with good solubility, cell permeability, and oral bioavailability.[33,34] These optimisation challenges are further compounded by the difficulty in predicting cooperative ternary complex formation; the driver of degradation efficiency.[35] Despite these challenges, several PROTACs are currently being developed, including orally bioavailable Vepdegestrant (ARV-471), which is jointly developed by Arvinas and Pfizer (Figure 1.5B). ARV-471 degrades the estrogen receptor (ER) and is currently in phase 3 clinical trials for the treatment of metastatic breast cancer.[36,37]

Molecular glues are monovalent molecules that also attempt to mediate unnatural ubiquitination events by facilitating non-native protein-protein interactions (PPIs) between an E3 ligase and a POI (Figure 1.5D).[38] Compared to PROTACs, these molecules tend to have better physicochemical properties due to their smaller size. However, molecular glues are often identified serendipitously, and the inability to predict neo-PPIs currently prevents the development of molecular glue degraders for a selected target protein.[39] Furthermore, once a chemical moiety is discovered, its potency can be very sensitive to alteration, making molecular glue optimisation particular challenging as well. Many molecular glues that have been identified are analogues of thalidomide, a compound which is able to bind to E3 ligase cereblon.[40] One such example is CC-122 (Avadomide) which is being developed by Salarius Pharmaceuticals to degrade hematopoietic transcription factors Aiolos and Ikaros, in order to promote antitumor and immunomodulatory activity (Figure 1.5C).[41]

One major shortcoming of TPD is that currently only a handful of E3 ligase small molecule binders have been reported, of which only a subset of these are frequently used, despite an estimated 600 E3 ligases that could be potentially used.[42] In recent years, there has been a concerted effort in the field to identify novel E3 ligase handles; including some work presented within the appendix chapter of this thesis. The hope is that by expanding the pool of E3 ligases that can be utilised for TPD, a range of E3 ligases with diverse characteristics will available. These E3 ligases could offer tissue specificity or be particularly suited to certain classes of POI, ultimately aiding in the development of more advanced degraders.

1.4 The KCTD Family

The potassium (**K**) **C**hannel **T**etramerization **D**omain family (KCTD) encompasses 25 members implicated in a broad portfolio of cellular processes.[43] Members are identified by possessing a conserved N-terminal short form BTB (bric-a-brac, tramtrack, broad complex) domain, and a highly variable C-terminal domain (CTD) (Figure 1.6). The BTB domain, which is sometimes referred to as the POZ (Poxvirus zinc finger) domain, is responsible for CUL3 recruitment, and oligomerisation. The short form version of this domain is homologous to that domains found in voltage-gated potassium (Kv) channel proteins, which self-assembles into a homo-tetrameric assembly to create a transmembrane channel.[44] Initially, it was thought that members of this family would behave in similar manner and have a similar function; hence why the family is named accordingly.

The CTDs are often responsible for interactor recognition and are quite distinct between members of this family.[45,46] Due to these variations, simple sequence classification of KCTD members were not that insightful. However, upon the advent of Alphafold2, L. Esposito *et al.* conducted a systematic structural survey of the KCTD family.[47] The researchers generated models for all family members (at the oligomeric state inferred by their BTB domain) and generated a 'structure-based pseudo-phylogenetic tree' by calculating similarity Z-scores of the CTD using the Dali server (Figure 1.6).[48,49] Molecular dynamics simulations of representative models in each cluster was used confirm the reliability of the Alphafold2 models.

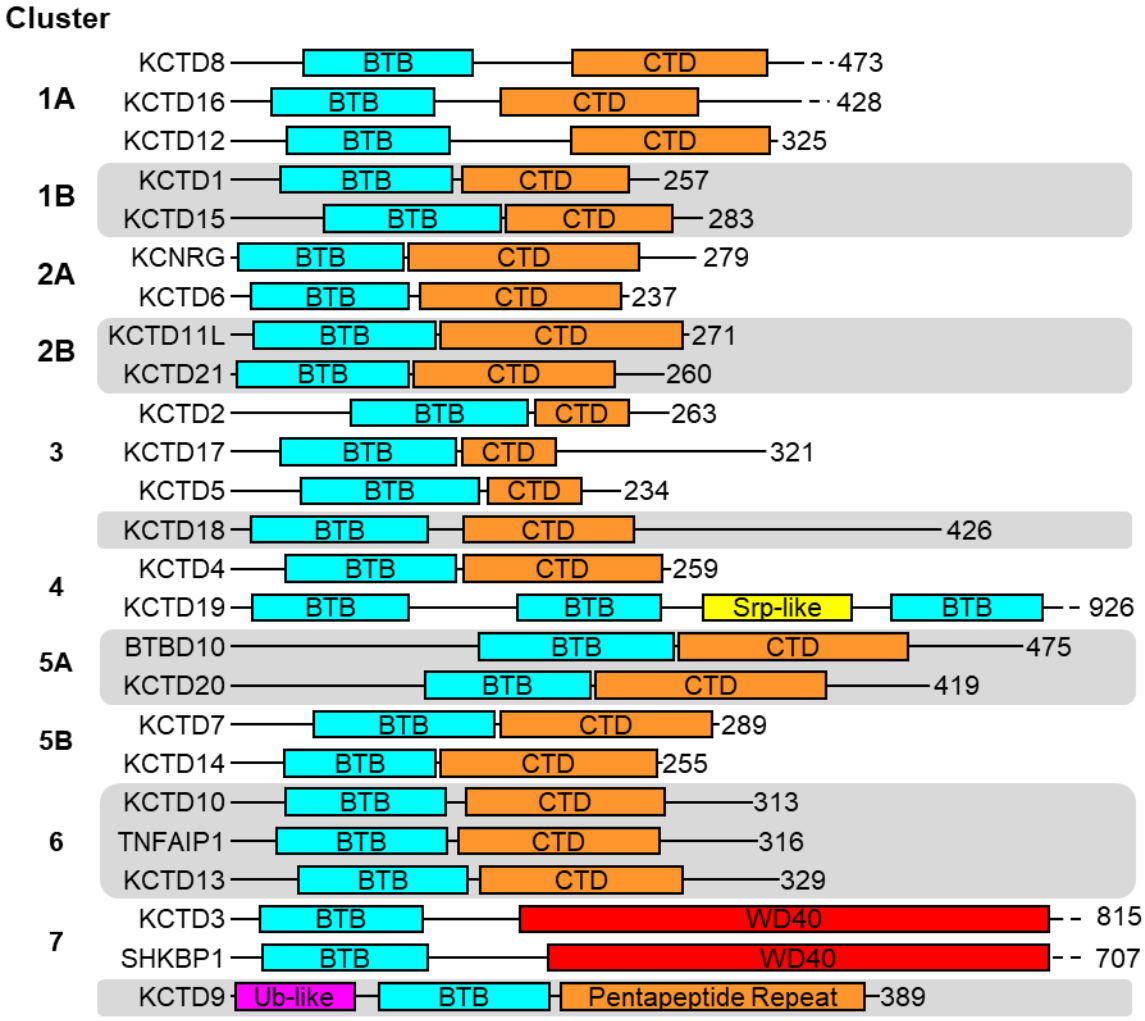


Figure 1.6: KCTD family summary. Structured-based clustering as reported by L. Esposito *et al.*[48]

Cluster 1-4 members saw all of the CTD adopt a five-point β-propeller-like structure (Figure 1.7A-D). The main variations between the clusters was seen at the loop sections between the β-sheets. Cluster 1A members had no predicted BTB/CTD interactions and as a result of contain a longer flexible linker region between the two domains. The models generated for KCTD18 and KCTD19 were deemed too unreliable by the authors to interpret critically. Models of cluster 5A members did not form reliable pentameric assemblies and this instead suggested that these members may adopt an open multimer state, where the BTB domain interacts with the CTD of an adjacent KCTD unit instead. A much larger cavity was predicted for cluster 5B members (Figure 1.7E) and dimeric models of cluster

6 members had higher confidence scores than pentameric models (Figure 1.7F). As the CTD of Cluster 7 members were predicted to have a large eight-bladed β-propeller domain, their limited computing resources prevented the authors from modelling a pentameric state for KCTD3 and SHKBP1. The researchers classified cluster 7 members and KCTD9 as non-canonical members of the KCTD family.[50]

The Goldstein group reported the first structure of this family, KCTD5$^{34-234}$ in high and low salt conditions (PDB: 3DRX and PDB: 3DRY respectively) (Figure 1.8).[51] These structures revealed that KCTD5 adopted a homopentameric state, around a five-fold axis, that is enabled by BTB surface interactions. The CTD domain was positioned above the BTB domains and separated by a short flexible linker region that allowed a near ~30° swivel between them. Since then, structures of human KCTD proteins have shown them adopting a range of monomeric, tetrameric, open/closed pentameric, and hexameric states (Figure 1.8, Table 1.2).

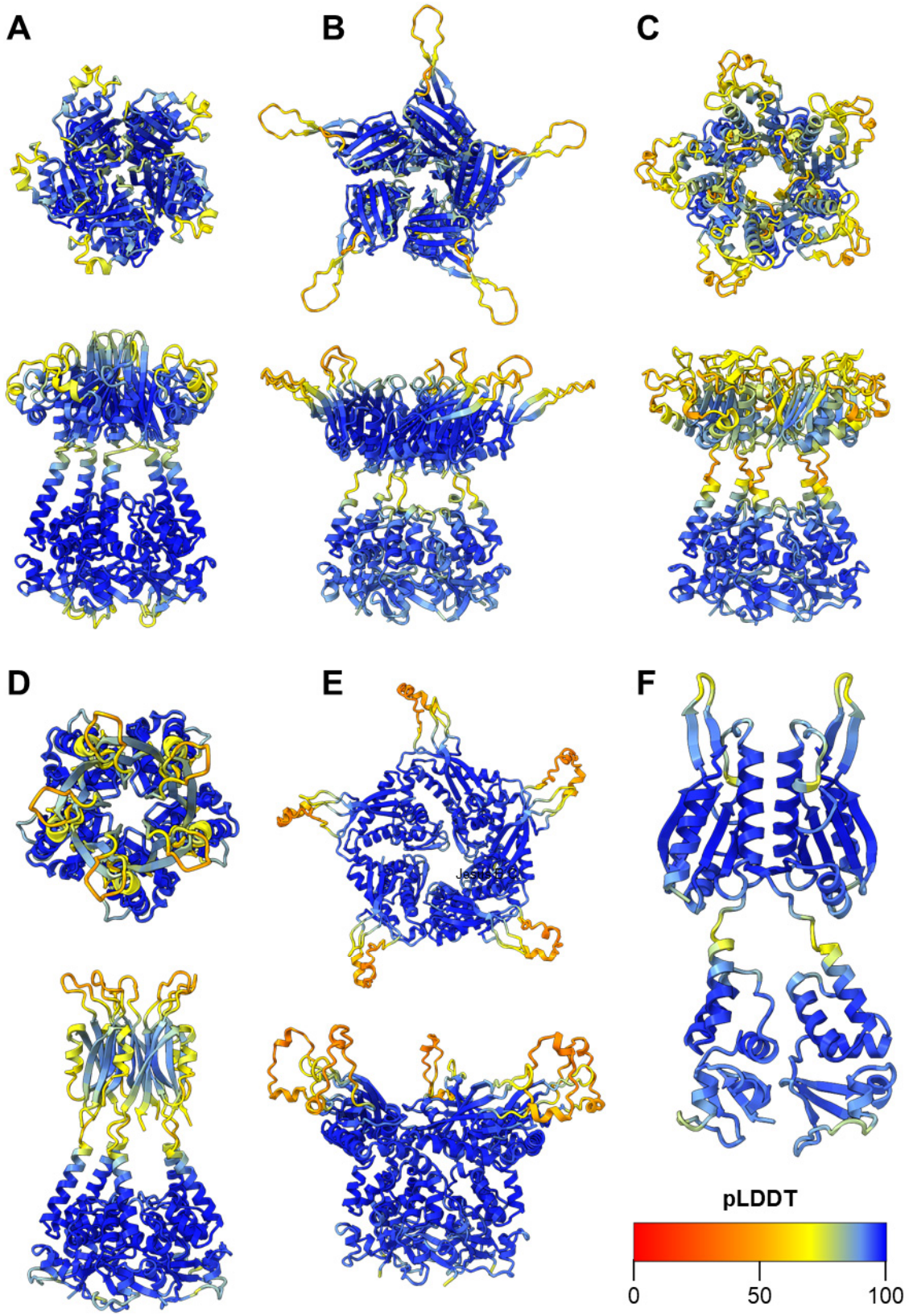


Figure 1.7: Top and side views of KCTD Alphafold2 models. Coloured by pLDDT score. A) KCTD15 pentamer (cluster 1B) B) KCNRG pentamer (cluster 2A). C) KCTD4 pentamer (cluster 2B). D) KCTD2 pentamer (cluster 3). E) KCTD7 pentamer (cluster 5B). F) KCTD10 dimer (cluster 6).

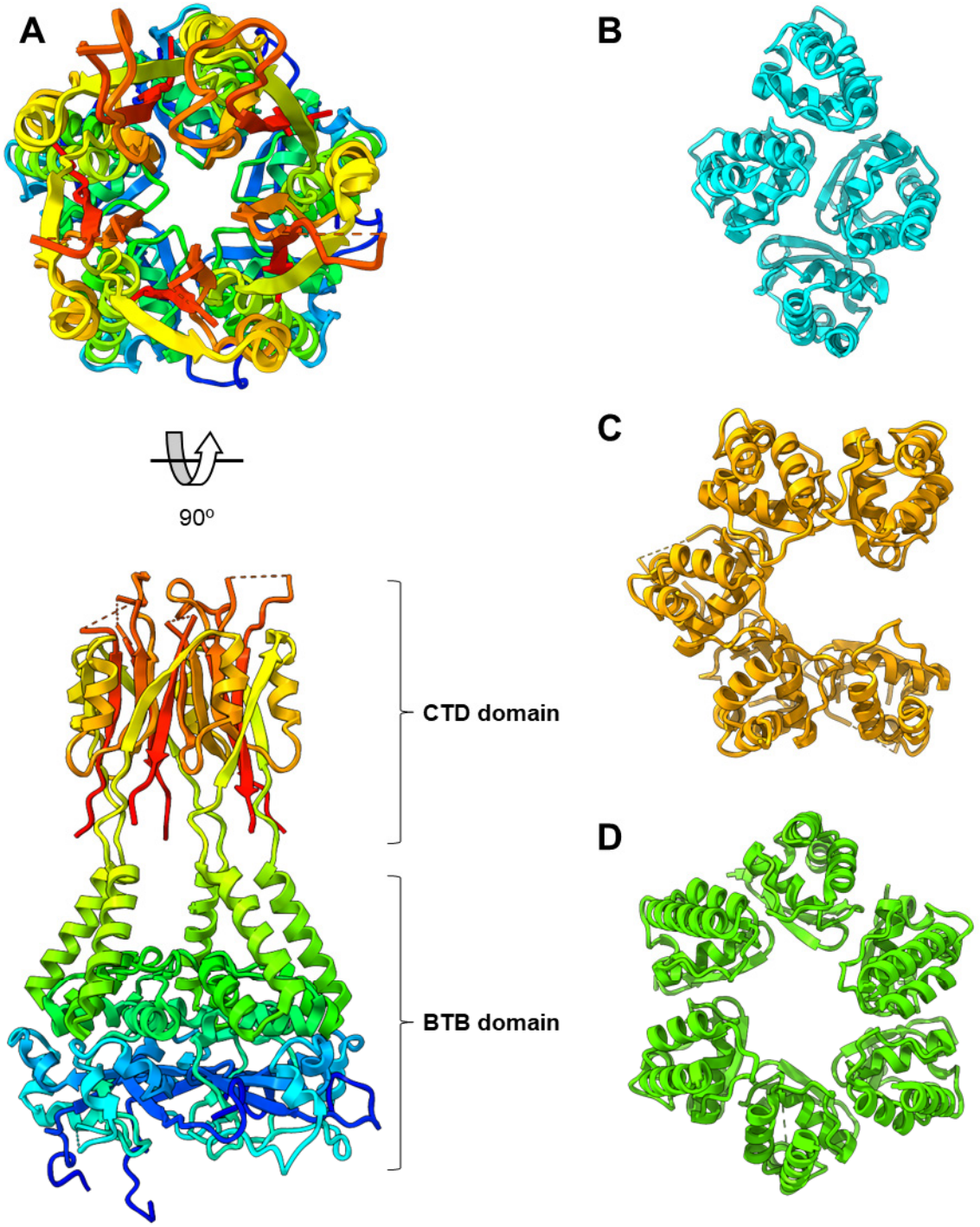


Figure 1.8: A) Closed pentameric KCTD5$^{34\text{-}234}$ in high-salt buffer (PDB: 3DRX).[51] Rainbow coloured from N (blue) to C (red) terminus. B) Tetrameric KCTD10 (PDB: 5FTA).[52] C) Open pentameric KCTD16 (PDB: 6OCP).[53] D) Hexameric KCTD15^{G88D} (PDB: 8PNM).[54]

Protein	Region	Oligomeric state	PDB Identifier
KCTD1	29-132	Open and closed pentamer	5BXB,[55] 5BXD,[55] etc.
KCTD5	34-234	Closed pentamer	3DRX,[51] 3DRY, etc.
KCTD7	1-289	Closed pentamer	8I79[45]
KCTD8	201-322	Dimer (H1 domain)	6G57
KCTD9	89-191	Closed pentamer	5BXH[55]
KCTD10	26-135	Tetramer	5FTA[52]
KCTD12	202-325	Closed pentamer	6M8S,[56] 6QZL
KCTD13	27-144	Monomer	4UIJ[52]
KCTD15^{G88D}	51-165	Hexamer	8PNM,[54] 8PNR[54]
KCTD16	23-123, 126-286	Open and closed pentamer	6OCP,[53] 6QB7, etc.
KCTD17	13-124	Closed pentamer	5A6R[52]
SHKBP1	18-120	Monomer	4CRH[52]

Table 1.2: Summary of previously reported KCTD structures.

KCTD proteins are implicated in striking range of unrelated cellular processes through ubiquitination and non-ubiquitination activities. KCTD5 is one example that is implicated in regulating the protein kinase B (AKT) signalling pathway.[57] The CTD of the E3 ligase recognises disassociated Gβγ (G-protein β and γ) dimers to facilitate their ubiquitination and proteasomal degradation; thereby preventing signal transmission. Additionally, this KCTD family member has also been shown to mono-ubiquitinate ΔNp63α (an isoform of p63 that is also a member of the p53 gene family) and impair its ability to bind DNA, in order to prevent transcription activation.[58]

KCTD6, KCTD11, and KCTD21 have all been demonstrated to mediate the ubiquitination, and subsequent degradation of histone deacetylase 1 (HDAC1).[59] HDAC1 deacetylates transcription factors Gli1 and Gli2 which promote Hedgehog signalling and dysregulation of this signalling pathway is the leading cause of medulloblastoma tumorigenesis. As a result, these KCTD members are found to be downregulated in malignant brain tumour cells.

KCTD1 is a transcription regulator that interacts with AP-2α transcription factors (TFAP2A, TFAP2B, and TFAP2C) through its non-CUL3 binding BTB domain.[60] Missense mutations of the BTB domain prevent TFAP2A repression during embryogenesis, and this causes Scalp-Ear-Nipple (SEN) syndrome. As KCTD1 is unable to recruit CUL3, it is an example of a KCTD family member that serves regulatory functions without serving as an E3 ligase. Other KCTD members have also been implicated in neurodevelopment, natural killer cell development, and progression of various cancers, although many of their specific interactors are still unknown.[61-63]

1.4.1 KCTD9 is a Non-Canonical Family Member

KCTD9 is 389 residue (42.5 kDa) protein that stands out from the rest of KCTD family. While it contains the conserved short BTB domain that defines this group, it also possesses an additional N-terminal Ubiquitin-like domain, and a CTD with uncommon β-solenoid pentapeptide repeat domain (PPR) (Figure 1.9A).[50]

The structure of KCTD9^{BTB} (PDB: 5BXH, Figure 1.9B) was first reported by G. Privé *et al.* in a study that investigated KCTD assembly and CUL3 recruitment.[55] KCTD9^{BTB} was shown to adopt a closed pentameric state with 'near-perfect' 5-fold symmetry and is similar to the KCTD5^{BTB} structure (PDB: 3DRZ).[51] The oligomeric state was confirmed by analytical size-exclusion chromatography (SEC) and a glutaraldehyde cross-linking experiment (Figure 1.9C). SDS-PAGE after the treatment revealed bands up to and beyond the pentameric state, though these additional bands are likely to be crosslinked aggregates and not representative of an in-solution oligomeric state.

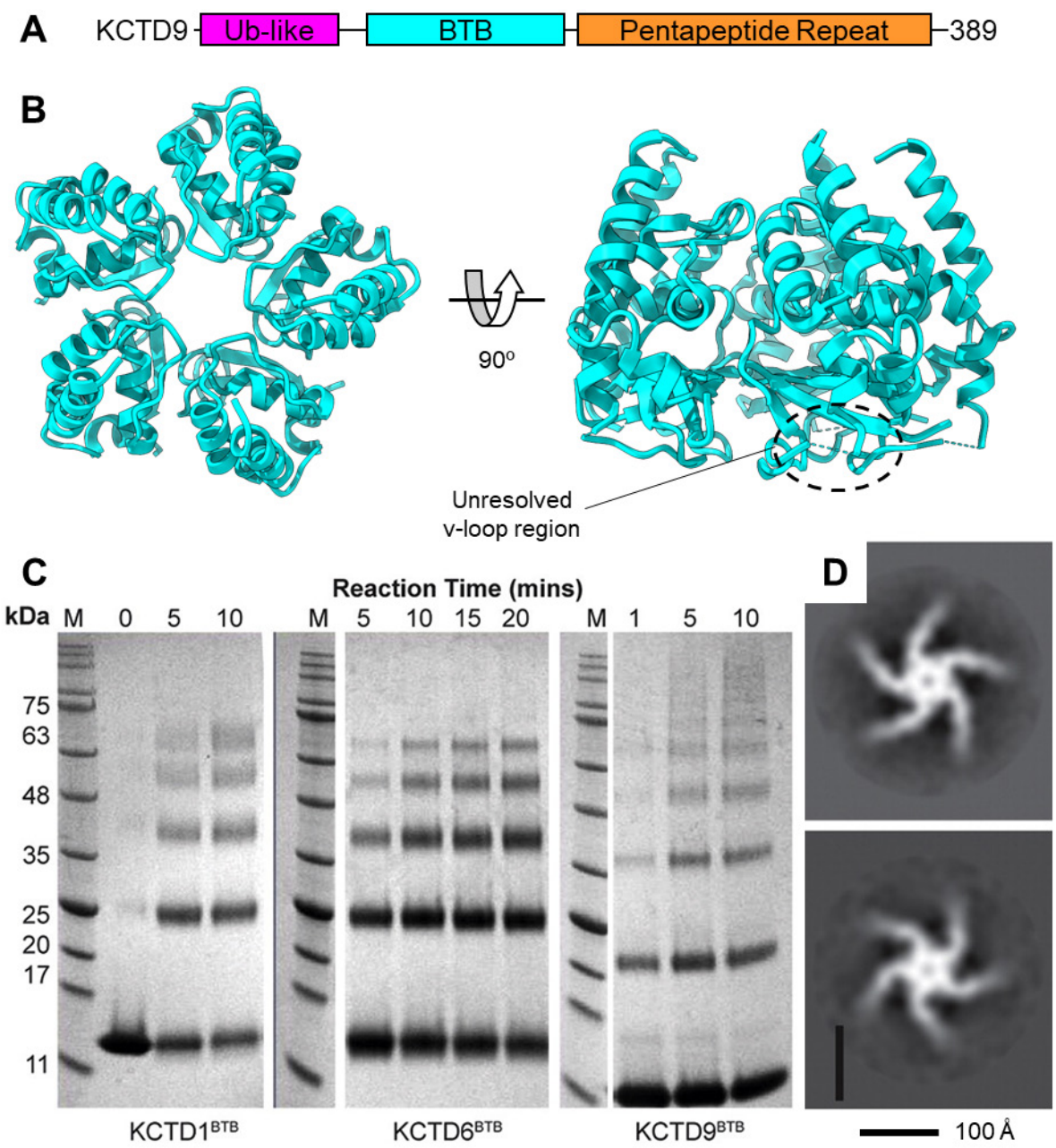


Figure 1.9: A) KCTD9 domain organisation. B) KCTD9^{BTB} Structure (PDB: 5BXH). C) Data published by G. Privé *et al.* KCTD9 glutaraldehyde cross-linking experiment (left). Reactions quenched at 1,5 and 10 minutes. 2D cryo-EM images of KCTD9^{BTB}/CUL3$^{1-381}$ (C5 symmetry applied, right).[55]

Nanomolar binding (K_d = 11 nM) of CUL3 to the KCTD9^{BTB} domain was measured by isothermal titration calorimetry (ITC) and 2D single-particle cryo-EM images indicated that the complex formed a distinctive five-point pin-wheel complex (Figure 1.9D). The researchers generated a 5:5 KCTD9^{BTB}/CUL3$^{31-381}$ model using the reported pentameric crystal structure and the 2:2 KLHL3/CUL3 dataset (PDB: 4HXI), which resembles a more recently published KCTD5/CUL3 structure (PDB:8U80, Figure 1.10A.).[64] From these models, two distinct sets of interactions were predicted between KCTD9 and CUL3. The primary interface of KCTD9 (α2/v-loop/β3 region and helix α4) contacts helices h2 and h5 of CUL3, in a manner analogous to other BTB/CUL3 complexes; where the weakly positive CUL3 surface is complementary to the negative charged surface of the BTB domain (Figure 1.10A). At the other interaction surface (the distal interface), KCTD9 engages CUL3 in a manner that is not observed by other CUL3 substrate receptors. While the BACK domain of BTB-KELCH proteins participate in further interactions with helices h2 and h5 of CUL3, the distal surface of KCTD9 instead engages the helix h1 and the alternative side of helix h2 (Figure 1.11).

The researchers identified two elements of the KCTD protein which are critical for CUL3 binding and defined an indicator for whether a KCTD family member can recruit CUL3.[55] Firstly, the electrostatic surface of the primary surface must complement the weakly positive charged CUL3 helices. KCTD1, KCTD16 and other KCTD family members that are not able to recruit CUL3 have positively charged or hydrophobic residues along the primary surface (Figure 1.10C), instead of a negative charged surface present in KCTD5 and KCTD9. Secondly, the variable loop region (v-loop) between α2 and β3 of BTB domain must dynamic. Examining long-form BTB (KLHL/SPOP) proteins, these proteins have

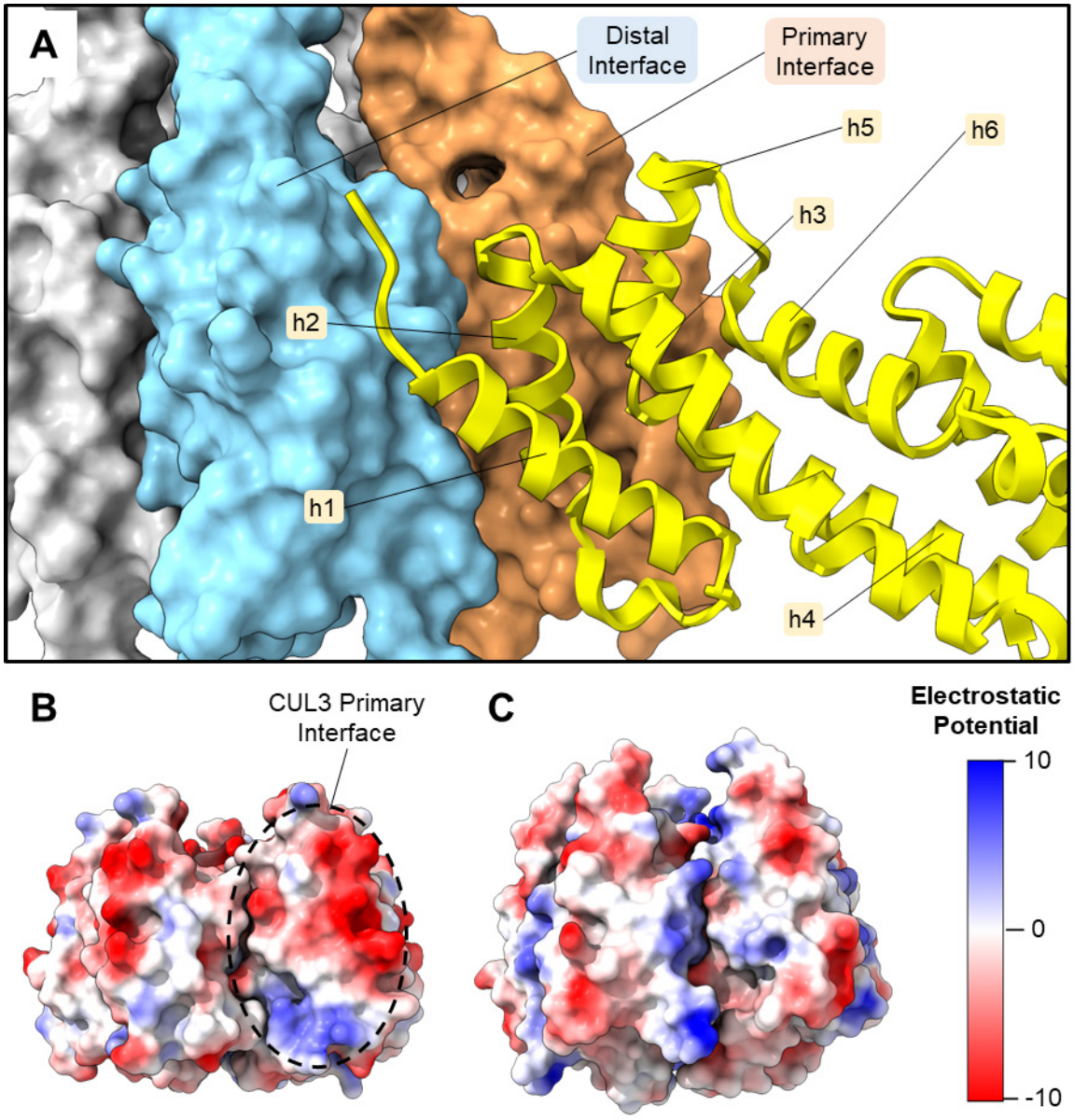


Figure 1.10: A) KCTD5/CUL3 complex (PDB: 8U80).[46] Primary and distal interfaces are coloured orange and cyan respectively. CUL3 is coloured yellow. B) Surface of KCTD9^{BTB} (PDB: 5BXH) coloured by electrostatic potential.[55] C) Surface of KCTD1^{BTB} (PDB: 5BXH) coloured by electrostatic potential.[55]

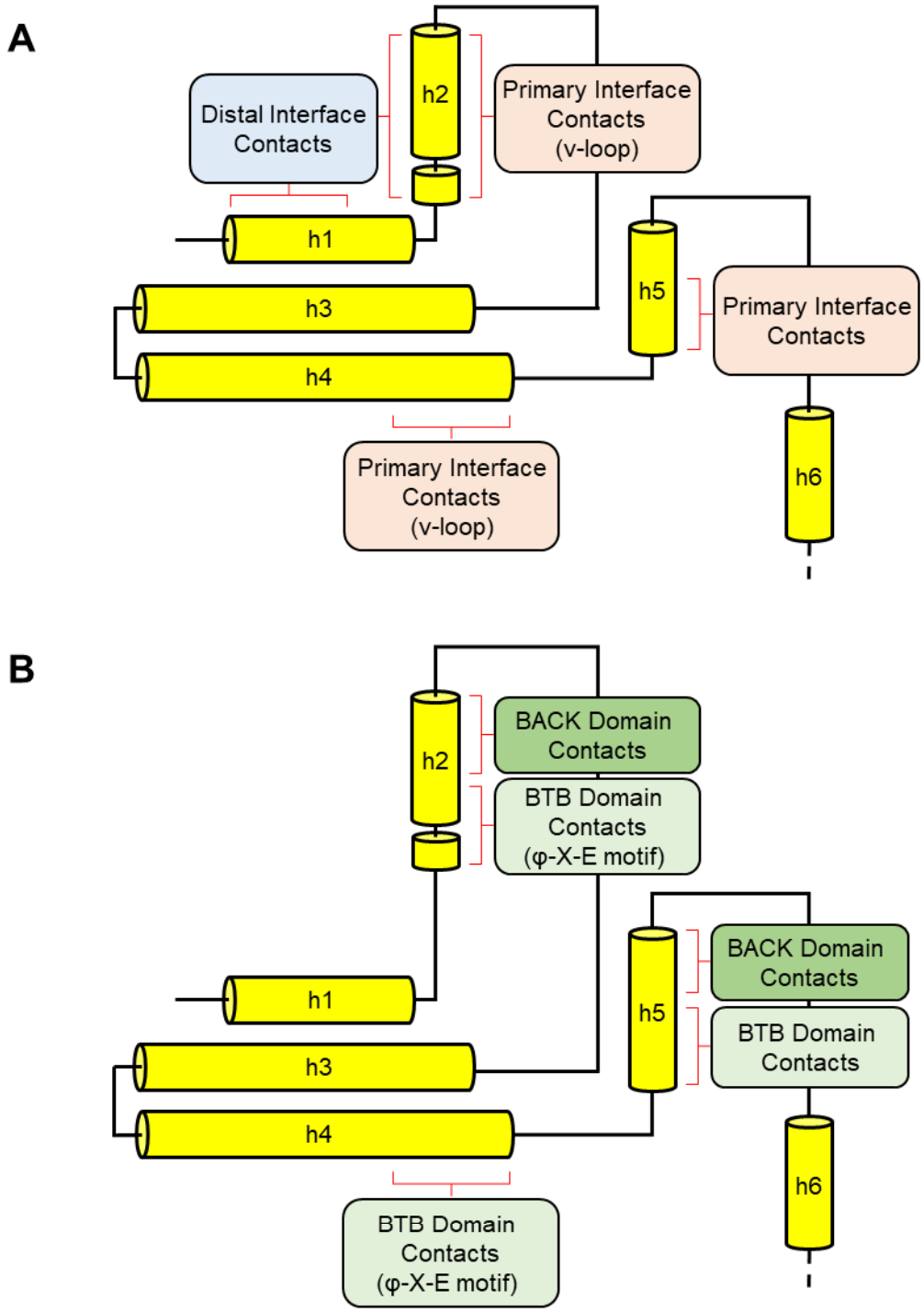


Figure 1.11: A) 2D topology diagram of CUL3 helices 1-6 (not to scale). Approximate contact regions with KCTD5, as shown in Figure 1.10 (PDB: 8U80), are shown. B) Contacts of KLHL3/CUL3.

a flexible φ-X-E motif which is responsible for interactions with helices h2 and h4 of CUL3, and becomes ordered upon binding to CUL3. While this φ-X-E motif is not present in KCTD proteins, CUL3 binding KCTD members have a flexible v-loop region that can participate in interactions with the same regions of CUL3 instead. In the case of KCTD1, this region is too rigid to adopt the complementary conformer for CUL3 binding. A KCTD9^{V125A} mutation of the v-loop region demonstrated partially disrupted CUL3 affinity (supported by ITC and SEC data), and the researchers postulated that this mutation hindered the ability of the v-loop to adopt the optimal conformation for binding.

The C-terminal pentapeptide repeat (PPR) domain not only makes KCTD9 distinct from the KCTD family, but also distinct in the human proteome. The PPR motif was first identified in Hglk, a cyanobacterium protein reported to regulate glycolipid localisation.[65,66] Initially, a simple AXLXX consensus repeat motif was suggested to represent this fold, but that motif has undergone several refinements to arrive at the currently recognised five-residue PPR motif (Figure 1.12B):

(A/C/S/V/T/L/I) *I* (D/N/S/K/E/I/R) *I* (L/F) *I* (S/T/R/E/Q/K/V/D) *I* (G/D/E/N/R/Q/K)

In October 2024, a search through Pfam (a protein families and domains database) for sequences that coded 8 copies of the PPR motif (PF00805) identified approximately 72,000 examples in ~15,000 taxa.[67] Around 83% of these sequences are found in bacteria, with 21,509 sequences found in cyanobacteria alone. Of these, only 30 structures with this motif have been deposited to the PDB and most remarkably the only protein found in *Homo sapiens* with this domain classification is KCTD9 (Uniprot: Q7L273).

M. tuberculosis MfpA was the first structure with a PPR domain to be reported and this structure displayed the domain to adopt a four-sided, right-handed, β-solenoid structure (PDB: 2BM4, Figure 1.12A).[68,69] A four-sided unit is made up of 20 residues and referred to as a coil (Figure 1.12C). The side chains of first (A/C/S/V/T/L/I) and third (L/F) residues of the motif contribute to the hydrophobic core of the solenoid structure, while the remaining residues in the motif are outward facing to confer the domains surface charge. Neighbouring coils sit almost directly on top one another (with only a slight left-handed helical turn) and stacks of hydrophobic sidechains preclude water from the central structure. Finally, these solenoid structures are often bookended by α-helices capping motif, or another PPR domain, to prevent exposure of the hydrophobic centre to solvent.

The MfpA expression has been reported to promote antibiotic resistance of fluoroquinoline compounds.[68] Normally these compounds will bind type II topoisomerases (DNA gyrases), which enable DNA replication and transcription by facilitating negative supercoiling (unwinding of DNA). The fluoroquinoline compounds inhibits DNA gyrase activity, causing a stall in DNA replication and ultimately promote cell death. The shape, size, and surface charge of the MfpA duplex is similar to B-form DNA (Figure 1.12A), and the structure is able to act as a T-segment DNA mimic. MfpA is able to bind to DNA gyrase in place of fluroquinolines without stalling DNA replication and this competitive binding results in reduced antibiotic efficacy.

Lastly, the N-terminal ubiquitin-like domain is another unique feature that KCTD9 possess in relation to the KCTD family.[50] None of the other members have an additional N-terminal domain prior to the BTB domain and the role of this extra domain is unknown function as there has been no prior investigation.

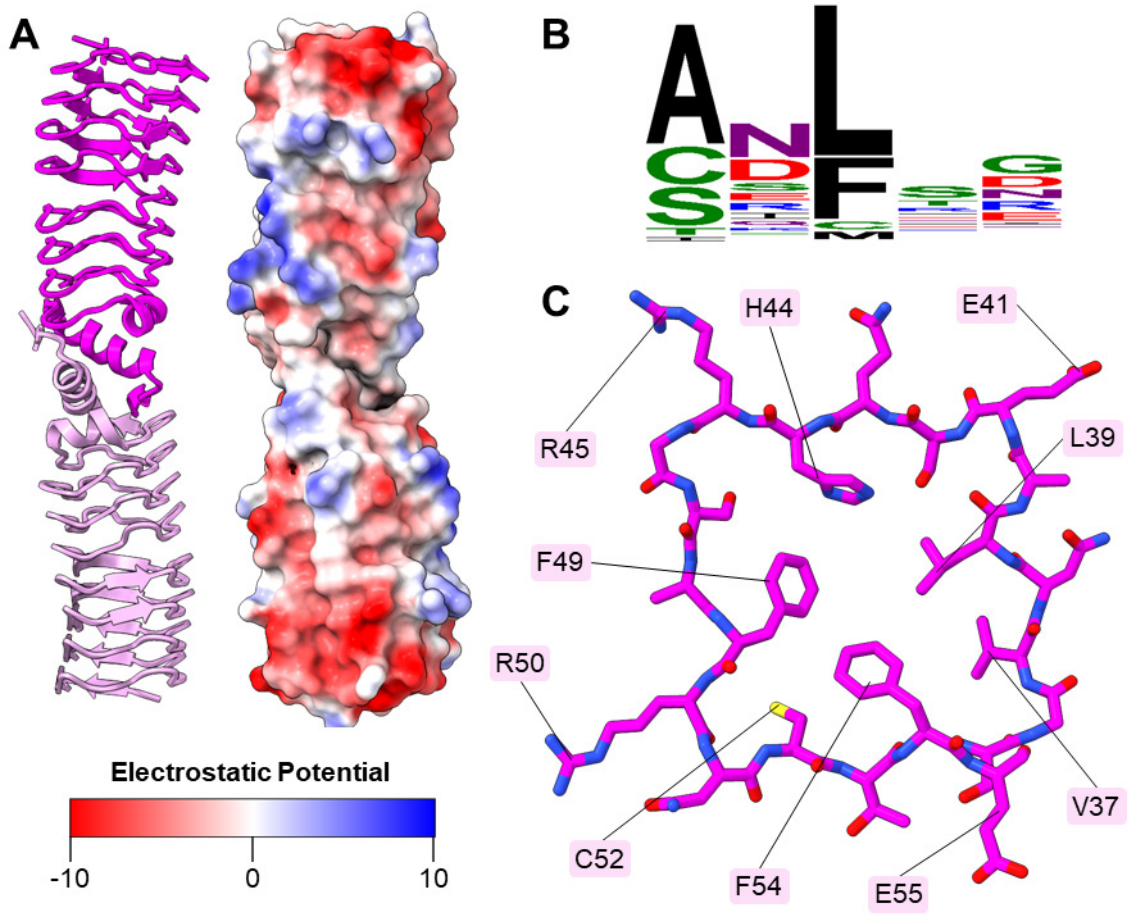


Figure 1.12: A) MfpA duplex (PDB: 2BM4) in cartoon (left) and surface representations (right).[69] Surface coloured by electrostatic potential. B) Sequence logo of the pentapeptide repeat motif using sequences identified in Pfam search (PF00805). Logo generated by WebLogo3.[70] C) A single coil of MfpA (S35-E55). Hydrophobic core residues, and corner residues are labelled.

1.4.2 KCTD9 is Implicated Natural Killer Cell Development

Natural killer (NK) cells are part of the innate immune system and the primary response to virus-infected cells. One study by T. Chen *et al.* compared blood and liver samples derived from patients with mild chronic hepatitis B (CHB) or hepatitis B virus-induced acute-on-chronic liver failure (HBV-ACLF).[71] The researchers demonstrated that KCTD9 expression was upregulated in peripheral blood mononuclear cells (PMBCs) and hepatic NK cells from HBC-ACLF patients. Additionally, KCTD9 was demonstrated to have been translocated from the cytoplasm to the nucleus by confocal microscopy analysis. Elevated KCTD9 levels (induced by HBV infection) promoted NK cell activation and increased cytotoxicity, which resulted in more liver injury and worse disease progression. The researchers later published data from murine hepatitis virus strain 3 (MHV-3) mouse models, which supported their initial observations.[72] In addition, they demonstrated that the effects of increased KCTD9 expression can be reversed with silencing RNA (siRNA).

In a final study, X. Zhang *et al.* sought to understand role KCTD9 plays in natural killer cell development and effector function using KCTD9$^{-/-}$ knockout mice.[63,72] These model species displayed reduced liver damage and increase survival rates when infect with MHV-3 in comparison to the wild-type control group. The researchers also demonstrated that KCTD9 regulates NK cell commitment, maturation and restricts cell proliferation. Furthermore, loss of KCTD9 impaired effector function and downregulated expression of transcription factors required for NK cell maturation and development (NFIL3, ETS1, Eomes and ID2). Despite all these investigations, no direct interactor of KCTD9 that is related to NK cell development has yet been identified.

1.4.3 KCTD9 Regulates Wnt Signalling

In recent years, H. Yao *et al.* identified that KCTD9 was consistently down regulated in several colorectal cancer (CRC) tissue and cell lines.[73] KCTD9 was shown to inhibit CRC cell proliferation and patients with high levels of KCTD9 expression had better survival outcomes. After gathering evidence that KCTD9 was inhibiting β-catenin/Wnt signalling in CRC, the researchers sought to identify the interactor of this E3 ligase by a mass spectrometry affinity proteomics experiment in HCT-116 cells. They managed to identify and confirm ZNT9 (zinc transporter 9), a binding partner of β-catenin, as a binding interactor of KCTD9. Interestingly, KCTD9 regulates the levels of β-catenin but in a manner independent of ubiquitination activity; instead, KCTD9 competes with β-catenin for binding of ZNT9. The KCTD9/ZNT9 interaction leaves β-catenin exposed to ubiquitination (by an unknown E3 ligase, potentially β-TrCP), and subsequent degradation (Figure 1.13B). The researchers showed that $KCTD9^{CTD}$ was able to bind to $ZNT9^{1-200}$ by immunoprecipitation (IP) and western blotting, but no structural datasets detailing these interactions have been generated thus far.

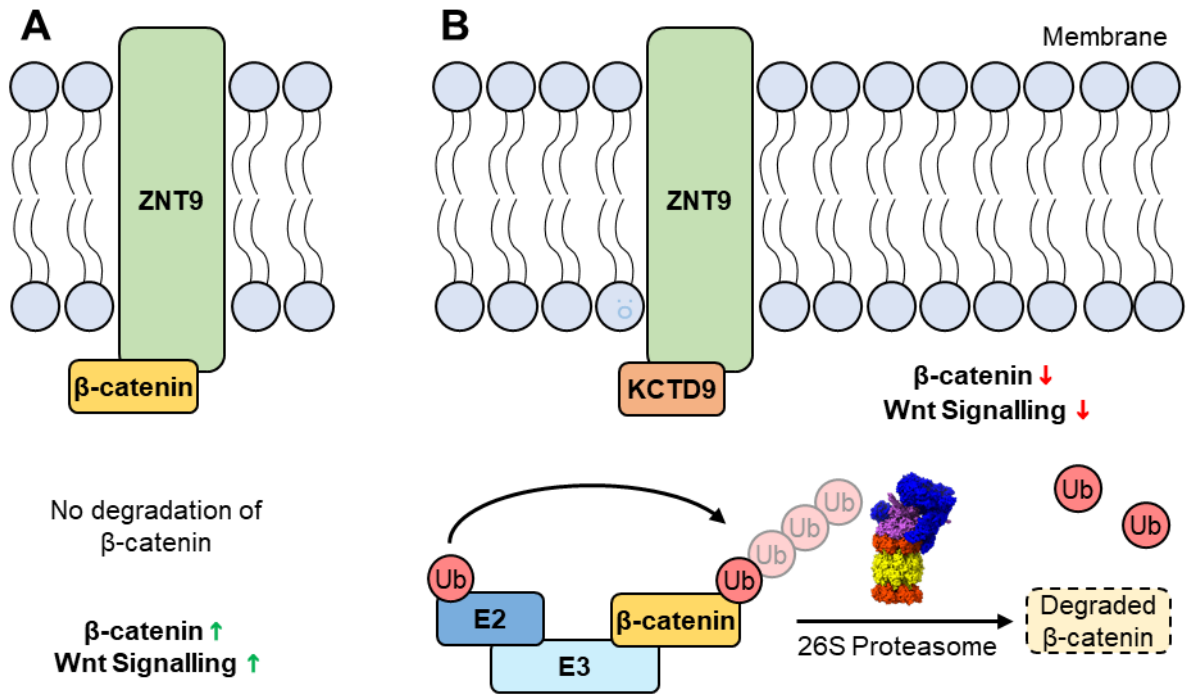


Figure 1.13: A) ZNT9/β-catenin stabilises β-catenin and promotes Wnt signalling. B) ZNT9/KCTD9 interaction leaves β-catenin exposed for ubiquitination and subsequent degradation.

1.5 The KLHL Family

The Kelch superfamily comprises 63 proteins that can be subdivided in three distinct groups: 42 KLHL (**K**el**c**h-**l**ike) type, 9 KBTBD (**K**elch repeat and **BTB** **d**omain-containing protein) type, and 12 KLHDC (**K**el**c**h **d**omain-**c**ontaining protein) type.[74] KLHL members are structured in a similar manner to the KCTD family, with an N-terminal long form BTB-BACK region and a C-terminal kelch repeat domain for substrate recognition. The KLHL family BTB domain is comparable in to the short form version seen in the KCTD family, but has an extra α-helix and β-sheet (α1 and β1) which contributes to dimeric interactions. This domain is flanked by a **B**TB **a**nd **C**-terminal **K**elch (BACK) region (Figure 1.14A), which contains a 3-box motif that participates in interaction with helices h2 and h5 of CUL3 (Figure 1.11B).[64,75-77]

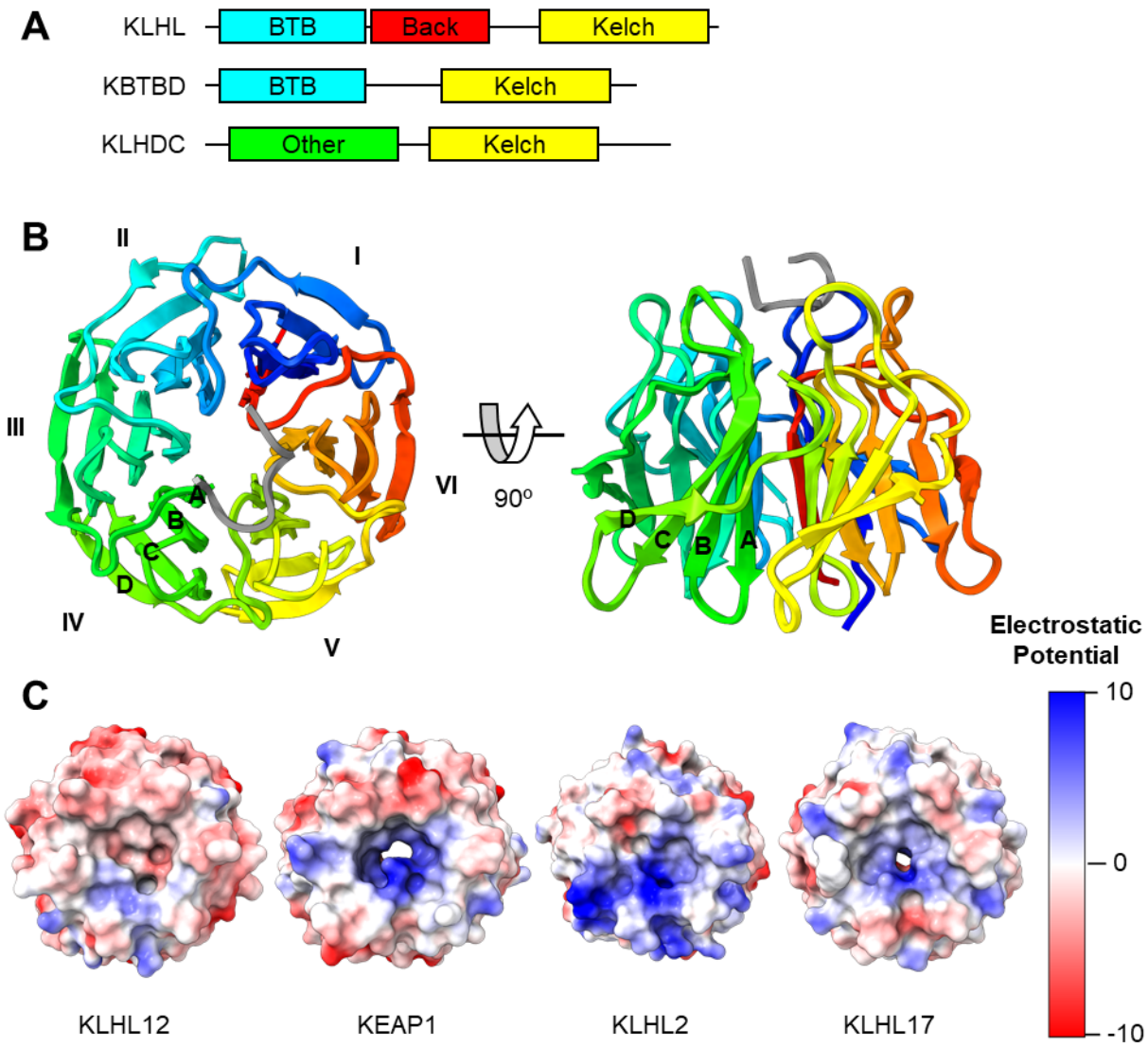


Figure 1.14: A) Domain organisation of the different kelch families. B) KLHL12 (rainbow) in complex with DVL1 peptide (grey, PDB: 6TTK). The KLHL12 kelch domain is made up of six blades (I-VI) in a β-propeller fold. The four β-strands of Blade I are labelled A-D (PDB: 6TTK). C) Comparison of the kelch pockets of several KLHL proteins (coloured by electrostatic potential). Left to right: KLHL12 (PDB: 6TTK),[78] KEAP1 (PDB: 1U6D),[79] KLHL2 (PDB: 2XN4),[80] and KLHL17 (PDB: 6HRL).

The Kelch domain adopts a six bladed β-propeller fold (I-VI), where each blade is made up of four twisted antiparallel β-strands (A-D), to resemble a goblet-like structure; hence the name kelch as it is German for goblet (Figure 1.14B).[74,78] Interactors bind at centre of this goblet structure and often forms interactions with all blades of the kelch domain. The sequence of the kelch domain varies across the family to give rise very different inner surface profiles (Figure 1.14C) which allow this family to recognise a wide range of interactors in a selective manner.

KBTBD members do not often have the BACK region. The KLHDC family have members which have alternative N-terminal domains (NTD) such as glycine rich regions (KLHDC10), or have no NTD whatsoever.[74]

1.5.1 KLHL12 Has Many Interaction Partners

Kel**c**h-**l**ike protein 12 is an E3 ligase substrate adaptor that is implicated in the regulation of many different systems and has a broad list of interactors. It regulates canonical and non-canonical Wnt signalling by ubiquitinating dishevelled proteins (DVL1-3), which results in the substrates being degraded by the 26S proteasome (Figure 1.15A).[81] The depletion of dishevelled proteins causes a reduction in classical and non-classical Wnt signalling by preventing β-catenin stabilisation and DAAM1/DVL1-3 interactions.[82,83] KLHL12 can also associate with co-adaptors PEF1/ALG2 to selectively recruit and mono-ubiquitinate SEC31, in order to promote COPII complex formation.[84] It is postulated that a dimeric KLHL12 complex recruits the complex at one of the kelch domain, and SEC31 is cooperatively recruited to the other (Figure 1.15B).[85] A recent study demonstrated that PEF1 is also monoubiquitinated to stabilise co-adaptor/substrate assembly.[85] Furthermore, PLEKHA4 and Lunapark have both been observed to sequester KLHL12 to plasma or ER membrane respectively (Figure 1.15C).[86] These antagonists prevent the E3 ligase form recruiting substrates and serve as a mechanism by the cell to regulate ubiquitination activity. Finally, dopamine D4 receptor is a KLHL12 substrate that is subjected to different types of ubiquitination, including non-lysine residues.[87]

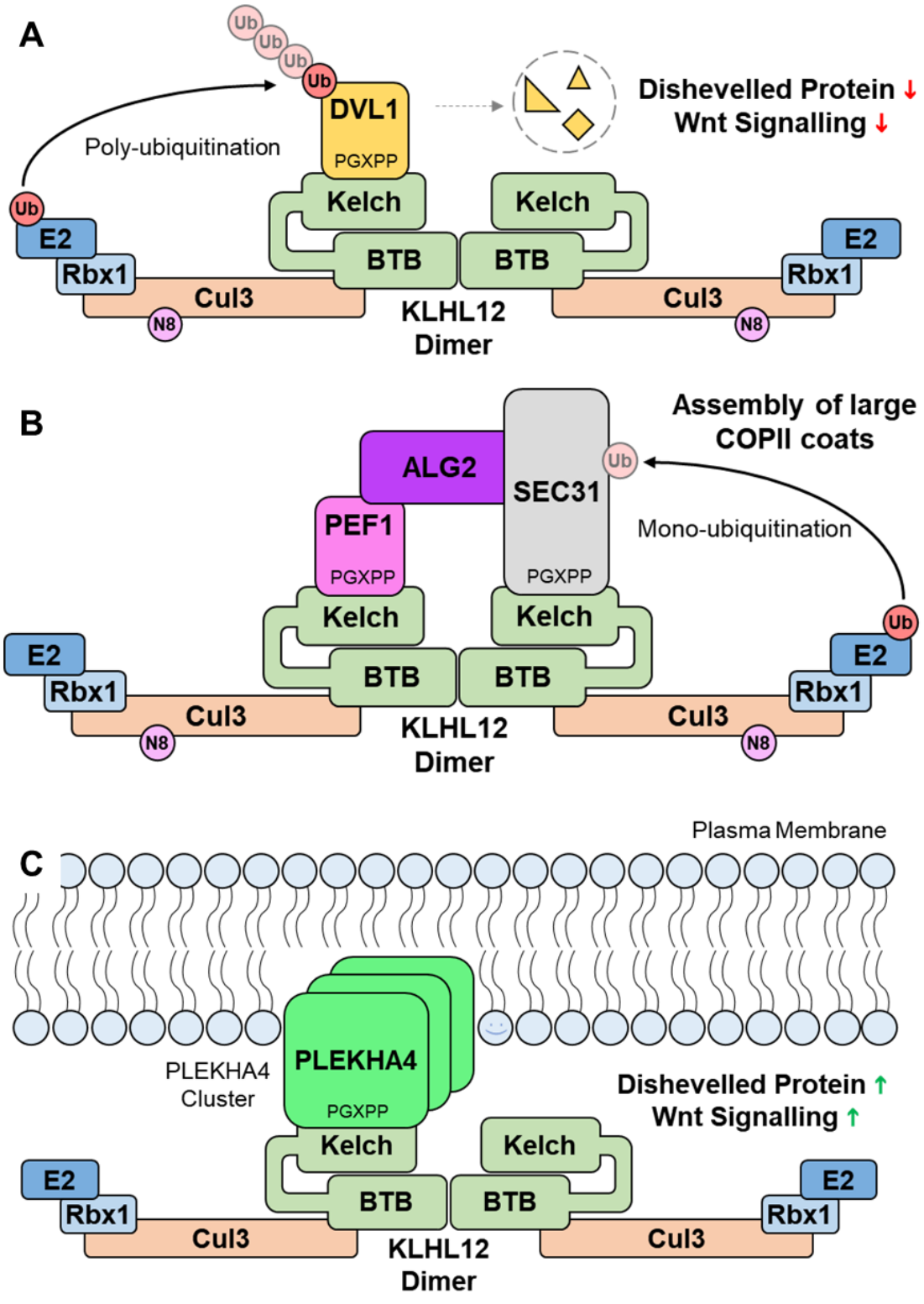


Figure 1.15: KLHL12 recognises a variety of partners with different functions through the same PGXPP recognition motif. A) DVL1-3 is a ubiquitinated by KLHL12 and ultimately degraded.[81] B) Co-adaptor PEF1 and ALG2 help recruit SEC31 to a dimeric KLHL12 complex for mono-ubiquitination.[84] C) PLEKHA4 sequesters KLHL12 to the plasma membrane and stops the ubiquitination activity of the E3 ligase.[86]

The Bullock group identified a consensus PGXPP degron motif in the interactors that bind to KLHL12 by a SPOT peptide array experiment.[78] The motif is present in all of the interactors listed above and the researchers generated a $KLHL12^{KELCH}$/DVL1 15-mer peptide co-crystal structure (PDB: 6TTK) to determine the structural mechanisms of interactor recognition. The peptide adopted a U-turn conformation within the kelch pocket, and a fluorescence polarization assay measured micromolar (K_D: 22 μM) binding affinity to KLHL12.

1.6 Aims of this Thesis

KCTD9 is a non-canonical KCTD family member with an uncommon PPR domain and unique Ub-like NTD.[50] There is only structural data for the pentameric BTB domain, and I aimed to generate additional datasets that cover the remaining regions. I therefore needed to clone new KCTD9 constructs, evaluate their expression yields, and optimise a purification protocol in order to generate samples that could be used to generate either X-ray crystallography or cryo-EM structural datasets. Additionally, potential interactors identified in the literature, or through my own experiments, could complexed with these KCTD9 constructs. Datasets of these complexes would allow me to determine the structural mechanisms that underpin KCTD9 interactor recognition and this information could serve as a foundation for future efforts in manipulating KCTD9 interactions.

Another aim of the thesis was to identify novel interactors of KCTD9 and determine its cellular role. At the start of this project, no binding partners of KCTD9 had been reported, but several detailed studies had implicated it in NK cell development and effector function; yet no direct interactor of KCTD9 had been determined.[63] Identifying these unknown interacting partner(s) in this role is

exciting because of the potential developments in cancer immunotherapy. In recent years, NK cells and other components of the immune system have garnered attention for their potential uses in immune-oncology (IO), where therapies seek to stimulate the body's natural immune system in order to identify and eliminate cancerous cells.[88] Some strategies seek to passively enhance the ability of the immune system by using immune checkpoint inhibitors, or immune system modulators. Other, more active methods, remove the immune cells from a patient, modify these cells to recognise the cancerous cells, and then reintroduce them back to the patient with a tumour.

Given that KCTD9 loss resulted in reduced NK cell development and effector function, identifying its interacting partner and understanding the pathway that it is a part of could generate a selection of new targets to be investigated; for the development of new immunotherapies using the strategies previously mention. To address this, affinity mass spectrometry proteomics was chosen to conduct this study, as this method is incredibly sensitive and can survey a wide coverage of the proteome.[89]

Finally, I also investigated the PGXPP recognition motif (identified by the Bullock group) that is common in multiple interactors of KLHL12. This E3 ligase is a rare example that recognises a common recognition motif that is shared between interactors implicated in different roles (substrate, co-factor, antagonist). I wanted to confirm that the U-turn binding conformer seen in the KLHL12$^{KELCH/}$DVL1 peptide structure was representative of the binding mode for other interactors.[78] Additionally, exploring the relationship between interactor sequence and KLHL12 binding affinity may provide overall insights into the many regulatory roles of this E3 ligase.

Chapter 2: Method and Materials

2.0 Methods and Materials

2.1 Cloning

2.1.1 Construct Design

KCTD9, FBXO31, SKP1, ZNT9, and TRIM constructs were designed with the help of protein disorder prediction tools (such as PSIPRED),[90] AlphaFold modelling,[47] and published literature.[91] KCTD9, ZNT9, and TRIM cDNA were sourced from the Mammalian Gene Collection (MGC).[92] FBXO31 and SKP1 cDNA were codon optimised and synthesised by GenScript.

The majority of constructs used were cloned through a high-throughput sequence and ligation-independent cloning method (SLIC).[93] 3XFLAG KCTD9 was produced by restriction enzyme cloning, and KCTD9 mutants by restriction-free cloning.[94]

2.1.2 Sequence and Ligation-Independent Cloning

Sequence and Ligation-independent cloning has been utilised by the SGC to develop a protocol where multiple targets and constructs can be cloned in parallel. It utilises the proofreading ability of T4 DNA polymerase to ‘chew-back’ along the backbone vector, in the presence of a single dNTP, in order to create large (~15 bp) overhangs. A DNA insert is generated by PCR of the gene DNA with the addition of appropriate SLIC extensions by primer design. The insert is treated in a similar manner as the backbone vector to produce complementary overhangs that can be annealed together. This method requires no ligation enzyme as the product is transformed into competent bacterial cells and be repaired by cell machinery instead.[93]

2.1.2.1 PCR and Vector Digestion

PCR primers were designed using a macro in an excel template written by Dr Jon Elkins in our department and then synthesised by Eurofins. The default varies for this macro were used:

- Minimum Primer Length = 32
- Minimum Primer T_m = 60°C
- Minimum Primer GC = 35%
- Maximum Primer GC = 75%
- Salt Concentration = 50 mM
- Primer Concentration = 0.5 μM

A 96-well primer plate containing 5 μM of the appropriate primer in each well and an entry clone plate with gene DNA at 2.5 ng/μL were prepared. The contents of each plate was transferred to a reaction plate containing Herculase II Fusion DNA polymerase (20 units, Agilent, 600675), Herculase II buffer, and 10 mM dNTP.

The reaction plate was sealed and reacted in a thermocycler with the following cycles (touch-down PCR):

- 95°C for 10 minutes.
- (95°C for 30 seconds; 68°C for 30 seconds; 68°C for 1-3 minute*) for 5 cycles.
- (95°C for 30 seconds; 60°C for 30 seconds; 68°C for 1-3 minute*) for 5 cycles.
- (95°C for 30 seconds; 55°C for 30 seconds; 68°C for 1-3 minute*) for 5 cycles.
- (95°C for 30 seconds; 50°C for 30 seconds; 68°C for 1-3 minute*) for 20 cycles.
- 68°C for 10 minutes.
- 4°C on hold.

*Extension time dependent on length of PCR product (30 seconds per 1 kb).

The PCR products were loaded onto a 1.5% agarose SYBR safe (Invitrogen, S33102) DNA gel and imaged with a ChemiDoc MP imager. Failed reactions where no band at the expected size was observed were optimised by the addition of DMSO (0-8% final concentration), GC-enhancers (20% final concentration,

NEB, B9028) or alternative thermocycler conditions. If the template DNA shared the same antibiotic resistance as the new backbone vector, the corresponding PCR product was treated with DpnI (20 units, NEB, R0176) to digest any of the original template DNA which will be methylated. The PCR products were purified using the PureLink^{TM} PCR purification kit (Invitrogen, K310001).

The backbone vectors were digested with the appropriate enzyme (such as BsaI). Digestion was analysed by 1.5% agarose gel and, where complete digestion had occurred, the product was purified by the PCR purification kit. For instances where incomplete digestion had occurred, the product was purified by QIAquick gel extraction (QIAGEN, 28704) using an 0.8% agarose gel.

2.1.2.2 T4 Polymerase Treatment and Annealing

The PCR products and digest vectors were separately treated with T4 DNA polymerase (0.075 units, NEB, M0203S), NEBuffer 2.1, and the appropriate dNTP (25 mM). The mixtures were incubated in a thermocycler with the following cycle:

- 22°C for 30 minutes.
- 75°C for 20 minutes.
- 4°C on hold.

2 µL of the T4 treated vector was added to 4 µL of the insert and incubated at room temperature for an hour at minimum. The annealed products were then transformed into chemically competent MACH1 cells (50 µL) and incubated on ice for 30 minutes. The transformant mixtures were ‘heat shocked’ at 42°C for 45 seconds, SOC media (100 µL) was added, and the mixtures were incubated at 37°C for 1.5 hours. The transformed cells were spread onto LB agar plates +5% sucrose containing the appropriate antibiotic and incubated overnight at 37°C.

2.1.2.3 Colony Screening

A 96-well PCR reaction plate containing MyTaq^{TM}Red (Bioline, BIO-250423) and appropriate sequencing primers (0.5 µM) was prepared. In addition, a 96-deep-well block containing 1 mL of LB media and the appropriate antibiotic was made. Successful colonies were picked with a sterile loop, briefly mixed in a PCR reaction plate well, and then use to inoculate the corresponding well of the 96-deep-well block. The block was sealed with a porous membrane and incubated overnight at 37°C with shaking at 700 rpm.

The PCR reaction plate was placed in a thermocycler and heated with the following conditions:

- 95°C for 10 minutes.
- (95°C for 30 seconds; 50°C for 30 seconds; 72°C for 1-3 minute*) for 25 cycles.
- 72°C for 5 minutes.
- 4°C hold.

*Extension time dependent on length of PCR product (30 seconds per 1 kb).

The PCR reactions was analysed by 1.5% agarose gel to see if there was successful incorporation of the insert into the vector. For the successful clones, the overnight culture (30 µL) was mixed with 60% glycerol (120 µL) and stored at -80°C.

2.1.3 Restriction Enzyme Cloning (REC)

This cloning method utilises a number of restriction enzymes that recognise specific recognition sites to create short complementary sticky ends (1-4 bp) on the vector and insert. The sticky ends are annealed together, covalently linked by a T4 ligase and transformed into competent bacterial cells. Unlike SLIC, the choice of backbone vector and restriction enzyme that can used is sometimes restricted by the composition of the template DNA.

2.1.3.1 Method

Primers for the PCR of the KCTD9 gene was designed using SnapGene and synthesised by Eurofins. The method outlined in **2.1.2.2** was used for the PCR and product purification.

Purified PCR products (1 µg) were digested with SalI-HF (20 units, NEB, R3138) and BglII (10 units, NEB, R0144) in NEBuffer r3.1 (NEB, B6003). The pCMV 3XFLAG vector was sourced from an alternative construct within the group. The vector (5 ug) was digested with NotI-HF (50 units, NEB, R0189) and PspOMI (50 units, NEB, R0653) in rCutSmart buffer (NEB, B6004). Digestion was analysed by 1.5% agarose gel and the products were purified by QIAquick gel extraction (QIAGEN, 28704).

A molar ratio of 1:3 vector/insert using 50 ng of insert was treated with T4 DNA ligase (40 units, NEB, M0202) in T4 DNA ligase buffer. The mixture was incubated overnight at 16°C and the enzyme was then inactivated by heating at 65°C for 10 minutes. The reaction mixture (2 µL) was transformed into MACH1 cells (50 µL) using the method outlined in **2.1.2.3**. The transformation mixture was spread onto LB agar plates containing the appropriate antibiotic and successful colonies were analysed using the method outlined in **2.1.2.4**.

DNA of a successful 3XFLAG KCTD9 clone was obtained using a Qiagen miniprep kit (Qiagen, 27104). The DNA sequence was confirmed using the Source BioScience Sanger sequence service.

2.1.4 Restriction Free Cloning (RF)

Restriction-free cloning is a two-step cloning method that can be used for the insertion of gene into a backbone vector using a high-fidelity polymerase and without the use of a restriction enzyme. A pair of hybrid primers is designed which contain a complementary sequence to both the gene DNA and the backbone vector. A primary PCR with the primers and the gene DNA generates a pair of 'mega-primers'. These mega-primers are then used in a secondary PCR with the backbone vector and high-fidelity polymerase to synthesise DNA along the whole vector. The reaction mixture is then treated with DpnI to digest any of the original (methylated) DNA sequences and the reaction mixture is transformed into competent bacterial cells.[94]

2.1.4.1 Method

Primers for generation of KCTD9 mutants were designed using www.rf-cloning.org/ and synthesised by Eurofins. The method outlined in **2.1.2.2** was used for the primary PCR and megaprimer purification.

A molar ratio of 1:20 vector/megaprimer (100 ng of vector) was added to Herculase II Fusion DNA polymerase (0.4 units, Agilent, 600675), Herculase II buffer, and 200 μM dNTP. The secondary PCR mixture was placed in a thermocycler and heated with the following conditions:

- 95°C for 30 seconds.
- (95°C for 8 seconds; 72°C for 1-3 minute*) for 15 cycles.
- 72°C for 5 minutes.
- 4°C hold.

*Extension time dependent on length of vector product (30 seconds per 1 kb).

The complete reaction mixture was transformed into MACH1 cells (50 µL) using the method outlined in **2.1.2.3**. The transformation mixture was spread onto LB agar plates containing the appropriate antibiotic; either kanamycin (50 µg/mL), chloramphenicol (100 µg/mL), streptomycin (50 µg/mL), and/or ampicillin (100 µg/mL).

Successful colonies were analysed using the method outlined in **2.1.2.4**. DNA of successful KCTD9 mutants were obtained using Qiagen miniprep kit (Qiagen, 27104). The DNA sequence was confirmed using Source BioScience Sanger sequence service.

2.1.5 Bacmid Generation

For insect cell expression, the Bac-to-Bac® baculovirus expression system was employed. This method involves the first cloning the gene of interest into a pFastBac donor vector in *E. coli* cells. The modified donor vector is then transformed into DH10Bac cells and the gene of interest is transposed into a large (>100 kbp) bacmid vector. Blue/white colony screening is used to identify the clones where transposition has successfully occurred and the purified recombinant bacmid DNA is then transfected into *Sf9* cells to generate baculovirus particles. These viruses can be amplified and used to infect *Sf9* insect cells for protein production.[95]

2.1.5.1 Blue/White Colony Screening

DNA of successful constructs was isolated using a Qiagen miniprep kit (Qiagen, 27104). The DNA (~550 ng) was transformed into chemically competent DH10Bac cells (30 µL) following the method outlined in **2.1.2.3**. SOC medium (900 µL) containing tetracycline (10 ug/mL), gentamycin (1 ug/mL), and the

appropriate antibiotic to the host vector was added to cells. The transformant mixture was incubated at 37°C with shaking at 700 rpm for 4-5 h and then 50 µL was spread onto LB-agar plates containing gentamycin (7 µg/mL), tetracycline (10 µg/mL), isopropyl β-D-1-thiogalactopyranoside (IPTG) (40 µg/mL), bluo-gal (100 µg/mL, ThermoFisher, 15519028), and the appropriate antibiotic. The plates were incubated at 37°C for 48 hours. White colonies were picked with a sterile loop and streaked onto fresh LB-agar plates. Successful colonies were analysed by colony PCR.

2.1.5.2 Transfection

Sf9 cells ($2x10^5$ cells/mL, 1 mL per well) in mid-log growth phase were seeded into 24-well flat TC plates (Costar, 3524), incubated at 27°C for 1 hour, and the media then removed. DNA of Bacmids was obtained using a Qiagen miniprep kit (Qiagen, 27014). The DNA (0.5-2.0 µg/µL) was added to *Sf9*-900™ II SFM medium (40 µL, Gibco, 10902088) containing GeneJuice (2 µL, Sigma-Aldrich, 707967), the mixture was incubated at room temperature for 30 minutes, and then Sf-900 II SFM medium (160 µL) was added. The GeneJuice mixture was added to the *Sf9* cells, and the plate was incubated 27°C for 4 hours. Sf-900 II SFM medium (400 µL) containing 2% FBS (Gibco, A5256801) was added to the cells and the plate was incubated at 27°C for 72 hours. The P0 virus was harvested by centrifuging the culture at 1,500 xg for 20 minutes, and then stored at 4°C.

2.1.5.3 Viral Amplification

Virus stock (0.5 mL per 100 mL of culture) was added to *Sf9* cells ($2x10^6$ cells/mL) containing 2% FBS. The culture was incubated at 27°C with shaking at 100 rpm for 48-72 hours. The amplified virus was harvested, by centrifuging the culture at 900 xg for 15 minutes, and then stored at 4°C.

2.1.6 Competent Cells Preparation

Competent bacterial cells were prepared to facilitate the uptake of foreign DNA more readily. Chemically competent cells are prepared by the treatment of calcium chloride at a high concentration to produce pores in the cell membrane. Heat-shock of these treated cells temporarily opens these pores and allows DNA to enter.[96]

2.1.6.1 Buffers and Media

SOB Media:

- 20 g/L Tryptone
- 5 g/L yeast extract
- 0.5 g/L NaCl
- 10 mL of 250 mM KCl (18.6 g per 100 mL)

The mixture was made up to 1L with de-ionized water and autoclaved. Once autoclaved, filtered $MgCl_2.6H_2O$ (2 M, 10 mL) was added.

SOC Media:

- 20 g/L Tryptone
- 5 g/L yeast extract
- 0.5 g/L NaCl
- 10 mL of 250 mM KCl (18.6 g per 100 mL)

The mixture was made up to 1L with de-ionized water and autoclaved. Once autoclaved, filtered $MgCl_2.6H_2O$ (2M, 10 mL), and filtered glucose (1M, 20 mL) was added.

TFB1:

- 2.945 g/L Potassium acetate
- 12.092 g/L RbCl
- 1.470 g/L $CaCl_2.2H_2O$
- 150 mL/L Glycerol

The mixture was made up to ~900 mL with de-ionized water. The pH of the solution was adjusted to 5.85 using acetic acid and KOH. $MnCl_2.4H_2O$ (9.896 g) was added, the solution made up to 1 L, and then sterilized using a filter sterilization unit.

TFB2:

- 2.093 g/L MOPS-free acid
- 11.027 g/L $CaCl_2.2H_2O$
- 1.209 g/L RbCl
- 150 mL/L glycerol

The mixture was made up to ~900 mL with de-ionized water. The pH of the solution was adjusted to 6.5 with KOH. The solution was made up to 1L and sterilized using a filter sterilization unit.

2.1.6.2 Method

Bacterial cells were streaked onto a LB-agar plate containing the appropriate antibiotics and incubated overnight at 37°C. 2X LB media (15 mL) containing the appropriate antibiotic was inoculated with a single colony and incubated overnight at 37°C and 180 rpm. SOB media (2 x 500 mL) was inoculated with 10 mL of overnight culture and grown at 37°C and 180 rpm until the OD_{600} had reached 0.4. The cultures were centrifuged at 1800 xg, 4°C for 15 minutes. The cell pellets were resuspended in TFB1 (160 mL), incubated on ice for 5 minutes, and then centrifuged again. The cell pellets were resuspended in TFB2 (70 mL), aliquoted into 500 µL volumes, flash-frozen in liquid nitrogen, and stored at 80°C.

2.2 Protein Expression

2.2.1 Cell maintenance

2.2.1.1 Media

Lysogeny Broth (LB) Media:

- 10 g Tryptone
- 10 g Sodium chloride
- 5 g Yeast extract

Made up to 1 L and then autoclaved.

2X Yeast tryptone (2XYT) Media:

- 16 g Tryptone
- 6 g Sodium chloride
- 10 g Yeast extract

Made up to 1 L and then autoclaved.

Terrific Broth (TB) Media:

- 12 g Tryptone
- 24 g Yeast extract
- 9.4 g Dipotassium hydrogen phosphate
- 2.2 g Potassium dihydrogen phosphate
- 5 % Glycerol

Made up to 1 L and then autoclaved.

HEK293 Complete Media:

- 500 mL DMEM (Gibco, 10741574)
- 10% FBS (Gibco, A5256801)
- 1% Pen-strep (Gibco, 15140122, 10,000 U/mL)

NK92mi Maintenance Media:

- 500 mL RPMI 1640 (Gibco, 11875093)
- 5 % FBS
- 5% Human Serum (Sigma, H4522)
- 100 Units/mL Pen/Strep
- 2 mM L-Glutamine (Gibco, 2503008)
- 10 mM HEPES (Gibco, 15630056)
- 1 mM Sodium Pyruvate (Gibco, 11360070)
- 50 µM 2-Mercaptoethanol (Gibco, 21985023)

2.2.1.2 *Sf9* Cell maintenance

Sf9 cells were revived and established by the Biotech SRF in our department. After the second passage, the cells were handed to me. The cells were split to $0.7x10^6$ cells/mL by the addition of Sf-900 II SFM medium and incubated at 27°C with shaking at 100 rpm for 72 hours before being split again.

2.2.1.3 HEK293 Cell Maintenance

Adherent HEK293 cells were revived from DMSO stocks and established using HEK293 complete media in appropriately size T-flasks. At >90% confluency, the cells were passaged; the media was removed, the cells were rinsed with DPBS (Gibco, 14190144, calcium and magnesium free), and then trypsinised using TrypLE™ Express Enzyme (Gibco, 12604021). The cells were centrifuged at 1000 xg for 3 minutes and the pellet was resuspended in complete media. The cells were seeded to 40% confluency and incubated at 37°C, 5% CO_2 for 48 hours.

2.2.1.3 NK92mi Cell Maintenance

Suspension NK92mi cells were revived and established by Dr Ashwin Jainarayanan. After two weeks of steady culture growth, the cells were handed to me. The cells were cultured between 0.4 - 0.8x10^6/mL by the addition of NK92mi maintenance media and incubated in stationary T-flasks at 37°C, 5% CO_2 for 48-96 hours. At every other occasion of cell splitting, IL2 (X1000) was added to the cultures. At no point was more than half the media replaced with fresh media.

2.2.1.4 Expi293 Cell maintenance

Expi293 cells were revived and established by Biotech SRF. After the second passage, the cells were handed to me. The cells were split to 0.6x10^6 cells/mL by the addition of Expi293 medium (Gibco, A1435101) and incubated at 37°C, 8% CO2, and 80% humidity with shaking at 120 rpm for 24-48 hours before being split again.

2.2.2 *E. coli* Expression

Construct DNA was isolated using a Qiagen miniprep kit (Qiagen, 27104) and transformed into chemically competent BL21(DE3)R3-pRARE2 (50 µL) following the method outlined in **2.1.2.3**. LB media (10 mL per L of large-scale expression) containing the appropriate antibiotic was inoculated with a single colony and incubated overnight at 37°C and 180 rpm. Autoclaved media (LB, 2YT, or TB) containing the appropriate antibiotic was inoculated with the overnight culture (10 mL per 1 L). The cultures were grown at 37°C and 180 rpm until the appropriate OD_{600} had been reached (LB and 2YT: 0.8, TB: 1.2). The cells were induced by

the addition of IPTG (0.4 mM), and the cultures were grown overnight at 18°C and 180 rpm.

The cultures were centrifuged at 5000 xg for 20 minutes at 4°C. The cell pellets were resuspended in a minimal volume binding buffer containing cOmplete™ protease inhibitor cocktail (Roche, 11836153001, 1 tablet per 25 mL) and stored at -20°C.

2.2.3 *Sf9* Expression

A day prior to infection, the *Sf9* cells were split to $1.1x10^6$ cells/mL and incubated at 27°C with shaking at 100 rpm. Amplified P2 virus (2-5 mL per L) was added to the cultures, and they were grown at 27°C with shaking at 100 rpm for 48-72 hours.

The cultures were centrifuged at 900 xg for 20 minutes at 4°C. The cell pellets were resuspended in a minimal volume binding buffer containing cOmplete™ protease inhibitor cocktail (Roche, 11836153001, 1 tablet per 25 mL) and stored at -20°C.

2.2.4 Mammalian Expression

2.2.4.1 Transient Transfection

A day prior to transfection, the mammalian cells were seeded (HEK293: 40% confluency. Expi293: $1x10^6$ cells/mL) and incubated with their normal growth conditions. 3XFLAG KCTD9 DNA was obtained from MACH1 cells using a QIAgen maxiprep kit (Qiagen, 12123). The DNA (1 mg per L of large-scale expression) and Linear-PEI (3 mg) were separately diluted in Opti-MEM (GIBCO, 31985062, 20 mL each). The two Opti-MEM mixtures were slowly combined and

incubated at room temperature for 30 minutes. Sodium butyrate (10 mM) and then the DNA/PEI mixture was slowly added to the cell cultures. The cells were incubated with their normal growth conditions for 48 hours.

2.2.4.2 Cell Harvest

The HEK293 cells were harvested by removing the media with an aspirator, rinsing cells with Dulbecco's phosphate-buffered saline (DPBS), and collecting the cells with a cell scraper in a minimal volume of ice-cold DPBS. The cells were centrifuged at 300 xg for 5 minutes at 4°C and the pellets were resuspended in ice-cold DPBS. The cells were centrifuged again, the supernatant removed, and the cell pellets were stored at -80°C.

The Expi293 cultures were centrifuged at 900 xg for 20 minutes at 4°C. The cell pellets were resuspended in a minimal volume lysis buffer containing cOmplete™ protease inhibitor cocktail (Roche, 11836153001, 1 tablet per 25 mL) and stored at -80°C.

The NK92mi cells were centrifuged at 300 xg for 20 minutes at 4°C. The cell pellets were resuspended in ice-cold PBS and centrifuged again. The supernatant was removed, and the cell pellets were stored at -80°C.

2.3 Protein Purification

2.3.1 Buffers and Consumables

Lysis Buffer:

- 50 mM HEPES pH 7.5
- 150 mM NaCl
- 0.1% NP40
- 1 mM EDTA
- 5 mM $MgCl_2$

- 10% Glycerol
- 1:500 PIN (Sigma-Aldrich, PPC1010)
- 1 mM DTT
- 1:1000 PMSF (ThermoFischer, 36978)
- 1:10000 Okadaic acid (Sigma-Aldrich, 39302)
- 1:100 Betaglycerolphosphate (Sigma Aldrich, G9422)

Binding buffer:

- 50 mM HEPES pH 7.5
- 500 mM NaCl
- 5% Glycerol
- 5 mM Imidazole

Wash Buffer:

- 50 mM HEPES pH 7.5
- 500 mM NaCl
- 5% Glycerol
- 30 mM Imidazole

Elution Buffer 1-4:

- 50 mM HEPES pH 7.5
- 500 mM NaCl
- 5% Glycerol
- Imidazole (EB1: 50 mM, EB2: 100 mM, EB3: 150 mM, EB4: 250 mM)

StreptactinXT Elution Buffer:

- 50 mM HEPES pH 7.5
- 500 mM NaCl
- 5% Glycerol
- 30 mM Imidazole
- 50 mM Biotin

Gel filtration buffer:

- 50 mM HEPES pH 7.5
- 300 mM NaCl
- 1 mM TCEP

Anion Exchange Buffer A:

- 50 mM HEPES pH 7.5
- 1 mM TCEP

Anion Exchange Buffer B:

- 50 mM HEPES pH 7.5
- 1M NaCl
- 1 mM TCEP

2.3.2 SDS-PAGE

Protein samples were analysed by SDS-PAGE gel, a qualitive electrophoresis method. The protein samples are denatured at high temperature in the presence of an anionic surfactant SDS (sodium dodecyl sulphate). The surfactant masks the surface of the denatured proteins in the sample to confer similar mass-to-charge ratios. The samples are applied to a NuPAGE™ Bis-Tris Midi protein gel (ThermoFischer, WG1403BOX, 4 to 12%) and a constant electric field (175 V for 45 minutes) separates components by mass.[97] The gel was stained with InstantBlue Coomassie (abcam, ab119211) and imaged using a ChemiDoc MP imager. Optionally, the gel was then stained with the Pierce™ Silver Stain Kit (ThermoFischer, 24612) and imaged.

2.3.3 Cell lysis

The cell pellets were defrosted on ice, benzonase was added, and then sonicated for 5 mins (5 seconds on, 10 seconds off, 35% amplitude) with occasional swirling. The lysate was centrifuged at 35000 xg, 4ºC for 30 minutes and the supernatant was passed through a 1.2 µm Minisart syringe filter (Sartorius, 17825Q).

2.3.4 Affinity Tag Purification

A difficulty in protein production is that the desired protein construct is produced alongside many other proteins that the cell requires.[98,99] Purification tags at the termini of constructs have been engineered to have high and reversible affinity to a solid phase resin. This allows the desired protein construct to be captured from the protein rich cell lysate. The bound resin is then washed to remove non-specific binding proteins, and then the construct is eluted by disrupting the interactions between the purification tag and the resin.

2.3.4.1 Histidine-tagged Protein Purification

Ni^{2+}-sepharose resin (Cytiva, 17526802, 0.5 mL bed volume per L of expression) was prepared by resuspending the resin in deionised water (10 CV), centrifuging the resin at 500 xg at 4ºC for 5 minutes, and then discarding the supernatant. The process was repeated twice more but with resuspension of the resin in binding buffer (10 CV) instead.

The filtered supernatant was applied to the prepared resin and the protein was allowed to bind through repeated inversion at 4ºC for 1 hour. The mixture was then centrifuged at 1000 xg, 4ºC for 10 minutes, and the supernatant was removed. The resin was resuspended in binding buffer (10 CV) and transferred to a gravity flow column. The resin was washed by gradient elution of binding buffer (10 CV), wash buffer (2 x 10-20 CV), and then elution buffer 1-4:

- EB1: 10 CV
- EB2: 10 CV
- EB3: 10 CV
- EB4: 3 x 10 CV

Samples of each fraction were collected and analysed by SDS-PAGE.

2.3.4.2 StrepII-tagged Protein Purification

Strep-TactinXT®4Flow resin (IBA Lifesciences, 2-5010-025, 0.5 mL bed volume per L of expression) was transferred to a gravity flow column and washed with wash buffer (30 CV).

The filtered supernatant was applied to the prepared resin and allowed to elute. The resin was washed by a two-step elution of wash buffer (20-30 CV), and StreptactinXT elution buffer (10 CV).

2.3.4.3 FLAG-tag Capture

ANTI-FLAG® M2 affinity gel resin (Sigma-Aldrich, A2220, 200 µL per L of expression) was prepared by resuspending the resin lysis buffer (10 CV), centrifuging the resin at 200 xg, 4ºC for 5 minutes, and discarding the supernatant. The process was repeated twice more.

The filtered supernatant was applied to the prepared resin and the protein was allowed to bind through repeated inversion at 4ºC for 2-3 hours. The resin was centrifuged at 200 xg, 4ºC for 5 minutes, and the supernatant was collected. The resin was resuspended in lysis buffer (10 CV), centrifuged as before, and the supernatant was collected. The washes were repeated four more times (3 x 10 CV lysis buffer, 1 x 10 CV water). Samples of each fraction, and the resin resuspended in sample buffer, were analysed by SDS-PAGE and western blot.

2.3.5 Polish Step

After the protein capture step, further purification steps are often required to remove contaminants (non-specific binding proteins, aggregates, etc.) and improve the sample quality.

2.3.5.1 TEV Protease Cleavage

TEV protease (150 µg per mg of sample) was added to the elution samples and incubated overnight at 4 °C. SDS-PAGE of the samples confirmed if complete cleavage had occurred.

2.3.5.2 Size Exclusion Chromatography

Size exclusion chromatography separates components by their size.[100] The porous matrix excludes larger species, which thus have a shorter mean path and elute from the column first. Smaller species will enter these pores and travel a longer path to elute later.

The protein sample was concentrated to half the loop volume using an equilibrated spin concentrator. The gel filtration column was prepared by washing the column with deionised water (1 CV), 0.1 M NaOH (0.25 CV), deionised water (1 CV), and then gel filtration buffer (1.5 CV). The sample was centrifuged at 21000 xg, 4^{o}C for 10 minutes to remove and precipitate aggregated species and then loaded onto the column. The column was eluted with an isocratic gradient of gel filtration buffer (1.25 CV). Fractions where absorbance at 280 nm was observed were analysed by SDS-PAGE.

2.3.5.3 Anion Exchange Chromatography

Ion exchange chromatography separates components by their net surface charge.[101] The protein material is reversibly bound to a solid matrix containing charged functional groups. The column is washed with buffer containing an increasing gradient of salt. At higher salt concentrations, the interactions between the charged surface residues and the charged functional groups of the matrix are disrupted, and the protein is eluted. Proteins with different surface residues will

have different affinities for the matrix and thus be eluted at lower/higher salt concentrations than other proteins.

The salt concentration of the protein sample was reduced by concentrating the sample using an equilibrated spin concentrator and then dilution in anion exchange buffer A. The anion exchange column was prepared by washing the column with anion exchange buffer A (5 CV). The sample was centrifuged at 21000 xg, 4°C for 10 minutes and then loaded onto the column. The column was washed anion exchange buffer A (5 CV) and then eluted with an increasing gradient of anion exchange buffer B (0-100% over 10-20 CV). Fractions where absorbance at 280 nm was observed were analysed by SDS-PAGE.

2.4 Co-immunoprecipitation

Co-immunoprecipitation (Co-IP) is a technique where a target protein and interacting partners are extracted from a sample using an antibody bound to a solid resin.[102] Various protease and phosphatase inhibitors (see lysis buffer in 2.3.2) are added to the sample to perserve the native states of the proteins at the time of harvest. For pulldown experiments and mass spectrometry proteomics, the protein complexes were not eluted off of the FLAG resin and instead used directly as SDS-PAGE/western blot samples, or on-bead digestion with trypsin.

ANTI-FLAG® M2 affinity gel resin (Sigma-Aldrich, A2220, 0.2 mL bed volume per L of expression) was prepared by resuspending the resin lysis buffer (10 CV), centrifuging the resin at 200 xg at 4ºC for 5 minutes, and discarding the supernatant. The process was repeated twice more.

The filtered supernatant was applied to the prepared resin and the protein was allowed to bind through repeated inversion at 4ºC for 2-3 hours. The mixture was transferred to a gravity flow column, washed with lysis buffer (5 x 10 CV), and then deionised water (10 CV). Samples of each fraction were collected and analysed by SDS-PAGE and western blot.

2.5 Western Blot

Western blotting allows a qualitative and semi-quantitative measurement of specific proteins within a sample at much lower levels than Coomassie staining. The sample components are separated by SDS-PAGE gel and then transferred to a membrane. The membrane is incubated with a primary antibody that will recognise and bind the protein of interest or purification tag. The membrane is then incubated with a secondary antibody that is conjugated with an enzyme for imaging (e.g. HRP - Horseradish peroxidase) and will recognise the primary antibody. For chemiluminescence imaging, the substrate is applied to the membrane and the conjugated enzyme catalyses the production of a luminescent signal. This signal is recorded and proportional to the specific protein present in the sample.[103]

2.5.1 Buffers and Materials

Transfer buffer:

- 25 mM TRIS base
- 190 mM Glycine
- 20% Methanol

PBST: 1x PBS (Phosphate-buffered saline) buffer with 1% Tween20

Antibodies:

- Mouse monoclonal ANTI-FLAG® M2 (Sigma-Aldrich, F1804, 1:3000)
- Mouse monoclonal GAPDH (Invitrogen, AM4300, 1:3000)
- Rabbit polyclonal CUL3 (Cell signalling, 2759, 1:2000)
- Rabbit polyclonal GOPC (Proteintech, 12163-1-AP, 1:1000)
- Goat Anti-Rabbit Immunoglobulins/HRP (Dako, P044801, 1:2000)
- Rabbit Anti-Mouse Immunoglobulins/HRP (Dako, P0260, 1:3000)

Western blot development: SuperSignal™ West Pico PLUS Chemiluminescent Substrate (ThermoFischer, 34580).

2.5.2 Method

The protein samples were run on an NuPAGE™ Bis-Tris Midi protein gel (150V for 1.5 hours). The gel samples were transferred to a PVDF membrane by a wet transfer protocol (300 mA for 1.5 hours) using transfer buffer. The membrane was washed with PBST (3 x 5 minutes), and then blocked with BSA (Bovine serum albumin, 2.5% in PBST) for 1 hour at room temperature at 55 rpm on a rolling platform. The membrane was washed again with PBST (3 x 1 minute), and then stained with primary antibody for 16 hours, at 4°C and 55 rpm. The membrane was washed with PBST (3 x 5 minutes), and then stained with the appropriate HRP-conjugated secondary antibody for 1 hour at room temperature and 55 rpm. The membrane was washed again with PBST (1 x 1 minute, 3 x 20 minutes) and a chemiluminescent signal was produced using the Pico Plus regents. The membrane was imaged using a ChemiDoc MP imager with exposures captured from 10 seconds up to 30 minutes.

2.6 Intact 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 de-convoluted using the MassHunter software (Agilent).

2.7 Mass Spectrometry Proteomics

I performed the FLAG Co-IP of samples derived from mammalian cells (HEK293 or NK92mi) that I cultured. The on-bead samples were handed to Dr Sarah Flannery and they performed to on-bead Smart digestion (Thermo Scientific) according to the manufacturer's instructions, using 2 µg trypsin. Peptide digests were then loaded onto Evotips (Evosep) and analysed by LC-MS using an Evosep One liquid chromatography system (Evosep) coupled to a TimsTOF Pro mass spectrometer (Bruker). Peptides were chromatographically separated using the 30 spd standard Evosep method. MS data were acquired in diaPASEF mode using 16 diaPASEF scans per TIMS-MS scan with an accumulation and ramp time of 100 ms, for a total cycle time of 1.80 seconds. The ion mobility range was set to 0.6 - 1.6 Vs/cm^2. Each mass window isolated was 26 m/z wide, ranging from 400-1200 m/z with an ion mobility-dependent collision energy that increased linearly from 20 eV to 59 eV between 0.6-1.6 Vs/cm^2.

Raw MS data were searched in DIA-NN v1.8.1 in library-free mode against the UniProt human proteome database (UP000005640, downloaded 13th May 2024) plus common contaminants.[104] Tryptic digestion (Trypsin/P) was permitted a maximum of 1 missed cleavage, with N-terminal M excision and C carbamidomethylation permitted as modifications. Match-between-runs and RT-dependent cross-run normalisation was enabled, with the neural network classifier set to double-pass mode. Using Perseus v2.0.11, proteins were filtered to include only those identified in at least 2 replicates of at least one experimental group. Intensity values were log2-transformed, and then missing values were imputed with random values generated from a downshifted normal distribution (width 0.3, downshift 1.8).[105] This data imputation is a standard practice in proteomics data analysis to allow statistical analysis across the whole dataset without the impact of low confidence data points within the recorded spectra.[106] Significant proteins were determined by Student's t-test with multiple testing correction using a permutation-based 5% FDR with 250 randomisations.

2.7 Structural Biology

2.7.1 Alphafold

Computational predictions of tertiary and quaternary protein structures from an amino acid sequence use to be challenging. For even small proteins, the calculations of the numerous physical interactions are computationally expensive and inaccurate. Homology modelling also presents a significant challenge when there is a lack of an appropriate homologue(s) with a solved structure. Alphafold utilises a neural network to consider homologous proteins identified by multisequence alignment and the thermodynamic properties using the Amber force field. It has been demonstrated to be highly accurate in critical assessment

of structure prediction assessment 2020 (CASP14) and recent developments allow oligomeric complexes, PTMs, and even protein-nucleic acid interactions to be modelled.[47,107]

Alongside these models, a predicted local distance difference test (pLDDT) score and predicted aligned error (PAE) are calculated.[108] The pLDDT metric is a per-residue, local confidence score that is scaled for 0-100. Residues with scores of 90+ are considered highly accurate for all parts of the residue. A pLDDT score between 70-90 indicates that while the backbone is likely to correct, the side chain maybe placed incorrectly. A score below 70 is to be treated cautiously and more sceptically as the score becomes lower. The PAE metric represents the confidence in the relative position of two residues within model. A low PAE score between two residues relates to a low predicted error (between 0 - 5 Å), and a high confidence in their relative positions. The modelling servers represent PAE scores in 2D PAE plots. When assessing Alphafold models, it is important to consider both of these metrics together to properly interpret the results.[47]

Alphafold2 models were initially generated by Colab-fold v1.5.2 using the default parameters on the ColabFold notebook (https://colab.research.google.com),[109] or on the COSMIC2 server (https://cosmic2.sdsc.edu), for larger models generated.[110] Later models were generated by Alphafold3 using the AlphaFold server (https://alphafoldserver.com/).[107]

2.7.2 X-ray Crystallography

Protein X-ray crystallography is one of the main techniques for determining the structure of protein samples up to atomic resolution. Protein crystallisation is promoted by incubation of a concentrated protein solution in the presence of precipitant. A wide variety of crystallisation conditions are assessed using course

crystal screens and initial hit conditions are then optimised by the use of fine screens or use of crystal seeding. Protein crystals are exposed to a strong source of X-rays and the elastic scattering of the X-ray by the crystal produces a diffraction pattern. Rotation of the crystal within the beamline samples the complete 3D space of the crystal to obtain a diffraction pattern series. The series is index, integrated, merged and scaled to make a reciprocal space-representation of the protein crystal. A real-space electron density map is initially produced by molecular replacement (MR) phasing using a homology model. This phasing method has the potential to introduce bias from the homology model but this can be limited by using MR models where the sidechains of every residue is removed, or models that only cover conserved regions. A model is then manually built by the crystallographer into the remaining unmodelled density. The final structure undergoes a series of optimisations and refinements to best account for the observed density following established geometric constraints.[111,112]

2.7.2.1 Protein Crystallisation

Protein samples were concentrated to ~10 mg/mL using a 10 kDa MWCO spin concentrator (Vivaspin, GE28-9323-60). The protein was seeded to a 3-well drop sitting drop vapour diffusion plate using a sptlabtech mosquito LV. Initially, pre-made coarse screens were used: Hampton Research Crystal Screen (Hampton Research, HR2-130, HCS), the Hampton Research Index (Hampton Research, HR2-134, HIN3) screen, the Joint Centre for Structural Genomics (Molecular Dimensions, MD1-40, JCSG) screen, Top 96 conditions screen (Molecular Dimensions, TOP96), and the Ligand Friendly Screen (Molecular Dimensions, MD1-122, LFS). Fine screens were designed in the Rockmaker software (Formulatrix) and made using a Formulatrix liquid handler. The crystal trays were

incubated at either 4°C or 20°C and monitored periodically by visible and UV imaging using a Rock Imager. On crystal harvest, reservoir solution (1 µL, optionally supplemented with cryoprotectant) was transferred to the drop, the crystals were mounted on appropriately sized loops, and vitrified in liquid nitrogen.

2.7.2.2 Data Processing

Diffraction data were collected on beamline I03 or I04 at the Diamond Light Source, Didcot, UK on Eiger2 XE 16M detectors. The data were pre-processed using the beamlines auto processing pipelines: Indexing and integration were performed using XIA2,[113] and merging by DIALS.[114] Phasing was solved by molecular replacement with Phaser-MR in Phenix using alphafold models as templates.[115] Models were refined utilising Phenix.refine and manually adjusted using COOT.[116,117] The models were validated using Molprobity and the protein data bank (PDB) data validation service before being deposited to the PDB.[118]

2.7.3 Cryo-EM

Cryogenic electron microscopy (cryo-EM) is an alternative structure determination technique to crystallography.[119] Large proteins or oligomers are applied to a grid mesh and rapidly frozen in liquid ethane. The state of the protein is preserved in a vitreous ice environment with protein ‘particles’ orientated randomly. The sample is exposed to a low dose of electrons and the contrast between the scattered and unscattered beams can be visualised by defocusing the objective lense. 2D images of protein ‘particles’ are identified and grouped together into 2D classes. These classes represent particular orientation or states of the protein sample. Once a good number of classes have been made which

represent many different orientations of the protein, the data classes are used to reconstruct a 3D model. Ab initio models are improved with homogenous, heterogenous, and local refinement methods to generate a real-space 3D volume map. Homology models are inserted into the volume and the model is then refined to best account for the observed coulomb potential map following established geometric constraints.

2.7.3.1 Method

Protein samples were concentrated to ~1 mg/mL using a spin concentrator. The sample (3 µL) was applied to a plasma cleaned Quantifoil Au R 1.2/1.3 grid, blotted with filter paper, and then plunge-frozen in liquid ethane using a Vitrobot. The grids were screened using a ThermoFischer Glacios electron microscope (200 kV), and then large datasets were collected using a ThermoFischer Krios electron microscope (300 kV) at Oxford Particle Imaging Centre (OPIC), Oxford.

The dataset was processed using Cryosparc v4.4.[120] Initial particles were selected using Cryosparc's blob picker function and 2D classification. Topaz particle picking models were trained using these particles for optimised particle picking. Topaz utilises neural networks to train models from a small particle set for more efficient particle identification.[121] Ab-initio maps were generated and used as a basis for homogenous and local refinement methods, where the appropriate applied symmetry. 3D variability assessments were used Cryosparc and static 3D classification was conducted on Relion.[122] Models were refined utilising Phenix.real-spacerefine and manually adjusted using Coot and ISOLDE.[123,124] The models were validated using PHENIX and the PDB data validation service before being deposited to the Electron Microscopy Data Bank (EMDB).

2.8 Biophysical Assays

2.8.1 AlphaScreen Peptide Displacement Assay

Peptide displacement assays were performed using a biotinylated peptide (ordered from LifeTein) and his-tagged target protein using the AlphaScreen® Histidine (Nickel Chelate) Detection Kit (Revity, 6760619C). Unlabelled peptides were dispensed into the assay plates (ProxiPlate-384 Plus, Perkin Elmer) using an Echo 525 Liquid Handler (Labcyte). Assays were performed in triplicate at room temperature in 20 µL samples containing optimised protein and peptide concentrations. The pre-mixed samples were added to the assay plate containing the unlabelled peptide dilution series, incubated for 10 minutes, and then AlphaScreen beads were added. Assay plates were incubated for 60 minutes before luminescence was measured using a Pherastar FS plate reader (BMG) with an excitation wavelength of 680 nm and an emission wavelength of 570 nm. Data were normalized to the non-biotinylated control sample and the IC_{50} determined from the nonlinear regression curve fit using GraphPad Prism.

2.8.2 Fluorescence Polarisation Peptide Displacement Assay

RhodamineB-labelled and unlabelled peptides were ordered from LifeTein. Assays were performed in triplicate at 25°C in 40 µL samples containing optimised protein and peptide concentrations within a 384-well plate. Preliminary assays identified the optimal peptide concentration by titration of the labelled peptide, the optimal protein concentration by titration of target protein, and the DMSO tolerance of the assay. The compounds were added to the assay plate containing pre-mixed samples, incubated for 10 minutes, and the polarisation measured using a HIDEX plate reader (Ex: 540 nm Em: 625 nm). Data were

normalized to the non-biotinylated control sample and the IC50 determined from the nonlinear regression curve fit using GraphPad Prism.

2.9 Molecular Graphics

Figures of structures and models were rendered using UCSF ChimeraX 1.7.1. RMSD measurements were calculated using the *matchmaker* command on ChimeraX, the RMSD for all atoms without filtering is reported. Columbic electrostatic potentials (ESP) were calculated using the *coulombic* command on ChimeraX with the default parameters. A surface model representation is generated and the local ESP is coloured with a red/white/blue palette (-10 to +10). Molecular lipophilicity potentials (MLP) were calculated using the *mlp* command on ChimeraX with the default parameters.[125,126] A surface model representation is generated and the local MLP is coloured with a teal/white/brown palette (-20 to +20).

Chapter 3: Structural Characterisation of KCTD9

3.0 Structural Characterisation of KCTD9

3.1 Introduction

As elaborated in chapter 1, KCTD9 is unusual in the KCTD family for containing three distinct domains. In addition to the conserved BTB (broad, tramtrack, bric-à-brac) and the highly variable CTD (C-terminal domain), it has an N-terminal ubiquitin-like (Ub-like) domain.[48] Prior to this work, the structural information of KCTD9 was just limited to the BTB domain, for which the Prive group determined a homo-pentameric structure by X-ray crystallography.[55] These researchers also demonstrated that CUL3 (cullin-3) was recruited by $KCTD9^{BTB}$ with nanomolar affinity and used homology modelling to hypothesize that CUL3 interacted with two adjacent KCTD9 subunits simultaneously through 'primary' and 'distal' interfaces. The C-terminal pentapeptide repeat (PPR) domain of KCTD9 is predicted to adopt an unusual solenoid structure that is uncommon in eukaryotic species.[66] Structures for other pentapeptide repeat containing non-human E3 ligases, such as SopA (PDB: 5JW7) and NleL (PDB: 3NAW), have been determined but these proteins appear unrelated to KCTD9.[127,128]. Examining the mode of action of family members such as KCTD5, the CTD is hypothesized to be responsible for substrate interactions.[46] The N-terminal Ub-like domain (also labelled as the DUF or KHA domain) is a mysterious outlier to the KCTD family and there has been no investigation into the role and function of this domain.

I aimed to determine the structure of KCTD9, particularly the pentapeptide repeat domain, in a homo-pentameric state and determine the substrate recognition motif. In this chapter, I describe the design, test expression, and observations from the purification of KCTD9 constructs used for structural studies. I solved the

structure of a KCTD9A-c073 (ubiquitin-like domain) construct by X-ray crystallography and I also produced two cryo-EM datasets: a KCTD9A-c047 only, and KCTD9A-c047/CUL3A-c012 complex. These datasets revealed the key interactions between the subunits that make up unexpected homo-decameric and eicosameric structures.

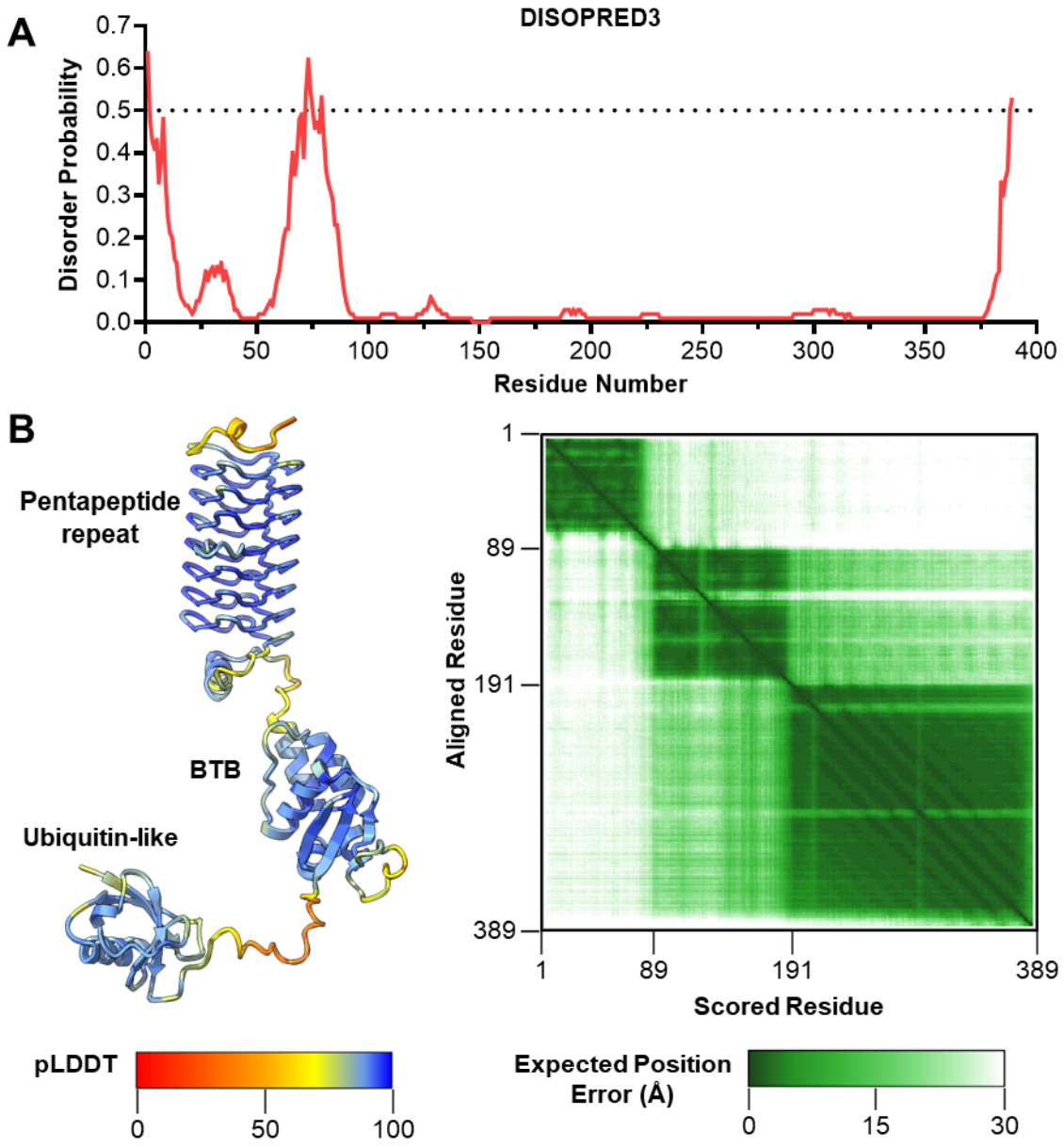


Figure 3.1: A) Disopred3 disorder prediction for KCTD9.[129] B) Left: KCTD9 monomer Alphafold model (AF-Q7L273-F1), coloured by pLDDT score. Right: 2D PAE graph, coloured by expected position error.

3.2 Results

3.2.1 Expression and Purification of New KCTD9 Constructs

It was thought previous attempts of bacterial expression of KCTD9 constructs were not successful due to an observed truncation of the protein. Additionally, I wanted to express the human construct of KCTD9 and the cellular machinery of prokaryote *E. coli* cells might not be best suited for the protein synthesis. With these considerations in mind, I designed constructs for expression in both *E. coli* and *Sf9* insect cells. Construct boundaries were guided using the Alphafold homology modelling and Disopred3 disorder prediction (Figure 3.1).[90] Some constructs contained a C-terminal glycine cap (GGGS) and/or C-terminal purification tags in an attempt to protect the construct from truncation (Figure 3.2). Additionally, constructs with functional tags such as Avi-tag and FLAG tag were designed to be used in characterisation assays of KCTD9 (Table 3.1 and 3.2).[130,131]

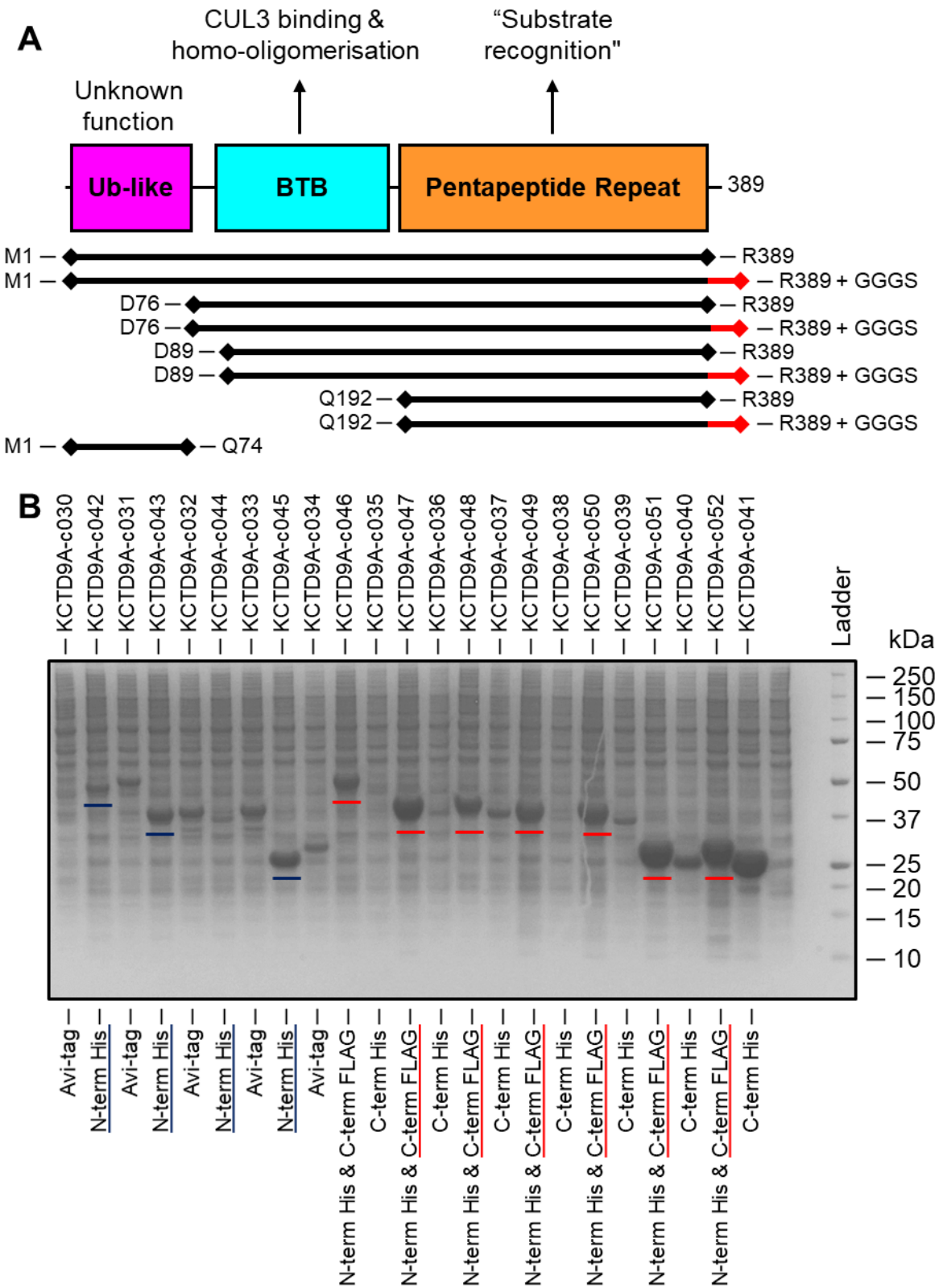


Figure 3.2: A) KCTD9 domain boundaries and designed constructs summary (See Table 3.1 and 3.2 for purification tags). B) KCTD9 *Sf9* constructs test expression results (3 mL). Coomassie-stained SDS PAGE gel of eluates after Ni^{2+}-sepharose purification of different KCTD9 constructs. Eluates correspond to the E4 elution buffer fraction (200 µL). Higher expressing C-terminal 3XFLAG tagged constructs underlined in red. Low expressing constructs that are without this tag underlined in blue.

Construct ID (Scarab)	Construct Boundaries	N-terminal Tag	C-terminal Tag	Test Purification Yield
KCTD9A-c013	M1-R389	His-tag	-	None
KCTD9A-c014	M1-R389	His-tag	Avi-tag	None
KCTD9A-c015	M1-R389 +GGGS	His-tag	Avi-tag	None
KCTD9A-c016	D89-R389	His-tag	Avi-tag	Low
KCTD9A-c017	D89-R389 +GGGS	His-tag	Avi-tag	Low
KCTD9A-c018	Q192-R389 +GGGS	His-tag	Avi-tag	Low
KCTD9A-c019	M1-R389	His-tag	StrepII-tag	None
KCTD9A-c020	D76-R389	His-tag	StrepII-tag	None
KCTD9A-c021	D76-R389 +GGGS	His-tag	StrepII-tag	None
KCTD9A-c022	D89-R389	His-tag	StrepII-tag	None
KCTD9A-c023	D89-R389 +GGGS	His-tag	StrepII-tag	None
KCTD9A-c024	Q192-R389	His-tag	StrepII-tag	None
KCTD9A-c025	Q192-R389 +GGGS	His-tag	StrepII-tag	None
KCTD9A-c026	M1-R389	-	His-tag	Not Tested
KCTD9A-c027	D76-R389	-	His-tag	None
KCTD9A-c028	D89-R389	-	His-tag	Not Tested
KCTD9A-c029	Q192-R389	-	His-tag	Not Tested
KCTD9A-c067	M1-L382 +QMSQSVR	His-tag	-	Low
KCTD9A-c068	M1-H383 +ISQSVR	His-tag	-	None
KCTD9A-c069	M1-L382	His-tag	-	Low
KCTD9A-c070	D76-L382 +QMSQSVR	His-tag	-	Low
KCTD9A-c071	D76-H383 +ISQSVR	His-tag	-	Low
KCTD9A-c072	D76-H383	His-tag	-	Low
KCTD9A-c073	M1-Q74	His-tag	-	High

Table 3.1: KCTD9 *E. coli* constructs summary table.

Construct ID (Scarab)	Construct Boundaries	N-terminal Tag	C-terminal Tag	Test Purification Yield
KCTD9A-c030	M1-R389	His-tag	Avi-tag	None
KCTD9A-c031	M1-R389 +GGGS	His-tag	Avi-tag	Low
KCTD9A-c032	D89-R389	His-tag	Avi-tag	Low
KCTD9A-c033	D89-R389 +GGGS	His-tag	Avi-tag	None
KCTD9A-c034	Q192-R389 +GGGS	His-tag	Avi-tag	Low
KCTD9A-c035	M1-R389	-	His-tag	None
KCTD9A-c036	D76-R389	-	His-tag	None
KCTD9A-c037	D76-R389 +GGGS	-	His-tag	Low
KCTD9A-c038	D89-R389	-	His-tag	None
KCTD9A-c039	D89-R389 +GGGS	-	His-tag	Low
KCTD9A-c040	Q192-R389	-	His-tag	Medium
KCTD9A-c041	Q192-R389 +GGGS	-	His-tag	High
KCTD9A-c042	M1-R389	His-tag	-	Medium
KCTD9A-c043	D76-R389	His-tag	-	Medium
KCTD9A-c044	D89-R389	His-tag	-	Medium
KCTD9A-c045	Q192-R389	His-tag	-	Medium
KCTD9A-c046	M1-R389	His-tag	FLAG-tag	Medium
KCTD9A-c047	D76-R389	His-tag	FLAG-tag	High
KCTD9A-c048	D76-R389 +GGGS	His-tag	FLAG-tag	High
KCTD9A-c049	D89-R389	His-tag	FLAG-tag	High
KCTD9A-c050	D89-R389 +GGGS	His-tag	FLAG-tag	High
KCTD9A-c051	Q192-R389	His-tag	FLAG-tag	High
KCTD9A-c052	Q192-R389 +GGGS	His-tag	FLAG-tag	High
KCTD9A-c053	D89-T380	His-tag	FLAG-tag	Not Tested
KCTD9A-c055	M1-L382 +QMSQSVR	His-tag	Avi-tag	Not Tested
KCTD9A-c056	M1-H383 +ISQSVR	His-tag	Avi-tag	Not Tested
KCTD9A-c057	D76-L382 +QMSQSVR	His-tag	Avi-tag	Low
KCTD9A-c058	D76-H383 +ISQSVR	His-tag	Avi-tag	Low
KCTD9A-c059	M1-L382 + QMSQSVR	His-tag	FLAG-tag	Not Tested
KCTD9A-c060	D76-L382 + QMSQSVR	His-tag	FLAG-tag	Not Tested
KCTD9A-c061	M1-H383 +ISQSVR	His-tag	FLAG-tag	Not Tested
KCTD9A-c062	D76-H383 +ISQSVR	His-tag	FLAG-tag	Not Tested
KCTD9A-c063	M1-L382 +QMSQSVR	His-tag	-	Not Tested
KCTD9A-c064	D76-L382 +QMSQSVR	His-tag	-	Not Tested
KCTD9A-c065	M1-H383 +ISQSVR	His-tag	-	Not Tested
KCTD9A-c066	D76-H383 +ISQSVR	His-tag	-	Not Tested

Table 3.2: KCTD9 *Sf9* constructs summary table.

All the multi-domain constructs were largely insoluble when expressed in *E. coli* cells on both a small (50 mL) and large scale (>1L). The Ub-like domain construct did yield soluble protein on a large scale and a crystal structure was obtained. Test purification (3 mL) of the constructs produced in *Sf9* cells were more promising as many constructs appeared to be expressed well and be soluble (Figure 3.2). Overall, full-length constructs were generally less well expressed when compared to the BTB-CTD length constructs. Additionally, constructs with a fused C-terminus FLAG tag (KCTD9A-c046 to -c052) appeared to be expressed much greater (and be more stable upon purification) than the equivalent 'FLAGless' counterpart (KCTD9A-c042 to -c045).

All of the full-length and BTB+CTD constructs were evaluated with large-scale expression and purification. On many occasions the yield was lower than first indicated, particularly the 'FLAGless' constructs or full-length constructs. Eventually, constructs KCTD9A-c047 (His-KCTD9$^{76-389}$-FLAG, 39.0 kDa) and KCTD9A-c048 (His-KCTD9$^{76-389}$-GGGS-FLAG, 39.3 kDa) were selected for structural studies as these constructs expressed reliably and high yields. The constructs were captured using Ni^{2+}-sepharose resin and the purification tag were not removed by TEV protease.

3.2.1.1 CUL3 Stabilises the KCTD9 Complex

During concentration steps and/or overnight incubation of the KCTD9 constructs, precipitation was observed. As KCTD9 is expected to form a multimeric cullin-RING assembly, I reasoned that this E3 ligase substrate adaptor would be less stable alone than in complex with CUL3 and other cullin-RING components.

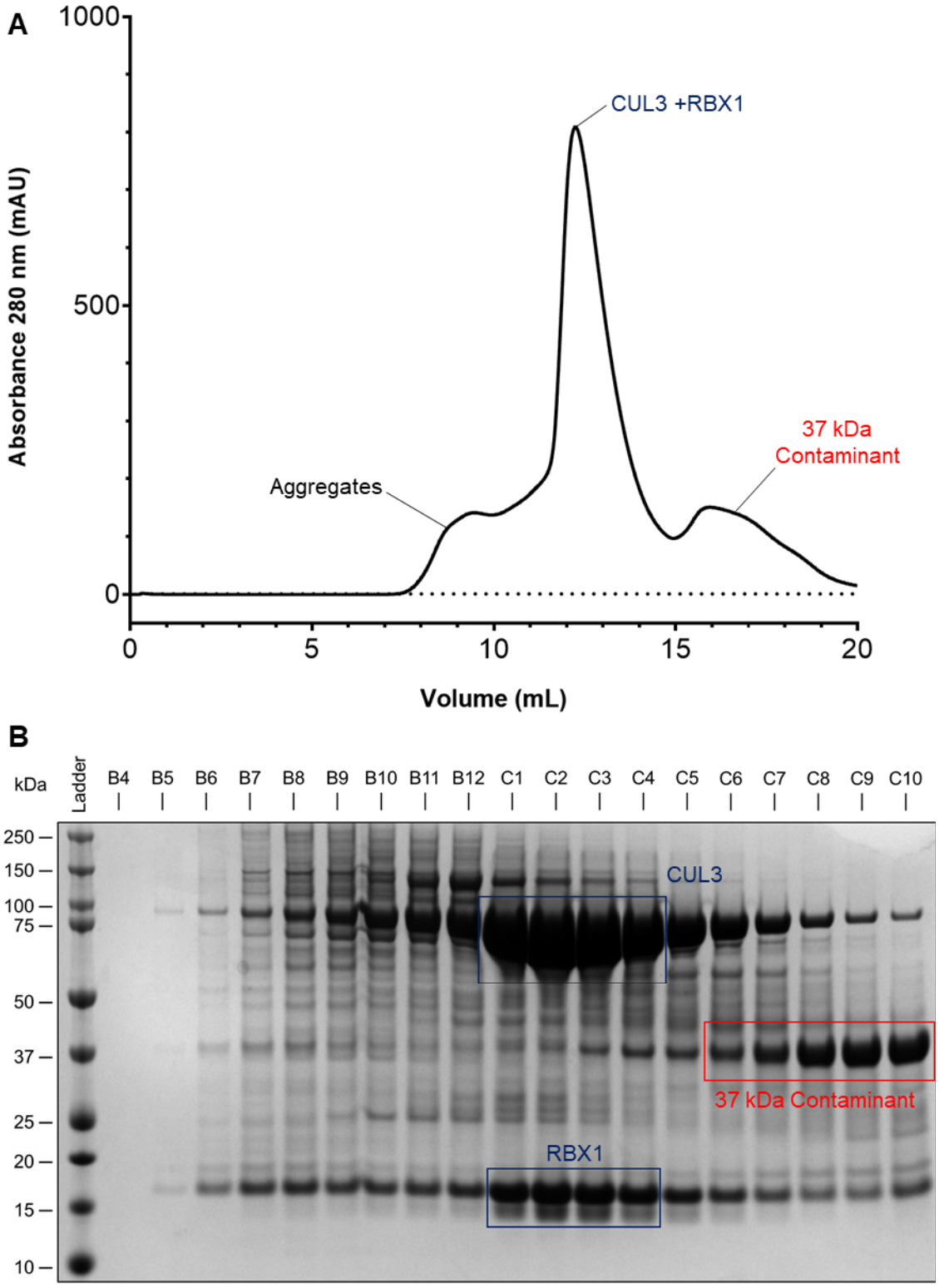


Figure 3.3: A) Size exclusion chromatography (Superose 6 Increase 10/300 GL) chromatogram of KCTD9A-c046, CUL3A-c038, and RBX1A-c002. Profile beyond 20 mL after sample injection is omitted. B) SDS-PAGE gel of relevant size exclusion chromatography fractions. 0.5 mL fraction sizes. 37 kDa contaminant highlighted in red.

Initially, I tried co-expression of full-length CUL3A-c038 (CUL3$^{1-768}$-His-FLAG, 91.7 kDa), RBX1A-c002 (His-RBX1$^{1-108}$, 14.9 kDa), and KCTD9A-c046 (His-KCTD9$^{1-389}$-FLAG, 46.9 kDa) by co-infection of *Sf9* with the appropriate bacmid viruses. Each litre of a *Sf9* culture (6 L in total) was infected with 1 mL of the CUL3 virus, 1 mL of the RBX1 virus, and 4 mL of the KCTD9 virus. The *Sf9* expression of KCTD9 did not improve, instead a CUL3-RBX1 complex was obtained in large excess, alongside an unknown 37 kDa contaminant (Figure 3.3).

A stable complex was made between CUL3A-c012 (CUL3$^{1-388}$-FLAG, 48.1 kDa) construct that was expressed in *E. coli* and KCTD9A-c047. The components were separately captured on Ni^{2+}-sepharose resin and then co-purified together by size exclusion chromatography (Figure 3.3). In comparison to a KCTD9 only sample, the KCTD9/CUL3 complex was less prone to precipitation and the overall yield for structural studies was greater.

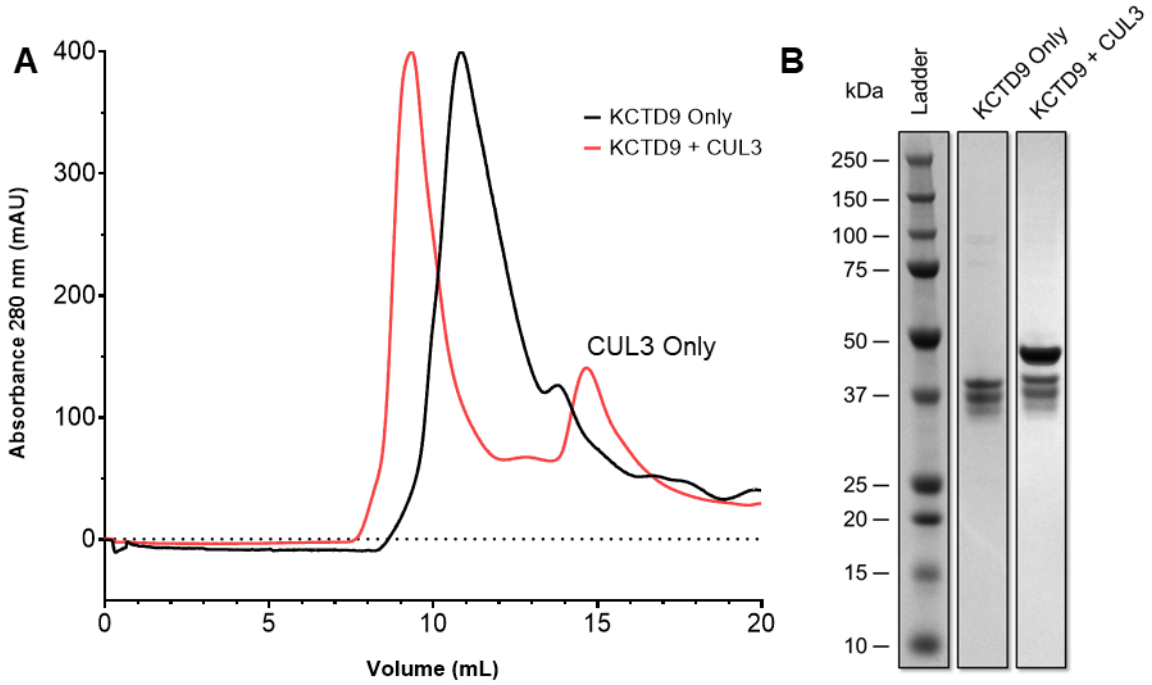


Figure 3.4: A) Comparison of size exclusion chromatography (Superose 6 Increase 10/300 GL) chromatograms for KCTD9 only and KCTD9/CUL3 complex samples. Profile beyond 20 mL after sample injection is omitted. B) SDS-PAGE gel of KCTD9 cryo-EM samples at 0.5 mg/mL.

3.2.1.2 Recombinant KCTD9 is Prone to Truncation

On large-scale purification, multiple bands were observed for the KCTD9 constructs; indicative of truncation. To characterise the truncated species, I purified three different constructs in parallel: KCTD9A-c047, KCTD9A-c048, and KCTD9A-c049 (His-KCTD9$^{89-389}$-FLAG, 37.6 kDa). The major ion peak from the intact mass spectrometry data (Figure 3.5) corresponded to a species which is the complete mass with an N-terminal methionine loss and acetylation (-89 Da, ■). Additionally, there were ion peaks corresponding to a M1 loss + acetylation + sodium adduct species (-6 Da, ●), and a M1 loss + acetylation + phosphorylation species (-67 Da, ▲). Finally, truncated species, with losses between 2376-2635 Da, was observed in all three samples. Comparison of the truncated masses to the construct sequences identified a consistent cleavage between H383-M384; which results in a loss of 6 residues at the C-terminus of the wild-type sequence, and any C-terminal extensions.

As KCTD9 is complexed to itself, I reasoned that it was unlikely that I would be able to separate the truncated species from the full-length construct as a pentameric complex would likely consist of a mixture of both full length and truncated KCTD9. Nevertheless, these samples were taken forward and used for structural studies.

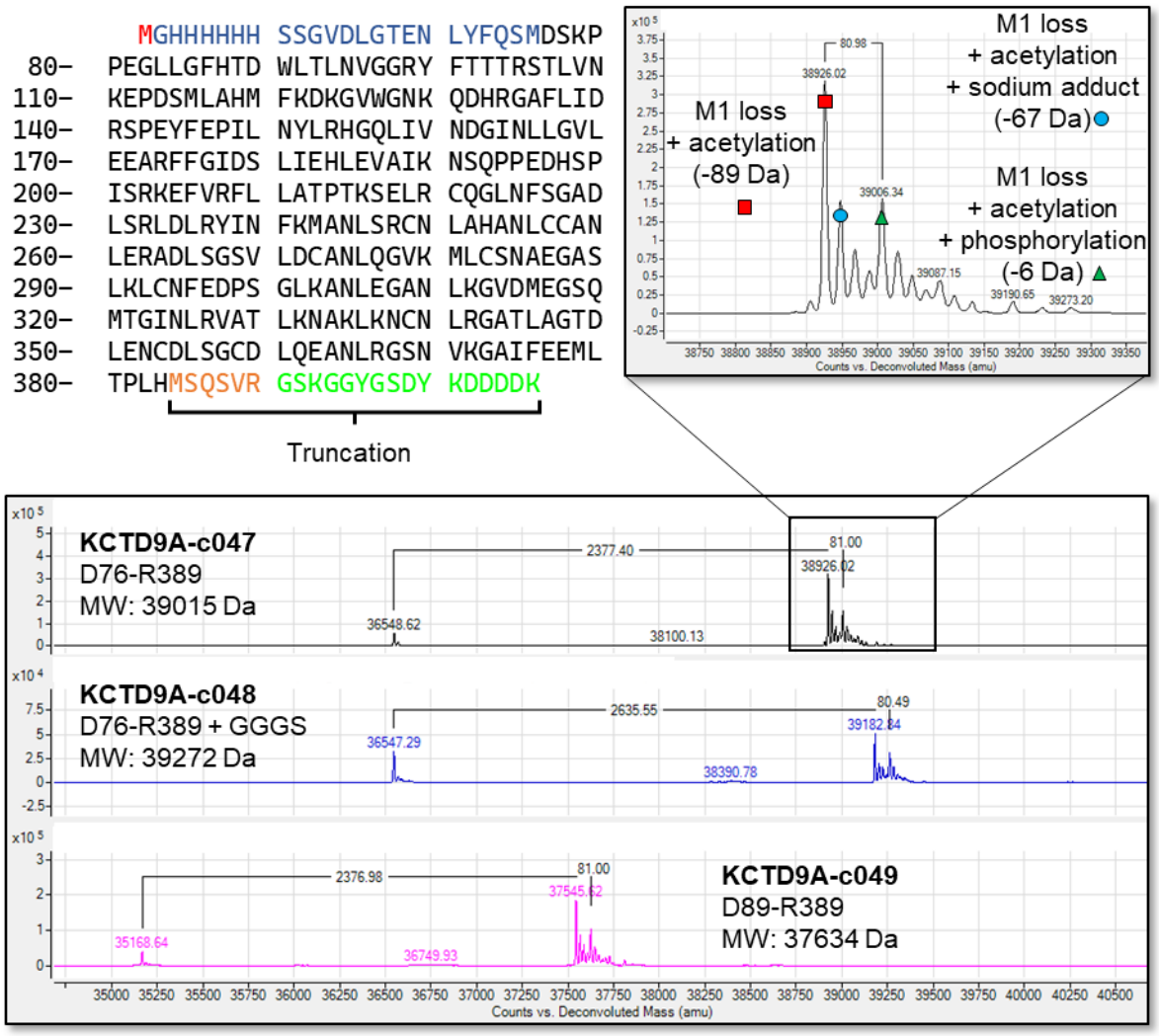


Figure 3.5: Intact mass spectrometry comparison of constructs KCTD9A-c047, -c048, and -c049. KCTD9A-c047 sequence is shown with N-terminal his-tag (blue), C-terminal FLAG tag (green), and missing KCTD9 residues (orange). Major observed post translational modifications of KCTD9A-c047 are identified and denoted with corresponding symbols (■, ●, ▲).

3.2.2 X-ray Crystallography of KCTD9

Initially, I tried to determine the structure of the pentapeptide repeat (PPR) domain by X-ray crystallography. Crystals of a KCTD9A-c007 (His-KCTD9$^{198-389}$, 23.4 kDa) had been previously obtained in our department, but they yielded low resolution, high mosaicity diffraction data that could not be phased by molecular replacement. I was able to produce similar quality datasets using KCTD9A-c005 (His-KCTD9$^{191-389}$, 24.1 kDa).

After cloning new multi-domain KCTD9 constructs, I attempted to purify several KCTD9 constructs (Table 3.3) for coarse crystal screens but the yields of these constructs were too low and too prone to precipitation at higher concentrations to achieve this. No crystal formation occurred when KCTD9A-c046 at 4.6 mg/mL was applied to coarse crystal screens. Crystal formation may have been hindered by the sample containing an inhomogeneous mixture of truncated/untruncated species (see section 3.2.1.4), and this prevented regular crystal contacts to be made.

3\.2.2.1 Structural Determination of the Ubiquitin-like Domain

KCTD9A-c073 (His-KCTD9$^{1-74}$, 23.4 kDa), a high yielding Ub-like domain construct was expressed in *E. coli* and purified by using Ni^{2+}-sepharose resin and size exclusion chromatography. The protein was concentrated to 10 mg/mL and applied to coarse crystal screens. The best diffracting crystal (Figure 3.6A) was obtained in a 150 nL sitting drop containing a 2:1 ratio of the protein sample and reservoir solution (0.1M Bis-tris propane pH 7.5, 0.2 M sodium sulphate, 20% w/v PEG 3350, 10% v/v ethylene glycol). The crystal was cryoprotected by addition of the reservoir solution (1 µL) containing 20 % glycerol to the well and vitrified in

liquid nitrogen. A 2.2 Å diffraction dataset was obtained by unattended data collection on the I04 beamline at Diamond Light Source, Didcot, UK. Datasets from two sweeps were merged and pre-processed by the xia2-dials automated pipeline at the beamline. R-merge and B-factor did not significant vary across the collection (Figure 3.6C), indicating that radiation damage of the crystal did affect the quality of the dataset. Additionally, reprocessing of the data to omit batches 1900-2300, where R-merge was slightly higher, only provided minor improvements to dataset quality. Therefore, molecular replacement was performed on the automated pipeline dataset using Phenix.phaser-MR and an Alphafold3 generated structure of KCTD9A-c073 as the search model. The model was refined and validated using COOT, Molprobity, and the PDB validation server (Table 3.4).[116,118]

Construct ID (Scarab)	Purification Result	Crystal Screen	Crystallisation Result	Diffraction Dataset
KCTD9A-c005	9.8 mg/mL	LFS, 20 ºC	Several	>4.0 Å, anisotropic
		HIN3, 20 ºC	Several	>4.0 Å, high mosaicity
KCTD9A-c007	6.8 mg/mL	LFS, 20 ºC	Several	>4.0 Å, high mosaicity
KCTD9A-c031	Low yielding	-	-	-
KCTD9A-c033	Low yielding	-	-	-
KCTD9A-c042	Low yield/aggregation	-	-	-
KCTD9A-c044	Low yield/aggregation	-	-	-
KCTD9A-c046	4.3 mg/mL	LFS, 20 ºC	None	-
		JCSG, 20 ºC	None	-
		HIN3, 4 ºC	None	-
		HCS, 4 ºC	None	-
KCTD9A-c047	Low yield/aggregation	-	-	-
KCTD9A-c048	Low yield/aggregation	-	-	-
KCTD9A-c073	10 mg/mL	LFS, 20 ºC	Several	2.2 Å, solved
		JCSG, 20 ºC	Several	3.8 Å, not processed
		HIN3, 4 ºC	None	-
		HCS, 4 ºC	None	-
		TOP96, 4 ºC	None	-
		TOP96, 20 ºC	None	-

LFS: Ligand-Friendly Screen
HIN3: Hampton Research Index Screen
HCS: Hampton Crystal Screen
TOP96: Top 96 conditions (PDB) Screen
JCSG: Joint Center for Structural Genomics Screen

Table 3.3: KCTD9 purification and crystallisation summary table. Best dataset is described in diffraction dataset column.

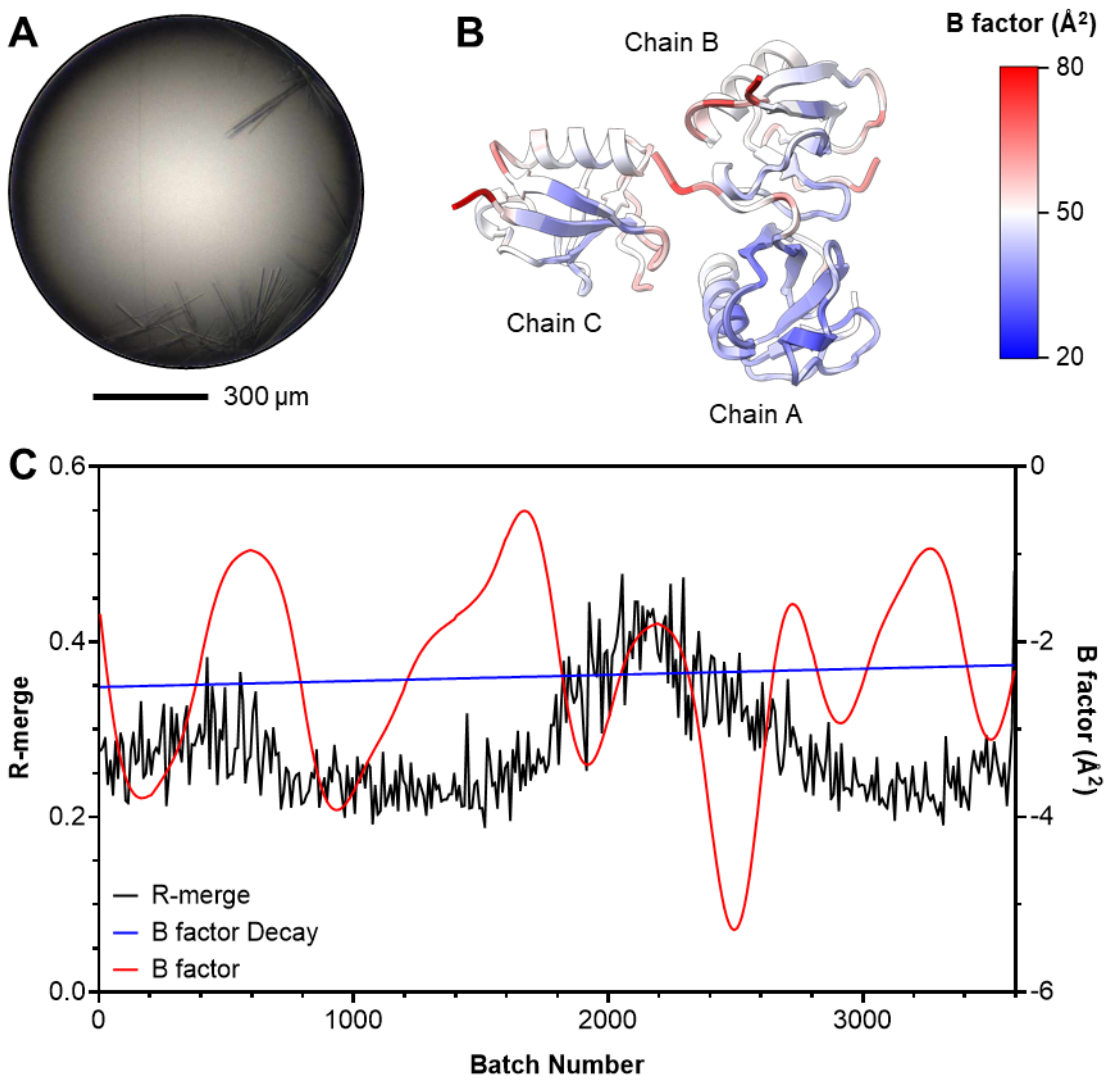


Figure 3.6: A) KCTD9A-c073 crystals in a 150 nL drop, 4 days after dispensing. Formulatrix Rock Imager. B) ASU of KCTD9A-c073, coloured by B factor. C) R-merge (black) and B factor (red and blue) statistics of the merged sweeps (3600 images total).

Wavelength (Å)	0.9763
Resolution range (Å)	35.29 - 2.21 (2.29 - 2.21)
Space group	*C 1 2 1*
Unit cell	
a, b, c (Å)	116.02, 57.28, 47.82
α, β, γ (º)	90.00 92.24 90.00
Total reflections	109,516 (10,853)
Unique reflections	15,839 (1,559)
Multiplicity	6.9 (7.0)
Completeness (%)	99.58 (98.27)
Mean I/σ(I)	5.47 (0.56)
Wilson B-factor (Å^2)	39.98
R-merge	0.2666 (2.542)
R-measure	0.2882 (2.748)
CC1/2	0.99 (0.315)
CC*	0.997 (0.692)
Reflections used in refinement	15781 (1536)
Reflections used for R-free	791 (79)
R-work	0.2265 (0.3882)
R-free	0.2817 (0.4044)
CC (work)	0.954 (0.562)
CC (free)	0.951 (0.406)
Number of non-hydrogen atoms	1771
macromolecules	1701
solvent	70
Protein residues	227
RMS bonds (Å)	0.009
RMS angles (º)	1.05
Ramachandran favoured (%)	96.8
Ramachandran allowed (%)	3.2
Ramachandran outliers (%)	0
Rotamer outliers (%)	3.14
Clash score	6.96
Average B-factor	53.1
macromolecules	53.42
solvent	45.21
Number of TLS groups	17

Table 3.4: Diffraction data collection and refinement statistics for KCTD9A-c073 domain. Statistics for the highest-resolution shell are shown in parentheses.

The structure contained 3 chains in the asymmetric unit (ASU) with an average root mean square deviation (RMSD) of 0.48 Å between the 3 chains (Figure 3.6B). This domain adopted a ubiquitin-like fold and was in strong agreement with the predicted structure by Alphafold3 (RMSD: 0.56).

KCTD9A-c073 was compared to ubiquitin (PDB: 1UBQ, RMSD: 7.70 Å) and NEDD8 (PDB: 1NDD, RMSD: 6.93 Å), a Ub-like protein which is responsible for cullin-RING activation. The α-helices and β-sheets of the structures could be well-aligned (Figure 3.7), however the loop regions β2-h1, and h1-β3 of KCTD9 are shorter than their equivalents present in ubiquitin and NEDD8, and β1-β2 is longer. Additionally, the electrostatic surfaces of the proteins are noticeably different. Ubiquitin and NEDD8 both have a negatively and a more positively charged region, whereas KCTD9 has a positively charged surface and a more neutral surface. Furthermore, the surface of KCTD9 does not resemble the cullin binding surface found on NEDD8.

The refined structure was uploaded to the DALI server and compared against structures deposited on the PDB.[49] Doublecortin (PDB: 5IKC), a microtubule-associated protein, was the most structurally similar human protein (RMSD: 2.54 Å) to KCTD9A-c073. Again, doublecortin has a distinct electrostatic surface when compared to KCTD9 and the loop regions of β2-h1 and h1-β3 are several residues longer than KCTD9. As ubiquitin, doublecortin, and NEDD8 all have different roles from each other, despite a similar 3D structure, I hypothesize that the Ub-like domain of KCTD9 has an unrelated function and/or binding partner to these proteins.

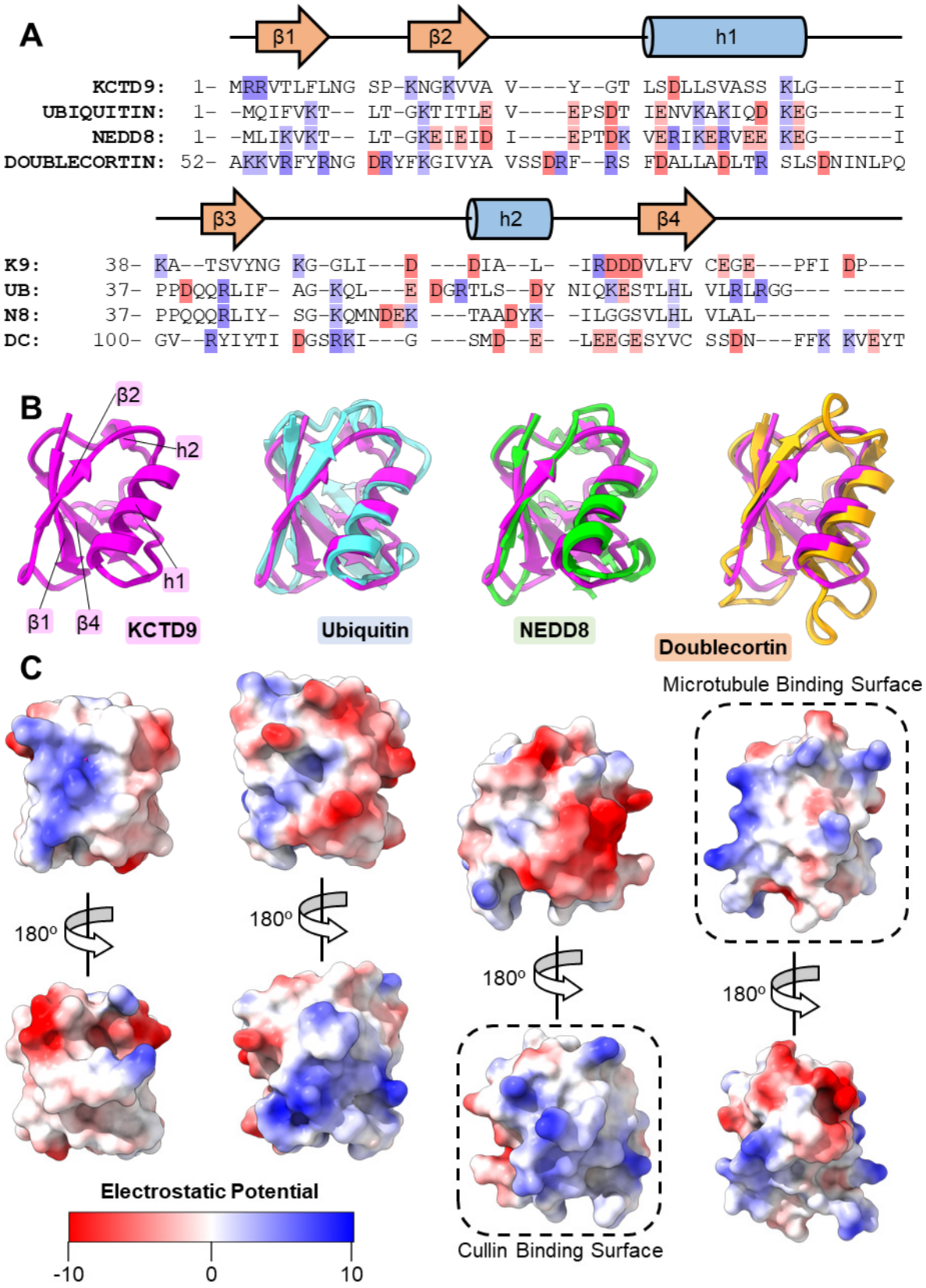


Figure 3.7: A) Dali structure-sequence alignment of KCTD9[1-73], ubiquitin, NEDD8, and doublecortin[52-139]. Positively and negatively charged residues are highlighted (blue and red respectively). B) Overlay of KCTD9A-c073 (purple) with ubiquitin (PDB: 1UBQ, cyan),[4] NEDD8 (PDB: 1NDD, green),[132] and doublecortin (PDB: 5IKC, orange),[133] from left to right. C) Surfaces of KCTD9A-c073, ubiquitin, NEDD8, and doubecortin (left to right). Coloured by electrostatic potential.

3.2.3 Cryogenic Electron Microscopy of KCTD9

As I was unable to obtain any X-ray crystallography datasets for the constructs I created (excluding KCTD9A-c073), I tried to obtain structural data by cryogenic electron microscopy (cryo-EM). The large size of a KCTD9$^{76-389}$ homopentamer (~200 kD, or ~440 kDa when it is associated with CUL3^{NTD}) and the predicted C5 symmetry of this complex, indicated that this sample could be well suited for this method. KCTD9A-c047 was expressed in *Sf9* cells and captured by Ni^{2+}-sepharose resin. Half of the eluted protein was complexed with CUL3A-c012, and the two samples were purified by size exclusion chromatography (Superose 6 Increase 10/300 GL) using gel filtration buffer containing 5% glycerol. The KCTD9 only sample and KCTD9/CUL3 sample were concentrated to 3 mg/mL and 6 mg/mL respectively. The samples were applied to glow-discharged Quantifoil Au R 1.2/1.3 grids and plunge frozen in liquid ethane using a Vitrobot Mk4. Pleasingly, grids for both samples with good-high particle density and suitable ice-thickness were made. After initial screening, a large data set of each sample was collected (Table 3.5).

3.2.3.1 Data Refinement

For the KCTD9 only sample dataset, initially 77,925 particles were selected from 456 micrographs (10%) after using the Cryosparc's blob picker and three rounds of 2D classification (Figure 3.8).[120] At this point, it was clear that an unexpected decameric state was observed, and no pentameric species could be identified.

Initially, three Topaz particle picking models were trained from the 2D classes: one side view, one top view, and one oblique view.[121] The models were then used to re-pick the complete dataset, and after 2D classification of these particles,

544,123 particles were selected for ab-initio reconstruction. Only one of the three ab-initio classes resembled a sensible volume map and this contained 30.2% of the re-picked particles. The ab-initio model was used in homogenous refinement and D5 symmetry was applied (C1 and C5 symmetry was also examined to confirm D5 symmetry was appropriate). After local and global contrast transfer function (CTF) refinement, a map with an estimated GSFSC (Gold standard Fourier shell correlation) resolution of 2.28 Å was obtained.

Particle distribution analysis of this map indicated that many oblique views were missing (Figure 3.9B). To obtain these views, the map was used to generate 2D templates for template picking (20 Å low-pass filter). After two rounds of 2D classification, 1,121,190 particles were selected for ab-initio reconstruction. While this number was higher than expected for this dataset, it did not appear to impact the quality of the final map. One of the two ab-initio classes resembled the expected map and this was refined as previously yield a final map with an estimated GSFSC resolution of 2.28 Å and more complete particle distribution (Figure 3.9C).

The resolution of the cryo-EM map at PPR region was sufficient to place both the backbone and side chains of the residues (Figure 3.12A). The resolution diminished away from the map centre and this compromised the BTB domain region (particularly at the N-terminus of the construct). Local refinement using a mask of the BTB domain did improve resolution of this region and 3D variability analysis (3DVA) did not identify any particular motion. I hypothesize that as a result of there not being any CUL3 to stabilise this domain, the BTB region undergoes a broad range of motions and this ultimately compromises the resolution of cryo-EM map for this area.

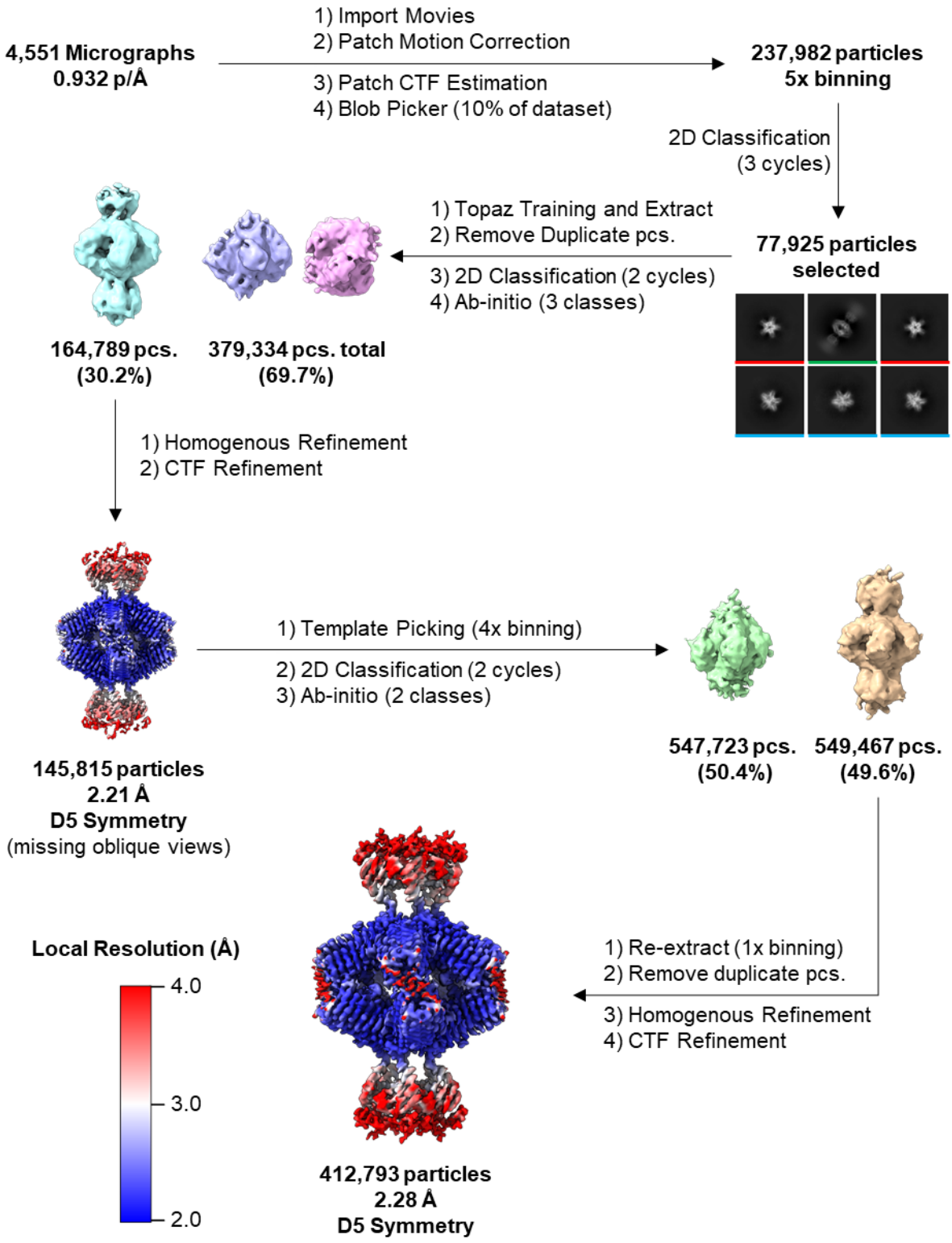


Figure 3.8: KCTD9 only cryo-EM dataset refinement pipeline. Particles were initially extracted from 10% of the dataset to train Topaz particle picking models. The map derived from Topaz picking of the full dataset lacked some oblique views. Template picking using this first map as a model manged to extract these oblique views and produce a final map where D5 symmetry was applied. 2D classes: top view (red), side view (green), and oblique view (blue).

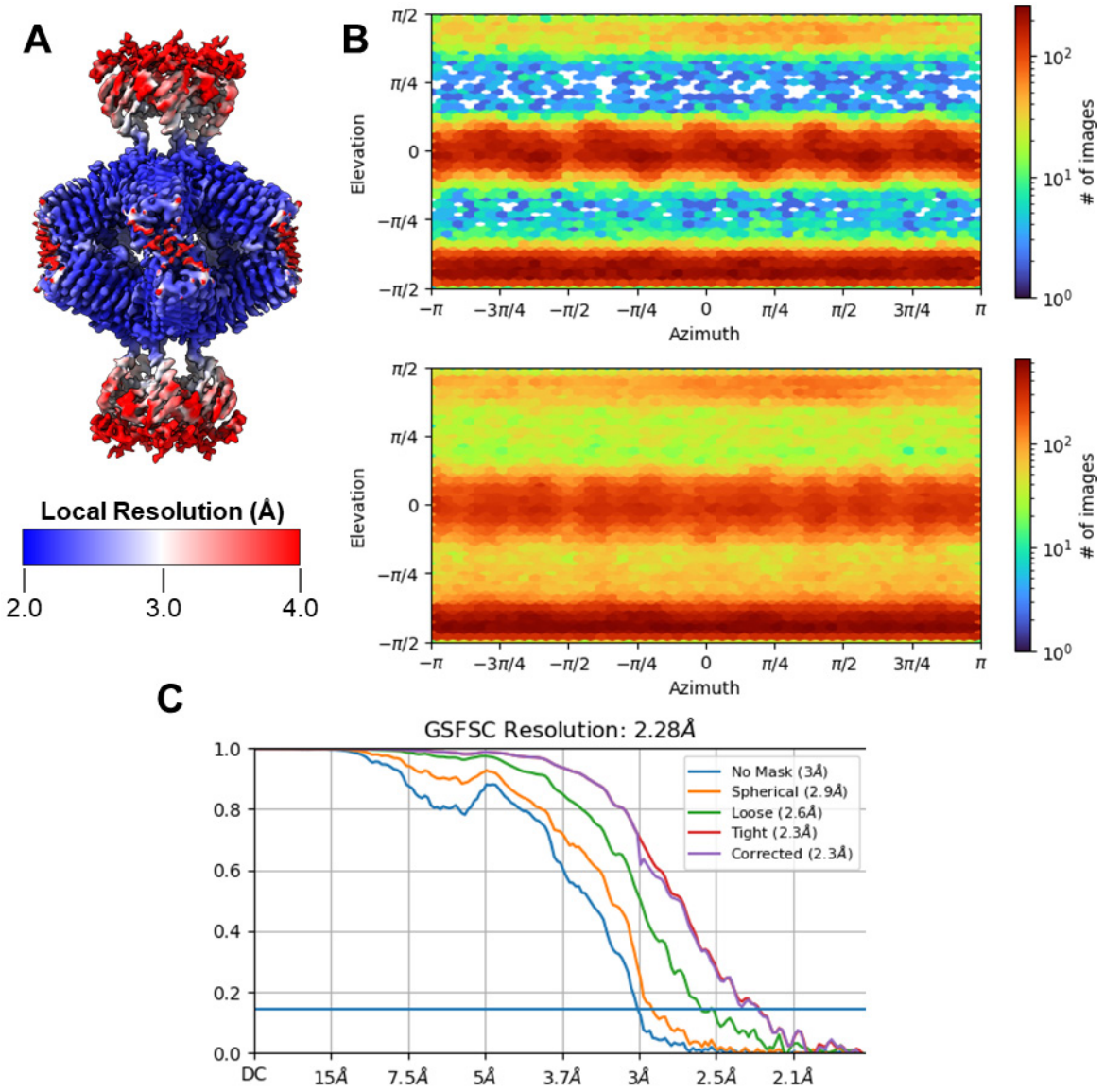


Figure 3.9: KCTD9 only volume map summary A) Local resolution estimation map coloured by local resolution (red-white-blue). B) CryoSPARC View Direction Distribution 2D plot. Top: Map derived from Topaz picking (oblique views missing), Bottom: Map derived from template picking. Coloured by number of images present (rainbow). C) CryoSPARC Gold Standard Fourier Shell Correlation (GSFSC) 2D plot. GSFSC resolution estimate at CC: 0.146.

A single PPR domain and a single BTB domain from a pentameric Alphafold model were fitted into the volume map and then linked together. After refinement of the model using phenix.real-space and ISOLDE, the model was replicated with D5 symmetry to generate a decameric model of KCTD9. 76% of the construct (G97-T101, and R140-E376) was confidently built but some portions of the BTB domain were left unmodelled (Figure 3.12A, Table 3.5).

A similar processing pipeline was used for the KCTD9/CUL3 dataset (Figure 3.10); 45,582 particles were selected from 1332 micrographs (10%) and used to train three Topaz particle picking models (top, side, and oblique views). Particles were picked using these models and, after 2D classification, 362,617 particles from the complete dataset were used in ab-initio reconstruction. One volume (that accounted for 57.4% of the re-picked particles) was used for homogenous refinement with D5 symmetry applied and CTF refinement. Again, a map with GSFSC resolution of 2.31 Å was obtained, but this map lacked some oblique particles (Figure 3.11B). This map was used for template picking and, after 2D classification, 416,405 were selected for ab-initio reconstruction. Only one of the models resembled the expected cryo-EM map (284,047 particles, 67.7%), but the as the other model possessed some features of the complex, all of the particles were used for homogenous refinement and CTF refinement.

The resolution of both KCTD9 domains was relatively good; however, the resolution at CUL3 components rapidly diminished for residues further away from the map centre. CryoSPARC 3DVA and static 3D classification using Relion revealed a 10° twisting motion of the BTB domain, relative to the PPR domain (Figure 3.12D).[122] This motion, around a small flexible tether region (Q192-D196), resulted in an 16 Å shift at the extremities of the CUL3 construct, which reduced

the resolution of the cryo-EM map. Local refinement using a pentameric KCTD9^{BTB}/CUL3^{NTD} mask and a decameric KCTD9^{PPR} mask generated two final cryo-EM maps with overall improvements of each region (GSFSC resolution of 2.67 Å and 2.12 Å respectively, Figure 3.11A).

The KCTD9/CUL3 complex model was into built two volume maps from local refinement using the same methods as the KCTD9 only dataset. The was T24-M219 and S231-M273 of CUL3 and D89-E204 of KCTD9 modelled using the KCTD9^{BTB}/CUL3^{NTD} local refinement map. Only G97-L110 and S141-E376 of KCTD9 could only be built into the KCTD9^{PPR} local refinement map. Combining these two models by superimposing the BTB domains to generated a KCTD9/CUL3 structure model that covered 96% of the KCTD9 construct and 61% of the CUL3 construct (Figure 3.12B-C, Table 3.5).

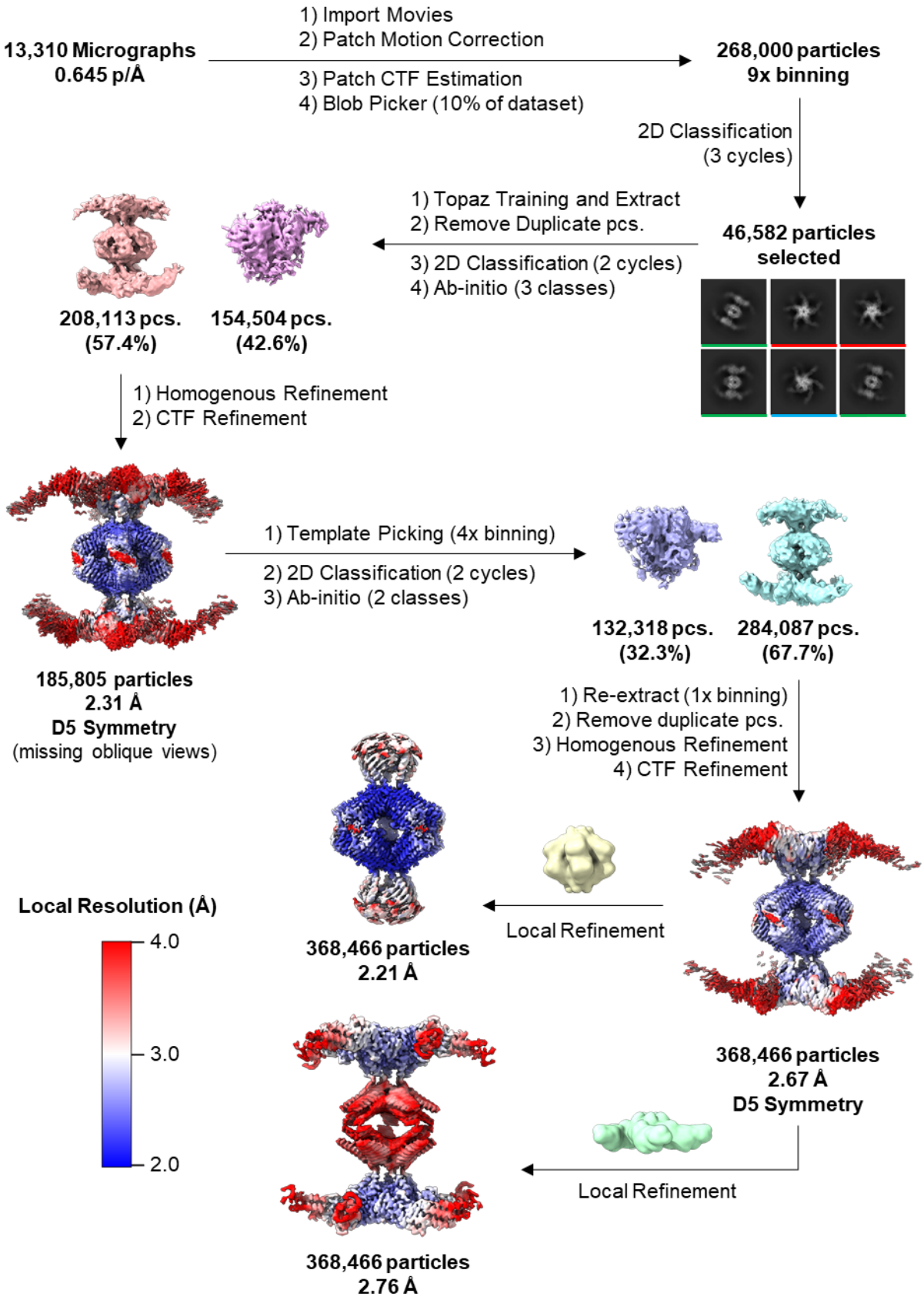


Figure 3.10: KCTD9/CUL3 cryo-EM dataset refinement pipeline. Particles were initially extracted from 10% of the dataset to train Topaz particle picking models. The map derived from Topaz picking of the full dataset lacked oblique views. Template picking, using this first map as a model, extracted these oblique views. Two local refinement maps were generated: Central KCTD9^{PPR} region, and the KCTD9^{BTB}/CUL3^{NTD} region. D5 symmetry was applied to both maps.

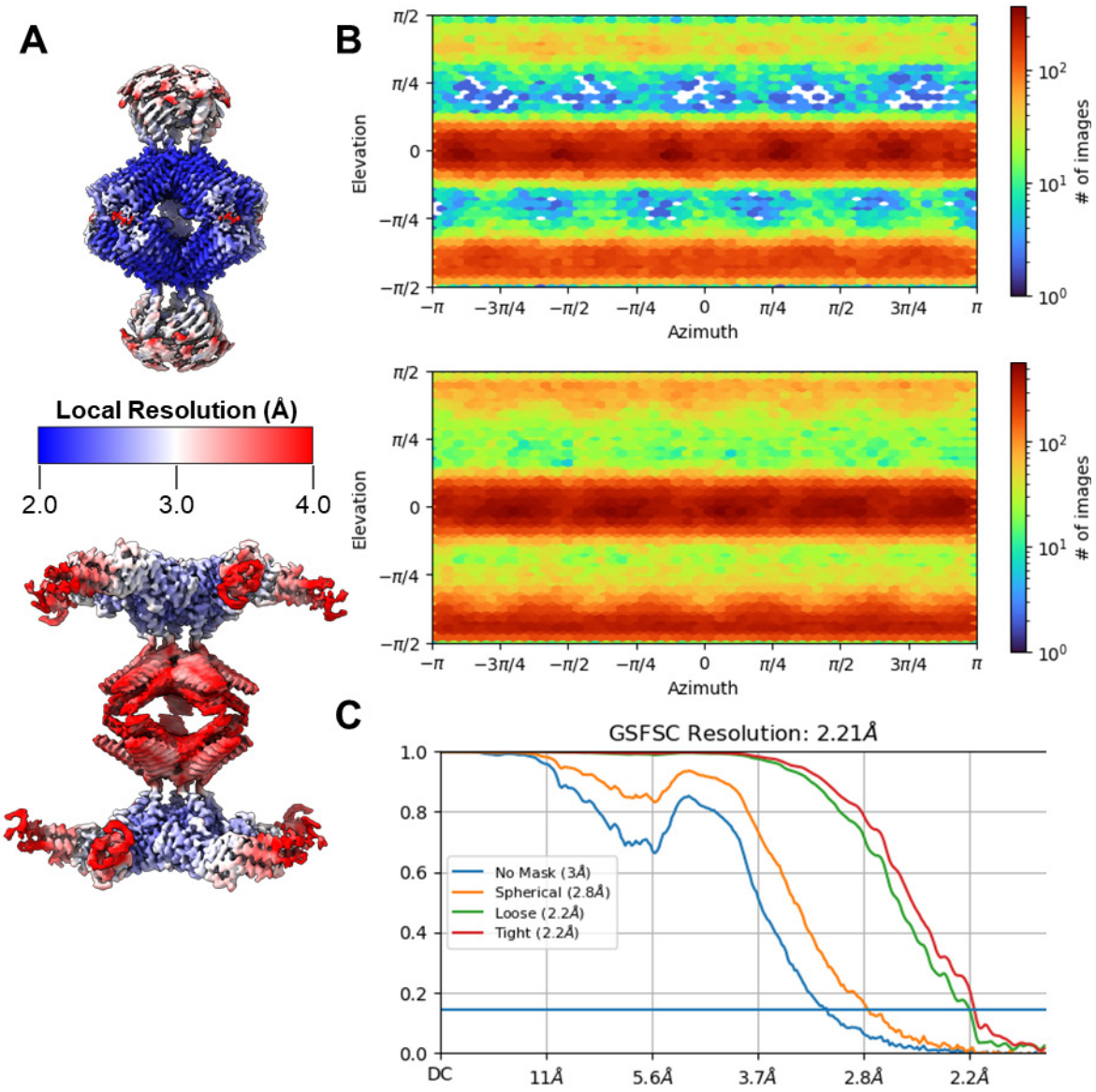


Figure 3.11: KCTD9/CUL3 volume maps summary A) Local resolution estimation maps coloured by local resolution (red-white-blue). B) CryoSPARC View Direction Distribution 2D plot. Top: Map derived from Topaz picking (oblique views missing), Bottom: Homogenous refinement map derived from template picking. Coloured by number of images present (rainbow). C) CryoSPARC Gold Standard Fourier Shell Correlation (GSFSC) 2D plot of the KCTD9^{PPR} domain local refinement map. GSFSC resolution estimate at CC: 0.146.

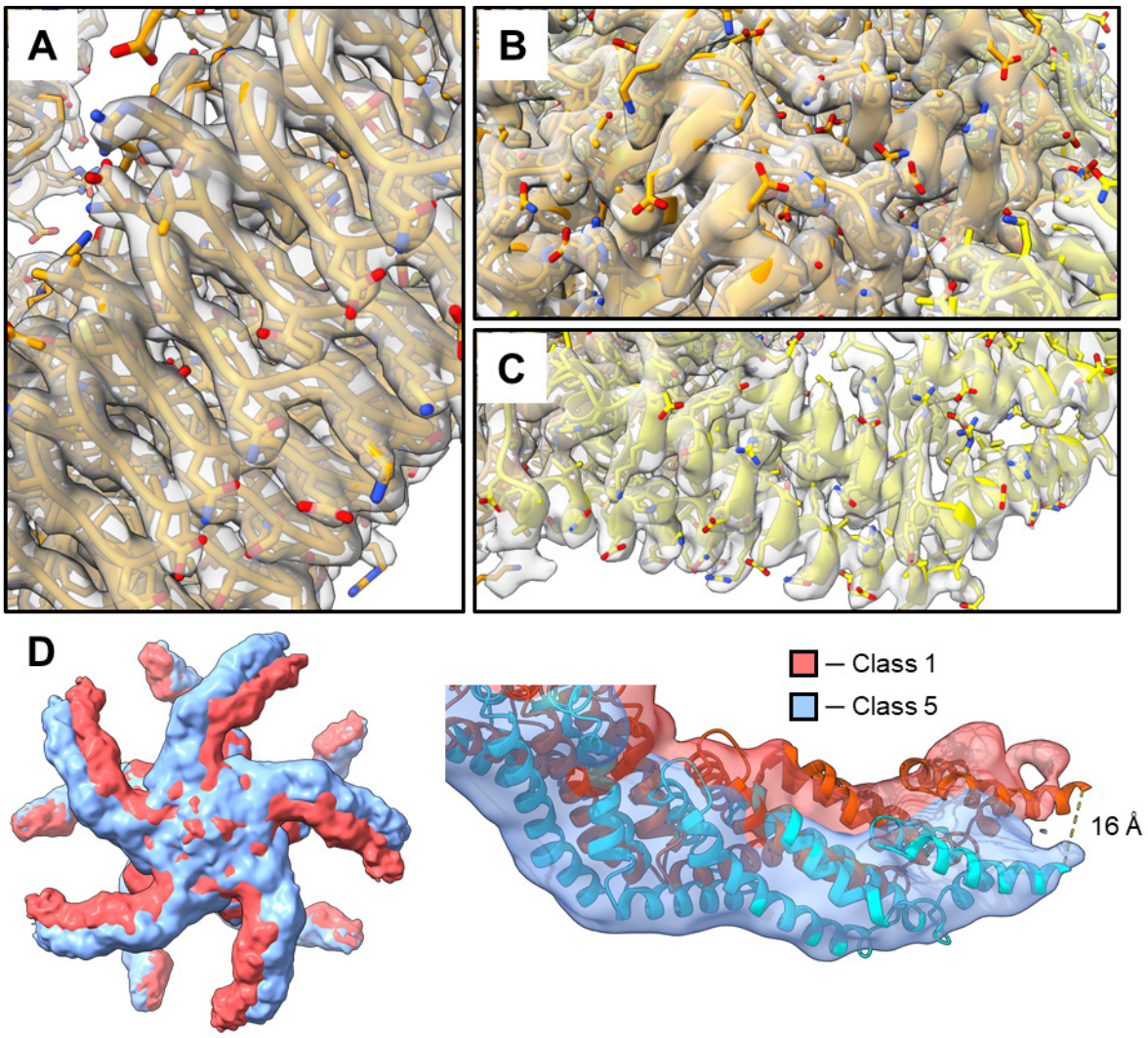


Figure 3.12: A) KCTD9 PPR domain region model in KCTD9 only map (threshold 0.1). B) KCTD9 BTB domain region model in $KCTD9^{BTB}/CUL3^{NTD}$ local resolution map (threshold 0.1). C) $CUL3^{NTD}$ region model in $KCTD9^{BTB}/CUL3^{NTD}$ local resolution map (threshold 0.08). KCTD9: Orange, CUL3: Yellow, Sidechain heteroatoms: coloured by element. D) Relion 3D classification of KCTD9/CUL3 dataset. Superimposition of class 1 and class 5 (coloured red and blue respectively).

	KCTD9A-c047 Only Dataset	KCTD9A-c047/CUL3A-c012 Dataset	
Data Collection and Processing			
Electron Microscope Model	Krios	Krios	
Voltage (kV)	300	300	
Detector	Falcon 4	K3	
Facility	OPIC[a]	eBIC[b]	
Nominal Magnification (X)	130,000	130,000	
Pixel Size (Å/px)	0.932	0.645	
Total Exposure Dose ($e^-/Å^2$)	50.00	50.02	
Fractions	50	50	
Defocus Range (μm)	-2.2 to -0.8 (0.2 increments)	-2.2 to -0.8 (0.2 increments)	
Micrographs (no.)	4,551	13,310	
Initial Particle Images (no.)[c]	237,982	268,000	
Final Particle Images (no.)	412,793	368,466	
Imposed Symmetry	D5	D5	
Map Name	P280-J329	P281-J409	P281-J412
Description	Homogenous Refinement Map	$KCTD9^{PPR}$ Local Refinement Map	$KCTD9^{BTB}/CUL3^{NTD}$ Local Refinement Map
Map Resolution (Å)[d]	2.28	2.21	2.76
Map Resolution Range (Å)	2.10-34.77	2.17-36.10	2.28-46.30
CC (Mask)	0.90	0.89	0.86
Refinement			
Initial Model Used	Alphafold3	Alphafold3	
Map Sharpening B Factor ($Å^2$)	67.9	55.3	78.8
Chains	10	10	20
Atoms	18,500	19,090	28290
Residues	2420	2500	3490
Bond (RMSD)			
Length (Å) (>4σ)	0.002 (0)	0.002 (0)	0.002 (0)
Angles (º) (>4σ)	0.480 (0)	0.492 (0)	0.472 (0)
Molprobity Score	1.30	1.26	1.71
Clash Score	5.60	4.89	7.48
Ramachandran Plot (%):			
Outliers	0	0	0
Allowed	1.68	1.63	2.05
Favoured	98.32	98.37	97.95
Ramachandran Plot Z-score:			
Whole (RMSD, N)	-0.93 (0.16, 2380)	-1.48 (0.16, 2460)	0.18 (0.13, 3410)

Helix (RMSD, N)	0.57 (0.25, 440)	-0.83 (0.22, 500)	0,51 (0.09, 2370)
Sheet (RMSD, N)	- (-,0)	- (-, 0)	-0.03 (0.53, 100)
Loop (RMSD, N)	-0.98 (0.13, 1940)	-1.10 (0.14, 1960)	-0.63 (0.19, 940)
Rotamer Outliers (%)	0.99	0.48	2.37
Box Dimensions:			
Length (Å)	120.23, 117.43, 179.88	120.94, 117.07, 180.92	198.34, 188.86, 208.01
Angles (º)	90, 90, 90	90, 90, 90	90, 90, 90

[a]OPIC: Oxford Particle Imaging Centre, Oxford.
[b]eBIC: electron Bio-Imaging Centre, Didcot (BI34631-3).
[c]Initial selected particles after cryoSPARC blob picking and 3 cycles of 2D classification (10% of data).
[d]GSFSC estimated resolution (FSC threshold = 0.143).

Table 3.5: KCTD9 cryo-EM model statistics. Calculated by Phenix (Molprobity) validation tool and cryoSPARC.[118]

3.2.3.2 KCTD9 Structural Observations

The two cryo-EM structures display KCTD9 in an unexpected homo-decameric state (or hetero-eicosameric state if CUL3 is also included); where two pentameric assemblies, chains A-E and chains F-J of KCTD9, assemble through interactions at the C-terminus of the PPR domain. The interactions of the KCTD9 protomer can be separated into three distinct sets: 1) between adjacent subunits in a KCTD9 pentamer (e.g. between chains A and B). 2) between KCTD9 and CUL3. 3) the unexpected interactions at the C-terminus of the KCTD9 (e.g. between chains A and F), which are responsible for the higher ordered oligomeric state. The buried surface area (BSA) of the structures were calculated using the PISA server to quantify which residues are at the interfaces of these subunits.[134]

3.2.3.2.1 Interactions Between Adjacent KCTD9 Subunits

Each KCTD9 monomer interacts with the two adjacent monomers through both the BTB domain and the N-terminal of the PPR. The interactions between the BTB domains have been previously characterised with a $KCTD9^{BTB}$ crystal structure but cryo-EM datasets allow characterisation of additional interactions between adjacent PPR domains. Where modelled, the BTB domains of the cryo-EM structures are in agreement with the $KCTD9^{BTB}$ crystal structure, with an RMSD of 1.10 and 0.86 for the structures (with and without CUL3 respectively).

The BTB domains form a series of hydrogen bonding interactions between V159 and N160 of one protomer to N164 and Y144 of an adjacent protomer (Figure 3.13C). Additionally, a salt bridge can be seen between R153 of chain A and D139 of chain B. There is a recess made by a coiled region (H132-D139) for tryptophan (W90) of the neighbouring unit to be situated within and provide hydrophobic contact with L137.

The pentapeptide repeat interactions have not been previously observed but are generally consistent with the interactions predicted by Alphafold3. The coiled and β-helix region (H197-L210) at the start of this domain provide a recess for the β-helix of an adjacent unit to be situated under the pentapeptide repeat (Figure 3.13B). The placement of these two helices relative to each other facilitates a couple of interactions: 1) hydrophobic contact of F208 and L223. 2) a salt bridge between E204 and R207 of the adjacent monomer. Several hydrogen bonding interactions are also observed between solenoid regions of PPR domain. The side chain of R219 extends deep into the cleft of an adjacent unit to interact with both D236 and T214. Additionally, H197 interactions are seen from of chain B forms a hydrogen with D229 of chain A.

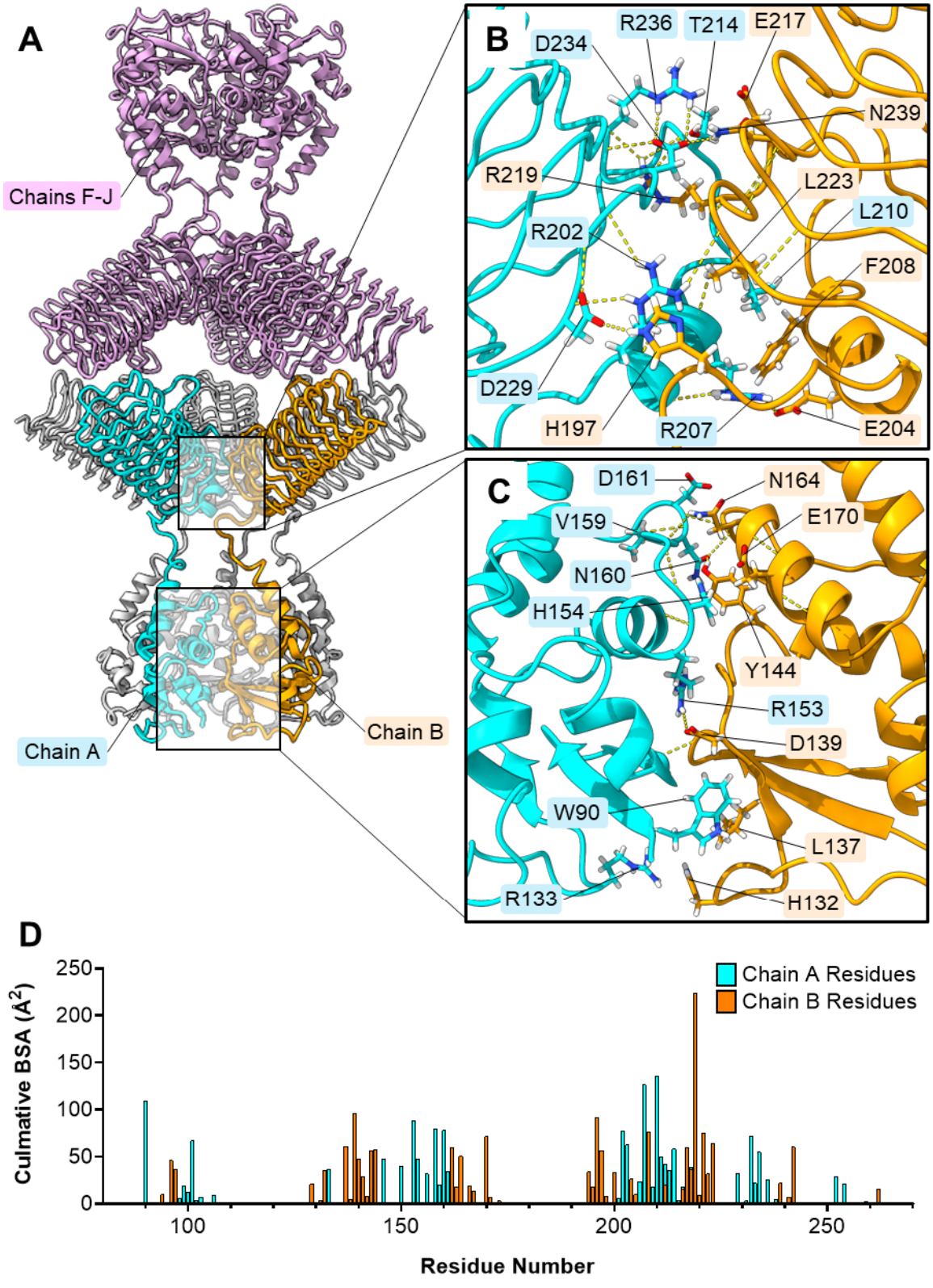


Figure 3.13: A) KCTD9 only decameric complex. Chain A (cyan), chain B (orange), chains C-E (grey), chains F-J (pink). B) Pentapeptide repeat domain interactions between chain A (cyan) and chain B (orange). C) BTB domain interactions between chain A and chain B. D) PISA buried surface area calculations for residues implicated in interactions between chains A and B.

3.2.3.2.3 KCTD9/CUL3 Interactions

The KCTD9/CUL3 structure elucidates the interactions between CUL3 and two adjacent BTB domains. At the primary interface (BSA: 779 $Å^2$,with KCTD9 chain B), the BTB domain contains negatively charged residues to complement the positively charged region (helices 1, 2, and 4) of the CUL3. The side chain of CUL3 R128 extends across to form a salt bridge with D178 and D113 of KCTD9. There are additional hydrogen bonding interactions between E55 of CUL3 to the mainchain of D139 (Figure 3.14B).

In comparison to the KCTD9 only dataset and the $KCTD9^{BTB}$ crystal structure, the electron density of a variable loop (v-loop) region of the BTB domain at the primary interface is resolved. This v-loop region is critical for CUL3 recruitment, despite lacking the φ-X-E sequence motif that is found in the long-form dimeric BTB-containing proteins (such as the KLHL/SPOP families).[76] In this structure, it is observed that the tryptophan side chain from KCTD9 (W126) is deeply buried between a loop connecting H1 to H2 and helix 4. The Prive group demonstrated a V125A mutant restricted the ability of the v-loop to adopt a conformation that provides optimal interactions with CUL3, and so partially disrupted the KCTD9/CUL3 complex.[55]

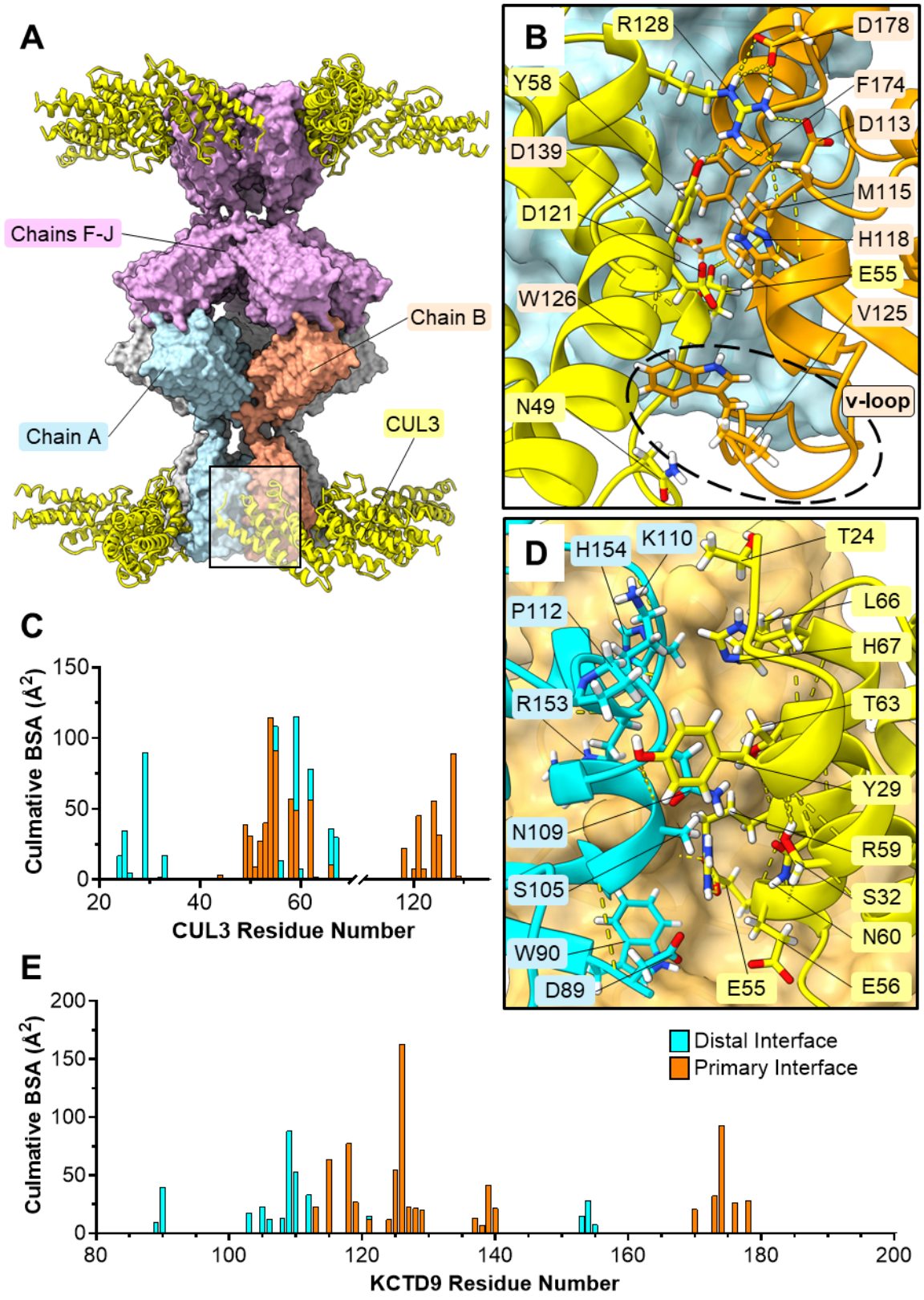


Figure 3.14: A) KCTD9/CUL3 decameric complex. CUL3 (yellow), Chain A (cyan), chain B (orange), chains C-E (grey), chains F-J (pink). B) Primary interface interactions with CUL3. Distal interface (light blue), v-loop circled. C) PISA culmative BSA calculations of CUL3 residues at the KCTD9/CUL3 interfaces. Primary interface interactions (orange) and distal interface interactions (cyan). D) Distal interface (KCTD9 chain A) interactions with CUL3. Primary interface (orange). E) PISA culmative BSA calculations of KCTD9 residues at the KCTD9/CUL3 interfaces.

At the distal surface (with KCTD9 chain A), interactions are somewhat less extensive, and there is an overall reduced contact surface area (BSA: 341 $Å^2$, Figure 3.15D). The side chains of a short helix of KCTD9 (R104-P112) slot between helices 1 and 2 of CUL3, but there are no hydrogen bonding interactions. Helix 2 of CUL3 (E56-L66) is wedged between both the primary and the distal interfaces; with the side chains of E55, R59, and L66 contacting both the primary and distal surface of adjacent KCTD9 subunits (Figure 3.14D).

3.2.3.2.4 Homo-decameric State Interactions

An unexpected homo-decameric state is observed where two of the expected KCTD9 pentameric complexes assemble together and form several hydrogen bonding interactions at the C-terminus of the pentapeptide repeat domain (Figure 3.15A). Firstly, the side chain of R366 one KCTD9 unit extends across (and above) the adjacent repeat domain to interact with both L360 and A363. The amide group of N364 forms a hydrogen bond interaction with the carbonyl of E362. Finally, the side chain of E362 extends downwards to interact with the extended side chain of R326 (Figure 3.15B). Although the decamer interface is much smaller than the other KCTD9/KCTD9 and KCTD9/CUL3 interfaces (BSA: 354 $Å^2$, Figure 3.15D), six of the seven contact residues are participating in hydrogen bond interactions.

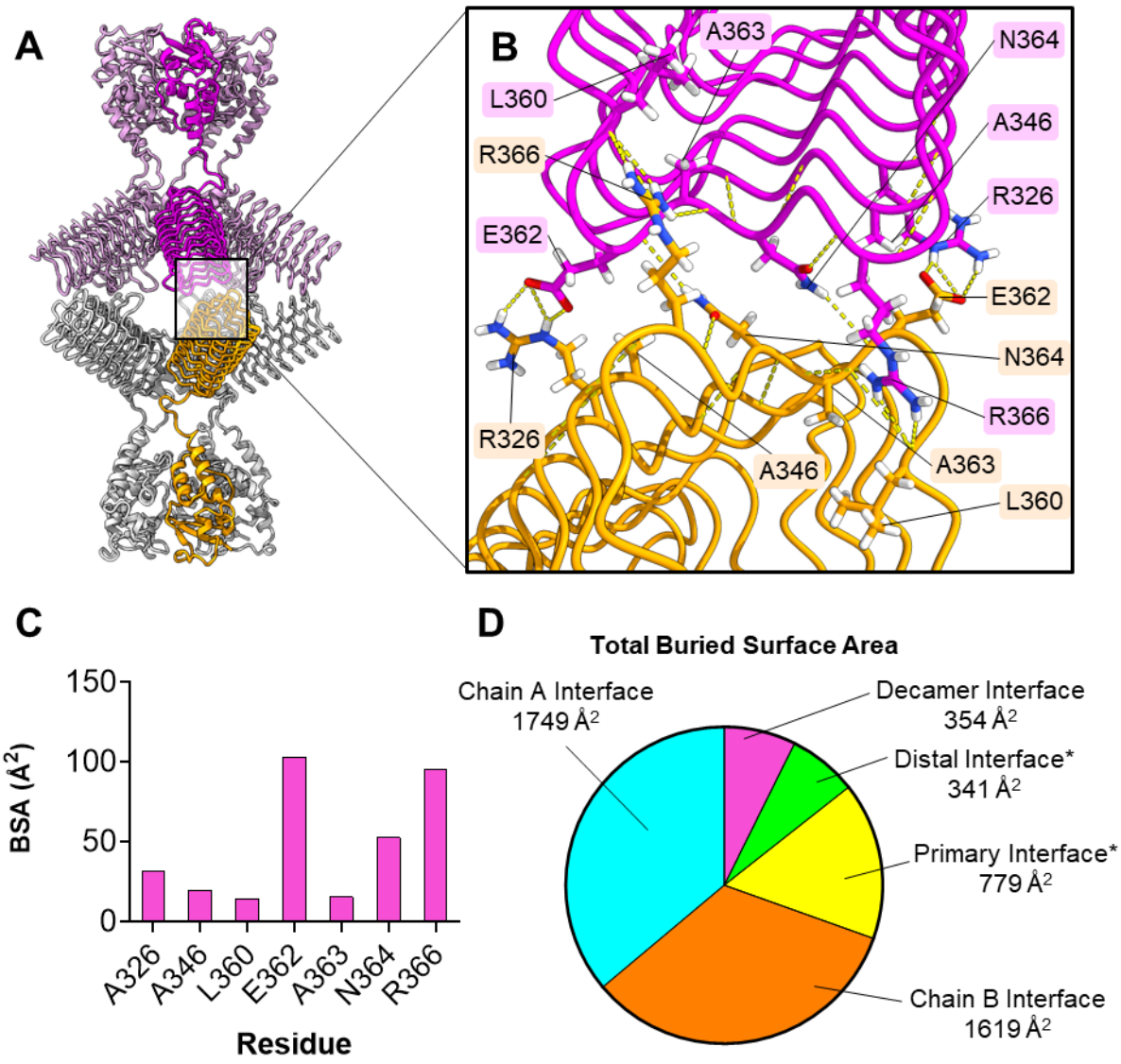


Figure 3.15: A) KCTD9 only decamer complex. Chain A (orange), chains B-E (grey), chains F (magenta), chains G-J (pink). B) Decameric interactions between chain A and chain B. C) PISA BSA calculations for KCTD9 residues at the decameric interface. D) Total BSA of each interface for a single KCTD9 chain. *KCTD9/CUL3 interfaces.

The cryo-EM map diminishes beyond E376 of KCTD9 and the remaining 13 residues of the C-terminus could not be resolved. Alphafold 3 predicts that these missing residues could adopt C-terminal capping motif structure. If this model was superimposed onto the decameric model, there would be clashes of the C-terminal residues (Figure 3.16B). It is worth noting that at this region, the Alphafold model pDDLT (confidence) score is lower than for the rest of KCTD9 and that C-terminal capping motif varies between predictive models (Figure 3.16A).

I hypothesize that the final few residues are more intrinsically disordered and that capping motif helix is not present in KCTD9, and that it can only be resolved when stabilised by an interactor. This type of behaviour is already present at the v-loop section of the BTB domain, which is dynamic when uncomplexed but becomes resolved upon binding to CUL3, as observed in the KCTD9/CUL3 complex dataset. Additionally, I identified a truncation site within the C-terminal residues of KCTD9 (H383-M384), and intrinsically disordered regions are more susceptible to proteolytic cleavage.

3.2.3.2.2 The Pentapeptide Repeat Domain

The PPR domain adopts the uncommon four-sided solenoid structure that is expected. (Figure 3.17). Each coil is made up of 20 residues (with the exception of coil 5) and each pentapeptide motif corresponds to one face of the solenoid. Face 1 and 4 make up the inner surface of the decameric complex, while face 2 and 3 are outward facing. At the midpoint of the domain (coil five), the regular solenoid structure is disrupted to accommodate a distinct five residue extension from the corner of face 1 and 4. This extension points towards the centre of the decameric complex, and towards the direction of another extension of a neighbouring PPR domain. The distance between the two is too large for any interactions.

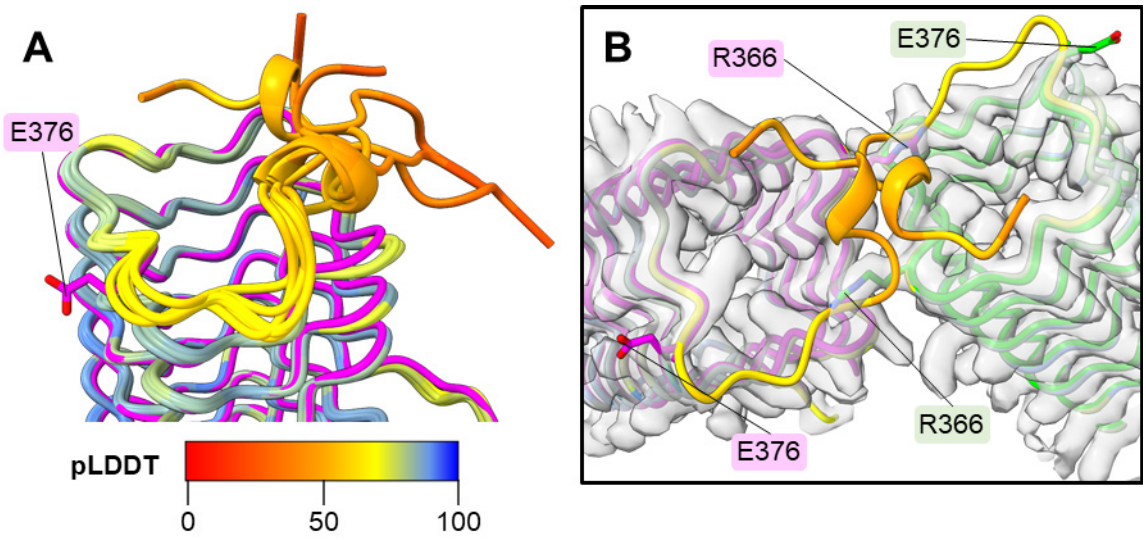


Figure 3.16: A) Overlay of KCTD9 (magenta) C-terminal domain region with multiple Alphafold3 models of KCTD9^{PPR}. The models are coloured by pLDDT score (confidence score). B) Super imposition of full-length KCTD9^{PPR} Alphafold3 model with the observed decameric species and cryo-EM volume map (threshold: 0.1).

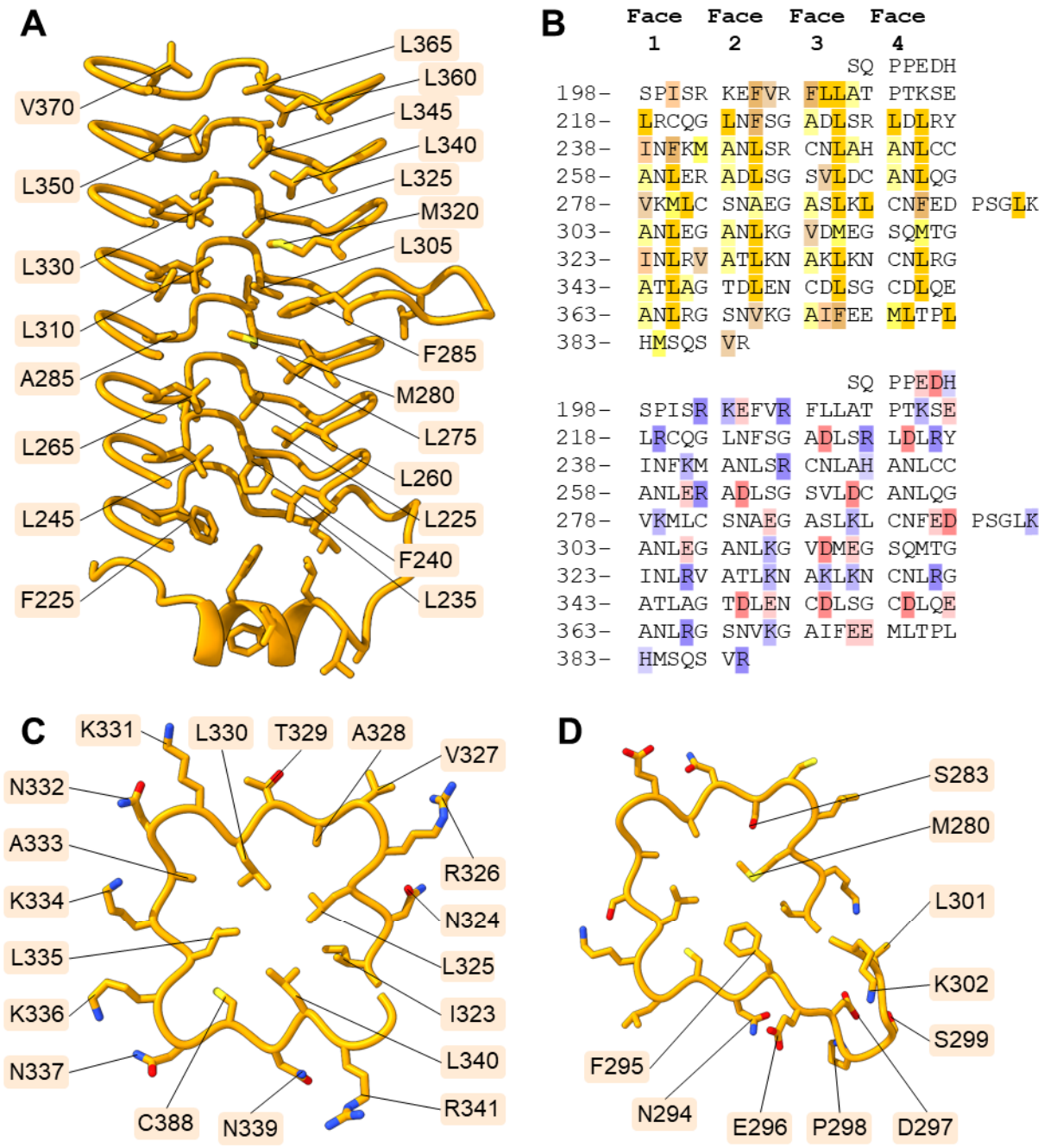


Figure 3.17: A) Cut through image of the pentapeptide repeat domain with residues contributing to the hydrophobic core shown. B) KCTD9 pentapeptide repeat domain sequence. Non-polar residues and charged residues are highlighted respectively. C) Standard single coil of a pentapeptide repeat (coil 6). D) Structure of rung 5 which contains a 5 amino acid extension.

The core of the solenoid structure lacks a channel for water molecules and is instead occupied by hydrophobic sidechains from the third residue of the pentapeptide motif. Often this residue is leucine but in sections that deviate from the consensus pentapeptide motif, phenylalanine and methionine are also observed at this position. The second and forth residues of the PPR motif are outward facing and contribute to the surface charges of the solenoid faces. Face 1, and thus the inner surface of the KCTD9 complex have a greater proportion of positively charged residues. In contrast, the central regions of face 2, 3 and 4 are slightly negatively charged, resulting in a negatively charged outward facing surface of the complex (Figure 3.18).

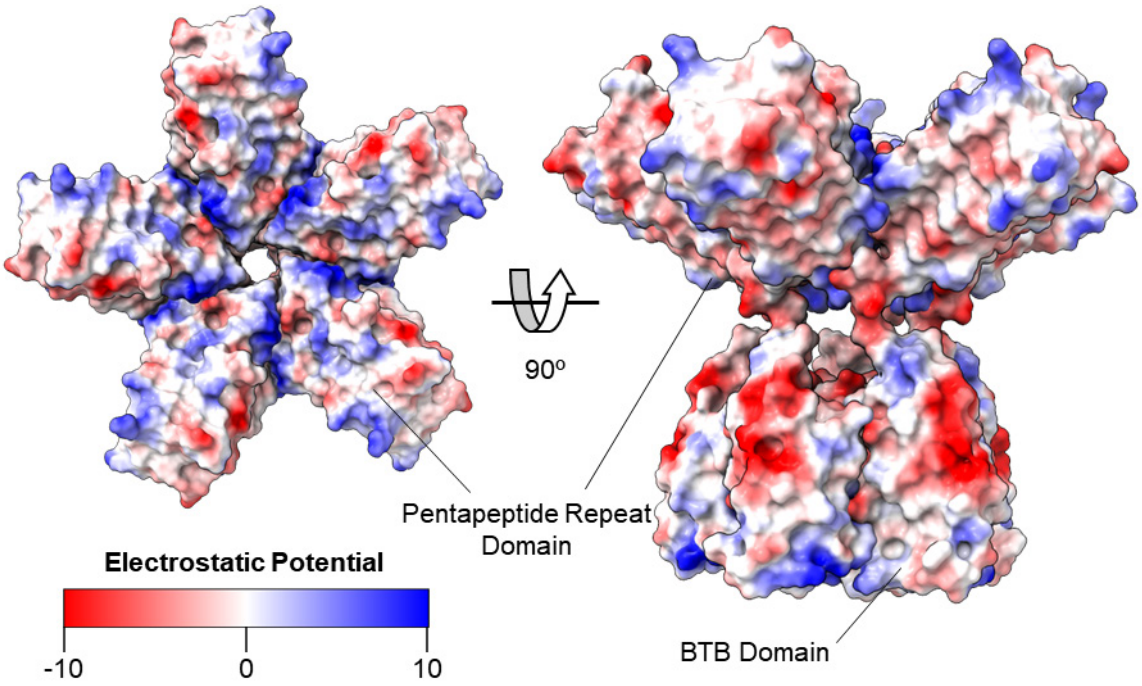


Figure 3.18: KCTD9 surface coloured by electrostatic potential. Chains F-J hidden.

3.3 Discussion

Determining the structures of Cullin-RING E3 ligases can help us to understand their mechanism of action and what interactions are involved in substrate recognition. I was able to generate constructs through *E. coli* and insect cell expression to be used for the structural studies. X-ray crystallography and cryo-EM, the two predominant techniques currently used in structural biology, were used to generate three datasets that were further analysed using servers, such as PISA and DALI, to gain insight into the structure of KCTD9.

From a high yielding construct (KCTD9A-c073), I was able to generate a 2.2 Å X-ray crystallography dataset. This presence of this Ub-like domain marks KCTD9 out from the rest of KCTD family and its role still remains unclear. Although I've shown it to have a Ub-like fold, it has vastly different surface properties to that of ubiquitin, NEDD8, or the Ub-like domains found in doublecortin.

Perhaps the role of this Ub-like domain is to localise the KCTD9 cullin-RING assembly to a specific region of the cell and so regulate its ubiquitination activity? The N-terminal copy of the tandem ubiquitin-like domains found in doublecortin associate with microtubules and anchors the rest of the protein there in order to promote microtubule polymerisation during neurogenesis.[135,136]

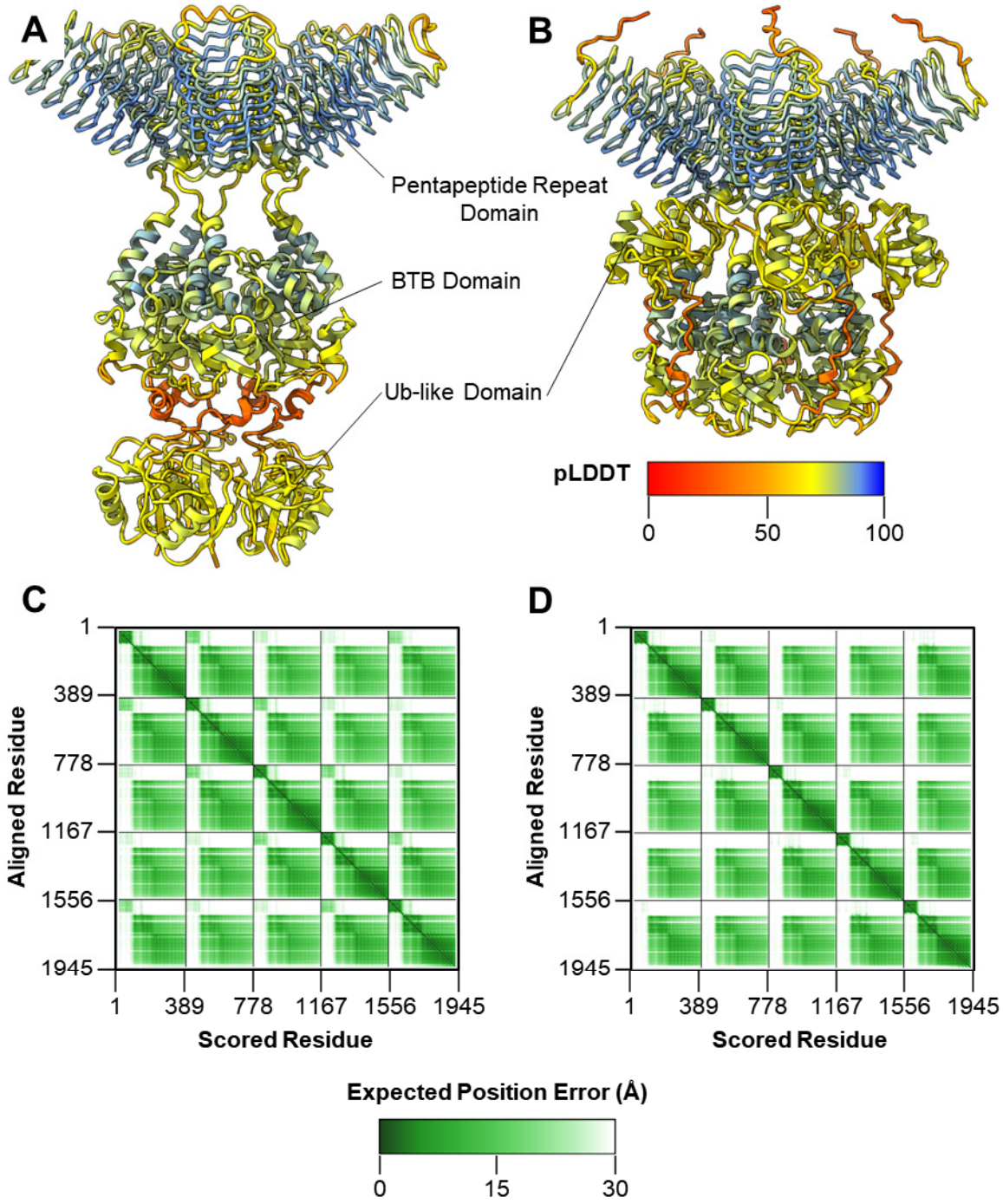


Figure 3.19: Different Alphafold3 predictions of a KCTD9 pentameric complex. A) Model 1: The Ub-like domains are positioned beneath the BTB domains. Coloured by pLDDT. B) Model 2: The Ub-like domains are positioned between the BTB and PPR domains. C) 2D PAE graph of model 1. Coloured by expected position error. D) 2D PAE graph of model 2.

The relative position of the Ub-like domain to the rest of KCTD9 complex is another mystery that remains to be solved. Disorder prediction suggests that the linker region between the Ub-like domain and the BTB domain is intrinsically disordered and this could allow wide range of movement the Ub-like region. Indeed, some Alphafold3 models place the Ub-like domain below the BTB pentamer, while other models predict it between the BTB and PPR domains. In either prediction, the 2D PAE graphs suggest there is low confidence in the placement of this domain.

I produced two cryo-EM datasets of ~2.5 angstrom resolutions using a KCTD9 construct expressed in *Sf9* insect cells (KCTD9A-c047) with and without CUL3A-c012. These datasets allowed the characterisation of a KCTD9 only and KCTD9/CUL3 complexes in an unexpected higher order oligomeric state; with additional hydrogen bond interactions at the C-terminus of the PPR domain. I was able to resolve the majority of the unusual solenoid structure of the pentapeptide repeat domain, except for the last 13 residues and the fused FLAG tag. The PPR domain is uncommon in eukaryotes and, according to the InterPro database, KCTD9 is the only example in *Homo sapiens* where this domain is found.

The interactions between the adjacent BTB domains in KCTD9 had been previously characterised by KCTD9^{BTB} crystal structure, but the cryo-EM datasets allowed the interactions between the N-terminus of the PPR domain to be determined. The PPR domains are observed to partially sit underneath an adjacent PPR domain, and this results in a fanning out of the solenoid structures. The distribution of charge residues across the PPR domains makes the inner surface of the decameric complex positively charged and the outer surface relatively negatively charged. In the KCTD9/CUL3 dataset, the variable loop

region of the BTB domain was able to be resolved, and this showed the side chain of W126 to be buried into the CUL3 surface. The Prive group had shown that a KCTD9^{V125A} mutant hindered the ability of the BTB domain to bind to CUL3 by preventing the variable loop domain from adopting an optimal conformation for binding. Based on my structural observations, I predict that a W126A mutant could show a similar or greater reduction in binding as this residue forms a much greater hydrophobic interaction with a larger BSA.

The majority of the KCTD family is thought to adopt a pentameric state and the KCTD9^{BTB} crystal structure suggested that KCTD9 did also. However, in both datasets, KCTD9 was in an unexpected homo-decameric (or hetero-eicosameric in the case of the KCTD9/CUL3 complex) state with C-terminal interactions between PPR domains. Models generate by Alphafold3 using the sequence of KCTD9A-c047, the truncated species, the region of KCTD9 observed in the cryo-EM dataset (D89-E376), and the wild-type protein did not predict these interactions or the decameric state; instead non-sensical models were produced.

Potentially the construct used for this dataset (KCTD9A-c047) contained a fused C-terminal FLAG tag which could have mediated the C-terminal interactions. It is notable that the FLAG tagged KCTD9 constructs (KCTD9A-c046 to -c052) had improved expression and purification yield when compared to equivalent ‘FLAG-less’ constructs (KCTD9A-c042 to -c045). If the FLAG tag was involved in these decameric interactions, I would expect to see a part of the tag in the cryo-EM volume maps, but it is not observed. Instead, I believe that the charged FLAG tag is shifting the pI of the constructs (KCTD9A-c043: pI = 6.16, KCTD9A-c047: pI = 5.96) and ultimately increasing the solubility. Additionally, the extension provided

by the FLAG tag is somewhat protecting the construct from the observed C-terminal truncation, and this also aids the stability of the constructs.

When a KCTD9 pentameric model generated by Alphafold3 is duplicated and imposed onto the observed structure, the predicted C-terminal capping motif clashed with adjacent chains. As these residues were not observed in the cryo-EM datasets, the pLDDT score for this region is low, and Disopred3 disorder prediction score is high, I hypothesise that this capping motif is incorrectly predicted and instead this region is intrinsically disordered. Potentially it is stabilised by interaction partner in an analogous manner as the v-loop section of the BTB domain upon CUL3 binding.

Although I was unable to measure the affinity of these decameric interactions, I predict that it will be similar to KCTD9/CUL3 binding affinity (nanomolar). Despite a single KCTD9 chain having a smaller BSA to another KCTD9 protomer at the decameric interface, once assembled, the combined BSA of the decameric complex would be comparable a KCTD9/CUL3 interface. This high affinity would rationalise why no pentameric states or partially dissociated states were observed in the cryo-EM datasets. If the affinity is very strong, it might be unlikely to measure the affinity by biophysical experiments, and instead in silico experiments may need to be conducted instead.

I believe that this decameric state could be representative of storage state that could regulate the ubiquitination activity of KCTD9. The Schulman group showed that KLHDC2, a CUL2-dependent BTB-containing E3 ligase, adopts a homo-tetrameric storage state and the substrate binding site is occupied by a degron mimic from the C-terminus of an adjacent subunit (Figure 3.20).[137] This inactivate

state undergoes slow dissociation to an active monomeric state that is then either stabilised by the binding of the native substrate, or self-associated back to the tetramer. This system serves to regulate the E3 ligase activity and control ubiquitination activity through a process referred to as kinetic amplification.

Although I think the specific interactions that regulate KCTD9 are different, the storage state of KLHDC2 system could be representative of similar state that KCTD9 adopts. Based on comparisons to KCTD5/Gβγ complex, I hypothesize that the PPR domain is responsible for substrate recognition and that the decameric state could block the substrate binding site.

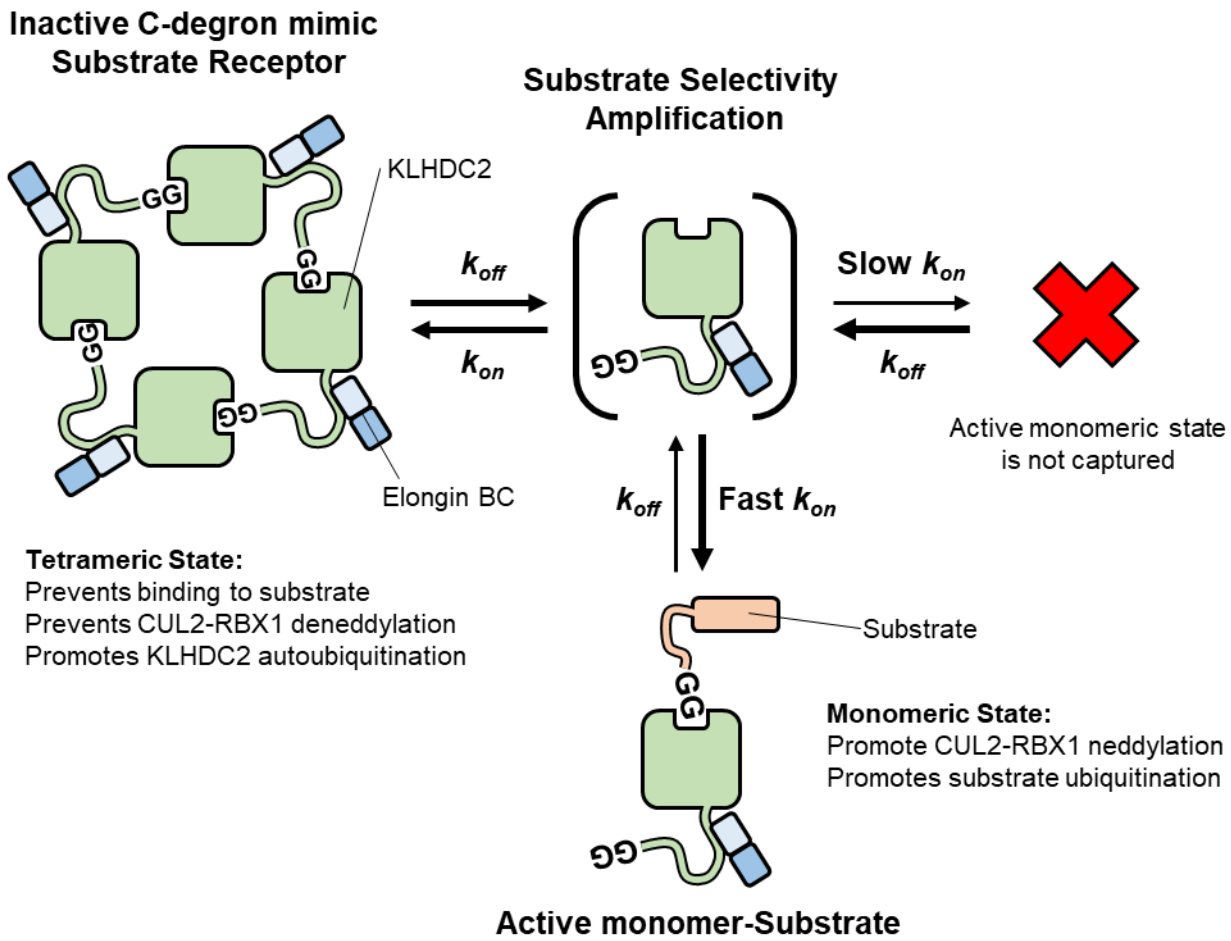


Figure 3.20: KLHDC2 tetrameric storage state to active monomer state by kinetic amplification. Mechanism diagram sourced from D.C. Scott *et al.* and adapted.[137]

Dissociation of this decameric state to an active pentameric state may require an unknown activation step, or could be concentration dependent. As the KCTD9 samples were at micromolar concentrations (higher than those thought to be biologically relevant), interactions that may occur vanishingly at cellular levels could instead occur much more readily at these elevated levels. To test this hypothesis in future work, in solution measurements of the KCTD9 complex using methods such SEC-SAXS (Size-exclusion chromatography small-angle X-ray scattering), AUC (Analytical ultracentrifugation), or Mass Photometry would be the most appropriate. These experiments could confirm that the decameric state is present in solution, and investigate the concentration dependence of this state. Indeed, I tried both SEC-SAXS and mass-photometry measurements of the KCTD9A-c047, however in both experiments I was unable to generate reliable, reasonable data due to poor stability of the construct during measurement.

The structural datasets presented in this chapter provide preliminary insights into how the whole of KCTD9 is structured and the interactions it participates in (with itself and CUL3). Further investigation of KCTD9, in complex with interactors identified by MS proteomics or alternative methods, could help us to understand its preferred recognition motifs and ubiquitination activity. The decameric state seen in the cryo-EM data raises the possibility that KCTD9 possess a storage state that is responsible for the regulation of its ubiquitination activity; the nature of which needs further investigation.

Chapter 4: Exploration for Novel KCTD9 Interactors

4.0 Exploration for Novel KCTD9 Interactors

4.1 Introduction

The role of KCTD9 is not understood but several studies have implicated this substrate adaptor in the innate immune system.[63,71,72] A KCTD9 knockdown study resulted in reduced natural killer (NK) cell development and effector function, but the interacting partners in this process are unknown. Additionally, it is unclear if KCTD9 is acting as a typical cullin-3 E3 ligase, or in an alternative manner within NK cells.

Typically in the KCTD family, the variable CTD domain is responsible for substrate interactions and this is exemplified by a cryo-EM structure of KCTD5/CUL3/Gβγ complex (Figure 4.1A).[45,46] Five Gβγ substrates associate with a pentameric KCTD5 complex, with each substrate interacting with an individual C-terminal domain of KCTD5 (Figure 4.1A). I hypothesis that KCTD9 also recognises their interactor(s) through the C-terminal pentapeptide repeat (PPR) domain in either a 5:5 stoichiometry, as seen for KCTD5, or a 1:5 stoichiometry; where a single interactor associates with a pentameric KCTD9.

The PPR domain has a distinct solenoid structure and it is uncommon in eukaryotic species, with KCTD9 being the only example of this domain in *Homo sapiens*.[65,66] Examples of the PPR domains in other species could the inform the types of interactions KCTD9^{PPR} could be involved in and potentially indicate the nature of any interactors. MfpA adopts a dimeric (end-on-end) solenoid structure to mimic T-segment DNA and bind DNA gyrase (PDB: 6ZT5, Figure 4.1B).[68] Could KCTD9^{PPR} mimic polynucleotide chains and interact with DNA/RNA binding proteins? Alternatively, the PPR domain of bacterial E3 ligase SopA has shown

to complex with the RING domain of homo sapien TRIM56 (PDB: 5JW7, Figure 4.1C).[127] Perhaps the RING domain of a human TRIM protein could be a interactor of KCTD9?

Finally, during these investigations, a study identified ZNT9 (Zinc transporter 9) as an interactor of KCTD9 by mass spectrometry experiments in HCT-116 cells.[73] The researchers demonstrated that the PPR domain of KCTD9 was able to bind to the N-terminal region of ZNT9 (M1-E200) and compete with β-catenin for binding by immunoblotting. The researchers did not identify any ubiquitination of ZNT9 by KCTD9, but instead demonstrated β-catenin was prone to proteasomal degradation when it was not able to bind with ZNT9.

In this chapter, I will verify potential interactors (TRIMs and ZNT9) that were cited in literature by performing pulldowns experiments using recombinant protein constructs. Additionally, I will try to identify novel interactors of KCTD9 through affinity mass spectrometry (MS) proteomics experiments. An initial pilot dataset was generated using HEK293 cells, and then a larger scale proteomics experiment was conducted in NK92mi cells. The data from the latter experiment is still undergoing analysis, but I have postulated a few potential targets that are worth validating DAVID (Database for Annotation, Visualization, and Integrated Discovery) analysis and modelling potential complexes using the Alphafold3 server.[107]

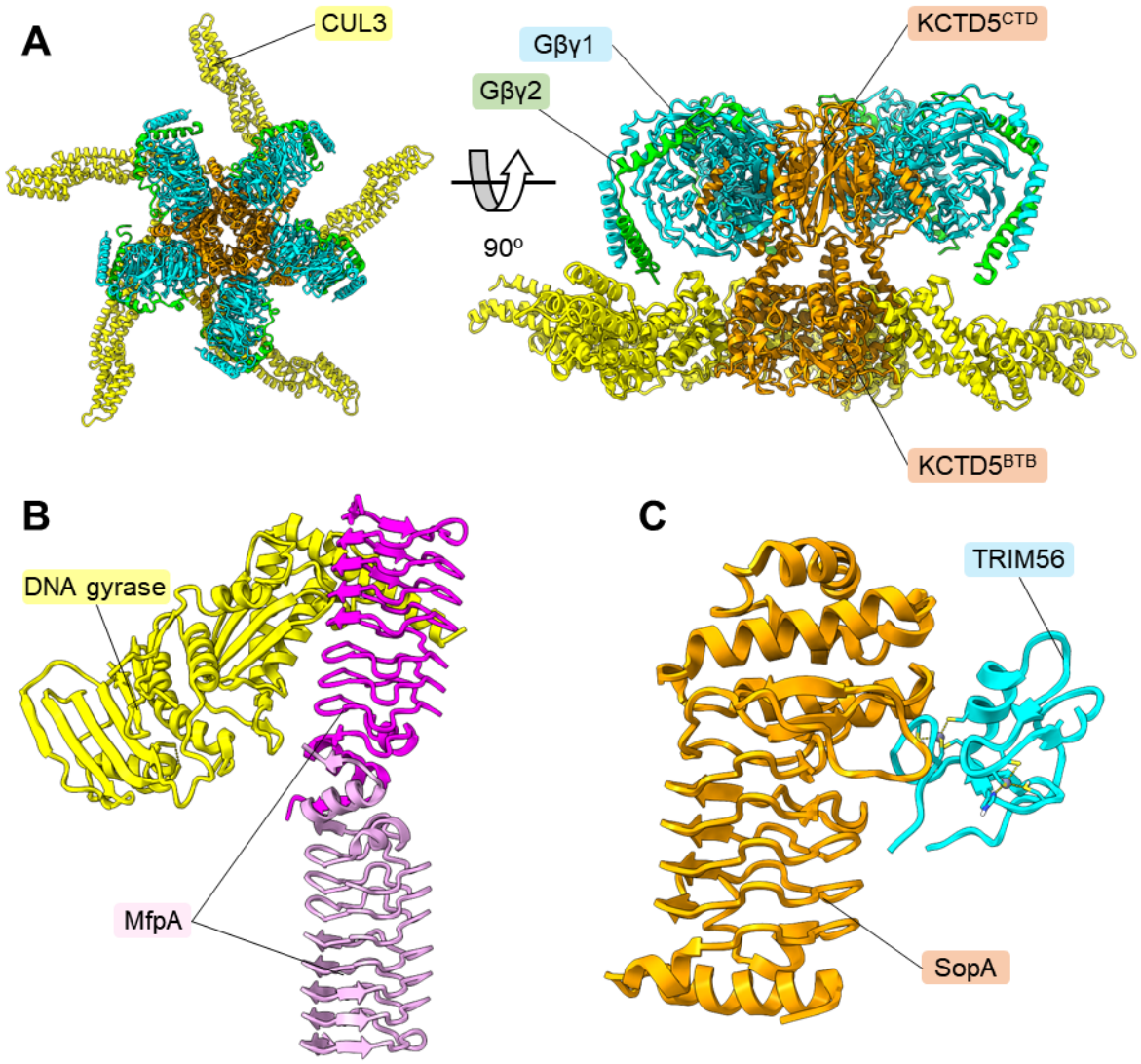


Figure 4.1: A) Pentameric KCTD5 (orange) in complex with five Gβγ (green and cyan) units and CUL3 (yellow). PDB: 8U81.[46] B) MfpA duplex (magenta and lilac) bound to a single DNA gyrase unit (yellow). PDB: 6ZT5.[68] C) SopA (orange) in complex with TRIM56^{RING} (cyan). PDB: 5JW7).[127]

4.2 Results

4.2.1 Pulldown Experiments of Recombinant Protein

4.2.1.1 Recombinant TRIM27 and TRIM32 Did Not Interact with KCTD9

A molecular interaction database (IntAct) listed 98 potential interactor proteins that were identified by yeast two-hybrid system (Y2H) screening.[138,139] In this assay, a DNA-binding domain is fused to one component of a potential protein/protein (or protein/DNA) interaction.[140] A transactivation domain is fused to the other interactor and activates a downstream reporter gene when the two components are brought into proximity by the binding of the test interactors. Without such protein/protein interactions, the reporter gene is not activated and no readout is measured. It is worth noting that these assays produce a high number of false positives, with estimates as high as 70%.[141]

Given that the pentapeptide repeat domain containing E3 ligase SopA has been co-crystallised with the RING domain of TRIM56 (PDB: 5JW7), I reasoned that any of TRIM proteins (TRIM27, TRIM32, TRIM42) identified by high-throughput Y2H assays have a greater chance of being bona fide interactors with KCTD9.[127] The RING domain constructs of these targets was designed using Uniprot, Disopred3, and Alphafold2 modelling (Table 4.1).[47,90] Additionally, TRIM56 RING domain constructs were designed to serve as a negative control for pulldown experiments; as it is not listed as an potential interactor of KCTD9 on the IntAct database. The constructs were cloned by high-throughput ligation independent cloning (LIC) and test expression in *E. coli* cells yielded soluble protein for TRIM27, 32 and 56 (TRIM42 constructs were insoluble).[93]

Construct ID (Scarab)	Construct Boundaries	N-terminal Tag	C-terminal Tag	Expression System	Test Purification Yield
TRIM27A-c001	S5-P86	His-tag	-	*E. coli*	High
TRIM27A-c002	S3-P86	His-tag	-	*E. coli*	High
TRIM27A-c003	S5-P86	-	His-tag	*E. coli*	High
TRIM27A-c004	S3-P86	-	His-tag	*E. coli*	Not Tested
TRIM32A-c001	S7-E93	His-tag	-	*E. coli*	High
TRIM32A-c002	S7-E93	-	His-tag	*E. coli*	None
TRIM42A-c001	I134-L225	His-tag	-	*E. coli*	Not Tested
TRIM42A-c002	I134-K216	His-tag	-	*E. coli*	None
TRIM42A-c003	I134-L225	-	His-tag	*E. coli*	None
TRIM42A-c004	I134-K216	-	His-tag	*E. coli*	None
TRIM56A-c001	M1-L94	His-tag	-	*E. coli*	Medium
TRIM56A-c002	M1-L94	-	His-tag	*E. coli*	Medium
ZNT9A-c006	M1-K223	His-tag	FLAG-tag	*Sf9*	Low
ZNT9A-c007	K124-K223	His-tag	FLAG-tag	*Sf9*	Low
ZNT9A-c008	M1-K223	His-tag	-	*Sf9*	Low
ZNT9A-c009	K124-K223	His-tag	-	*Sf9*	Low
ZNT9A-c010	M1-K223	His-tag	-	*E. coli*	None
ZNT9A-c011	K124-K223	His-tag	-	*E. coli*	None
KCTD9A-c067	M1-R389	3XFLAG-tag	-	*HEK293*	Not Tested

Table 4.1: KCTD9 Interactor constructs summary table.

TRIM27A-c002 (His-TRIM27$^{3-86}$, 12.2 kDa), TRIM32-c001 (His-TRIM32$^{5-93}$, 12.4 kDa), and TRIM56A-c001 (His-TRIM32^{-1-94}, 12.7 kDa) were expressed in parallel on a large scale (2 L each) and captured using Ni^{2+}-sepharose resin. The RING domains were eluted from the resin using imidazole and incubated overnight at 4ºC with purified FLAG-tagged KCTD9A-c047 (His-KCTD9$^{76-389}$-FLAG, 39.0 kDa) and CUL3A-c012 (CUL3$^{1-388}$-HIS-FLAG, 48.1 kDa). The mixture was then incubated with pre-washed Anti-FLAG® M2 affinity gel (herein referred to as FLAG resin) for 4-5 hours, and non-binding proteins were then eluted with wash buffer. The resin was boiled in sample buffer and all of the fractions were analysed by SDS-PAGE (Figure 4.2).

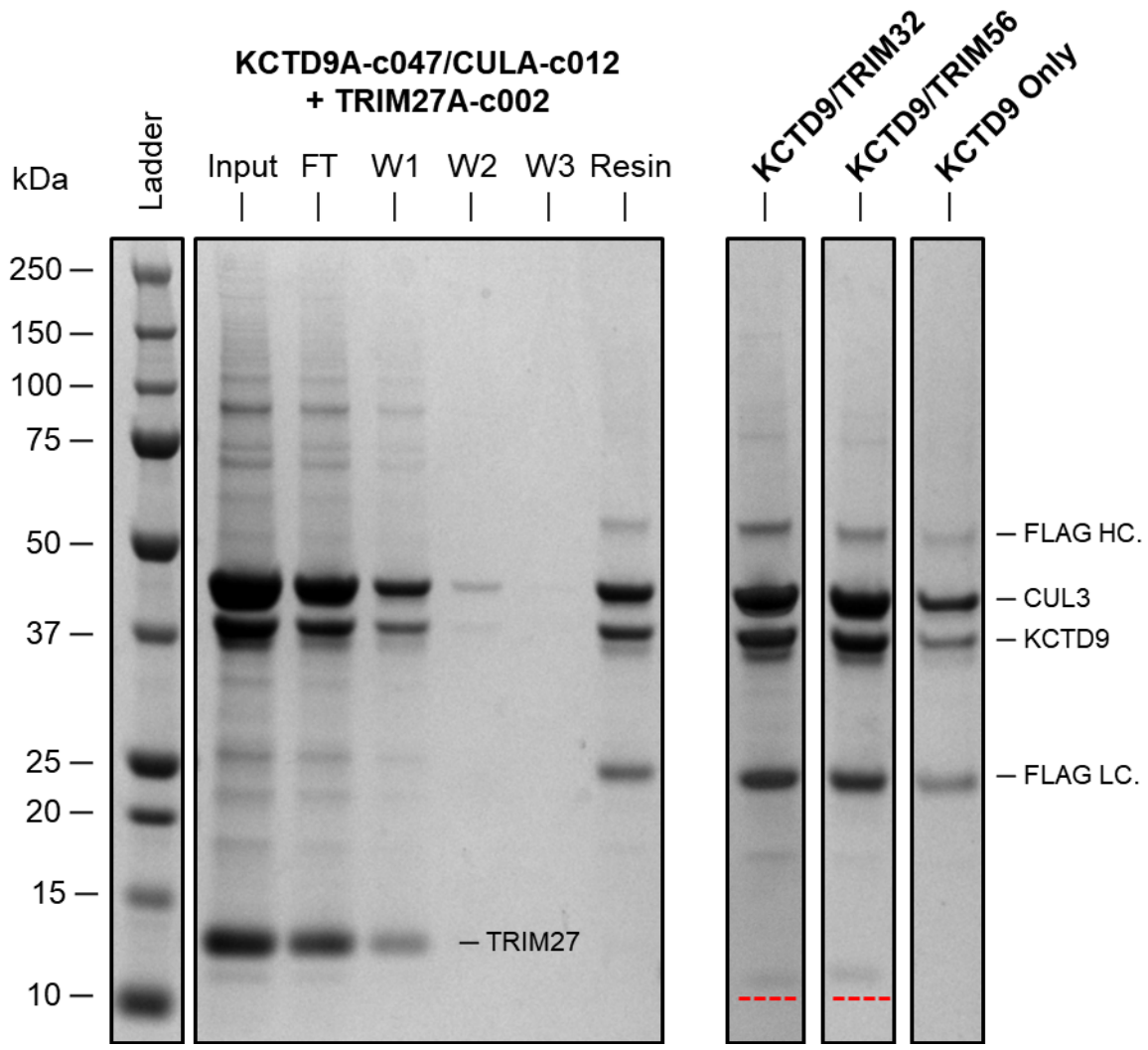


Figure 4.2: SDS-PAGE gel of KCTD9/TRIM FLAG pulldown fractions. FT: flow-through, W1-3: Washes 1-3, Resin: FLAG resin boiled in 1x sample buffer. Only the resin samples for TRIM32, TRIM56, and KCTD9 experiments are shown. Faint TRIM bands in resin samples were underscored in red.

None of the TRIM constructs were present in the resin sample and instead eluted with the wash buffer. Faint bands were observed in the KCTD9/TRIM32 and KCTD9/TRIM56 resin samples, however upon replication of these experiments, it was determined that these results were artifacts due to incomplete washing of the FLAG resin with the wash buffer. As the TRIM proteins could not be retained by the resin bound KCTD9, any interactions between the recombinant proteins was not verified.

4.2.1.2 Recombinant ZNT9 Did Not Bind to KCTD9

H. Yao *et al.* identified that ZNT9 was an interactor of KCTD9 by MS proteomics experiments conducted in HCT-116 cells.[73] The researchers then demonstrated that the N-terminal region of ZNT9 (M1-E200) specifically bound to the C-terminal region of KCTD9 by immunoprecipitation in SW620 cells. I wanted to verify these results by FLAG pulldown experiments and potentially characterise a KCTD9/ZNT9 complex by cryo-EM.

As a large proportion of the N-terminal region of ZNT9 was indicated to be disordered (Figure 4.3), several ZNT9 constructs were designed for *E. coli* and *Sf9* cell expression with either M1-K223 or K124-K223 construct boundaries (Table 4.1). The constructs were cloned by high-throughput LIC and recombinant baculoviruses for *Sf9* expression were generated by the Bac-to-Bac system. Initial test expression suggested these constructs were largely insoluble when expressed in both *E. coli* and *Sf9* cells.

To try an alternative strategy, I tried co-infecting *Sf9* cells (6L) with 3 mL of a P2 virus for KCTD9A-c042 (His-KCTD9$^{1-389}$, 45.3 kDa), and 3 mL of a P2 virus for ZNT9A-c007 (His-ZNT9$^{124-223}$, 16.4 kDa). The KCTD9 construct lacked a C-terminal FLAG tag, and had been too unstable to express/purify on previous attempts, but I wanted to investigate whether co-expression of these two constructs could stabilise and improve the yield of one another. The constructs were captured using Ni^{2+}-sepharose resin and SDS-PAGE analysis of the elution fractions indicted that only small quantities of KCTD9 were observed (Figure 4.4). Additionally, when compared to the test expression results, there was a much greater yield of ZNT9 than expected.

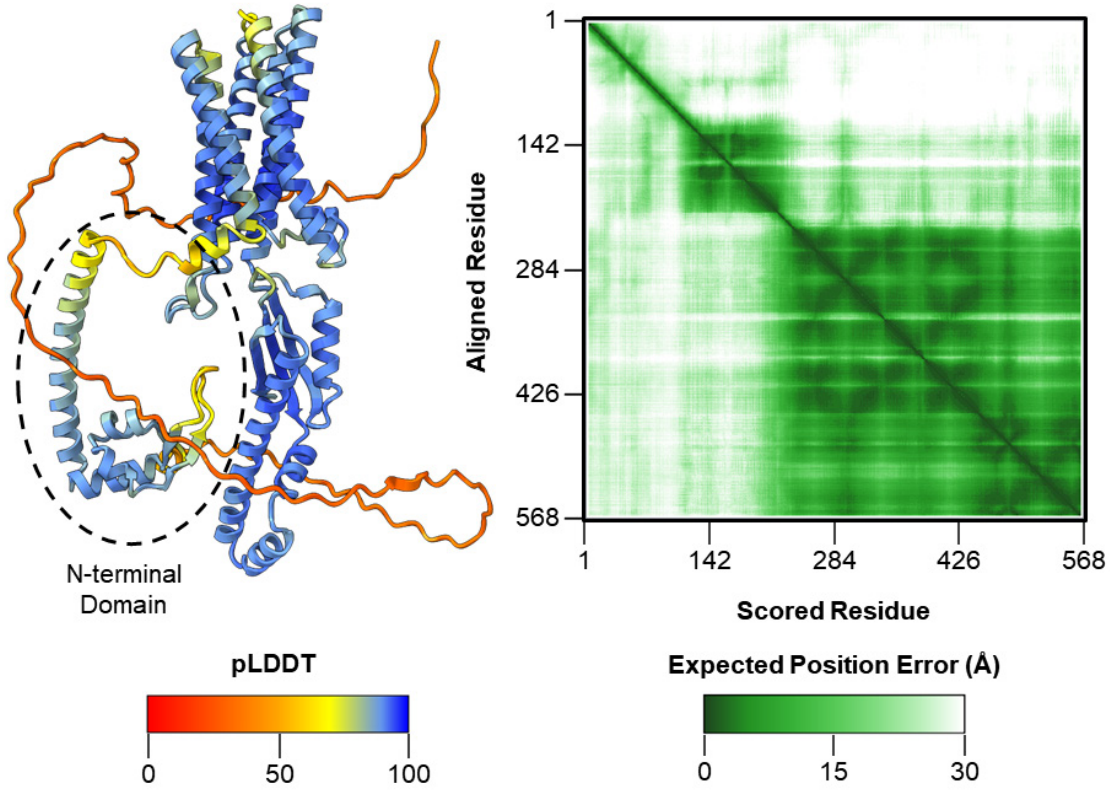


Figure 4.3: Alphafold2 model of ZNT9 coloured by pLDDT score. N-terminal domain circled. 2D PAE graph of Alphafold prediction (AF-Q6PML9-F1_v4). Coloured by PAE score.

The elution fractions containing KCTD9 and ZNT9 were pooled and an equimolar amount of CUL3A-c012 was added in an attempt to stabilise any potential complex. The mixture was concentrated, applied to a Superdex 200 Increase 10/300 GL gel filtration column, and eluted with gel filtration buffer. The elution profile displayed two major peaks: the first one corresponding to CUL3A-c012, and the second to ZNT9A-c007 (Figure 4.5A). These fractions were separately pooled and their identities confirmed by intact-MS (not shown). No KCTD9A-c042 could be identified by the chromatogram or SDS-PAGE gel of the fractions (Figure 4.5B). I suspect that this unstable construct may have been lost upon concentration and/or gel filtration, and that the presence of ZNT9 did not stabilise the KCTD9 construct. Although I was not able to produce a recombinant KCTD9/ZNT9, attempts to complex ZNT9 with a higher yielding KCTD9 construct (such as KCTD9A-c047) could potentially yield the desired complex.

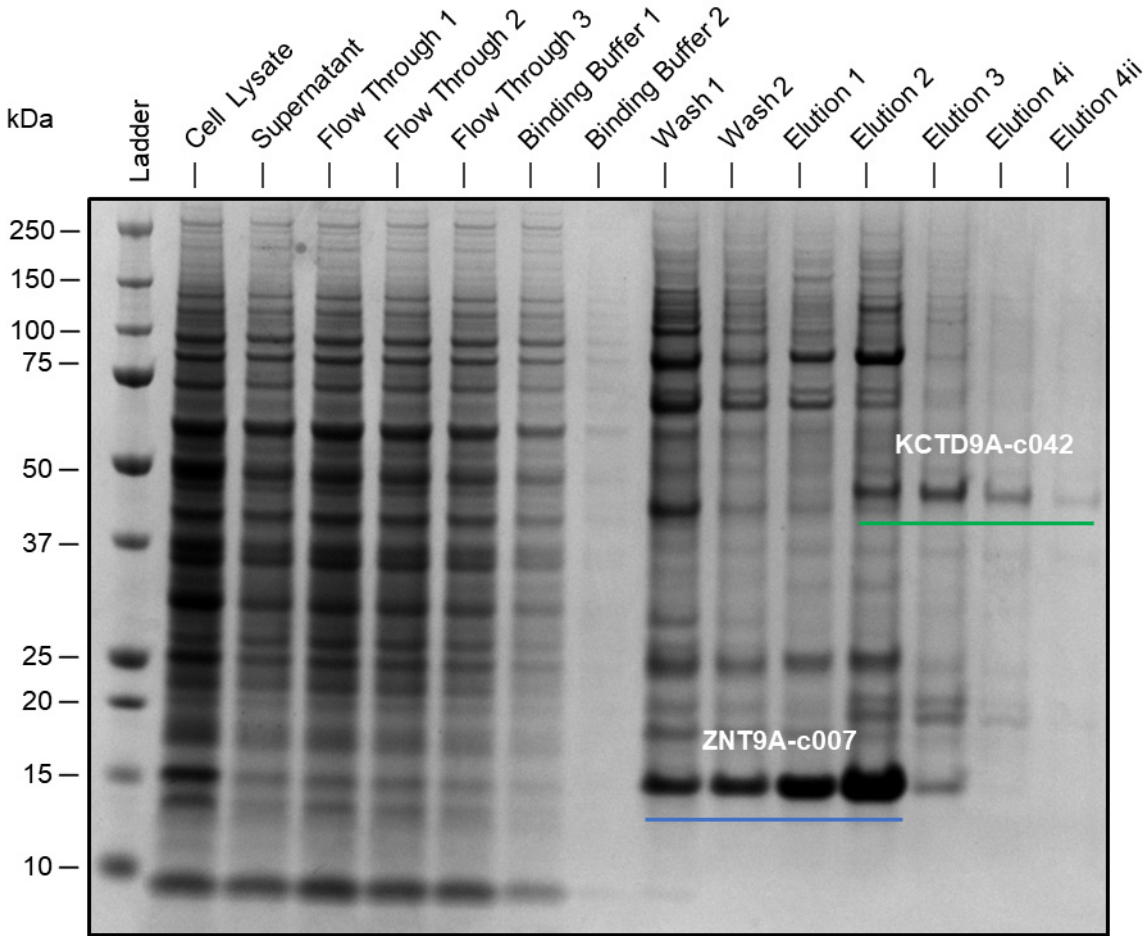


Figure 4.4: Ni^{2+}-sepharose resin capture of KCTD9A-c042 (underlined in green) co-expressed with ZNT9A-c007 (underlined in blue).

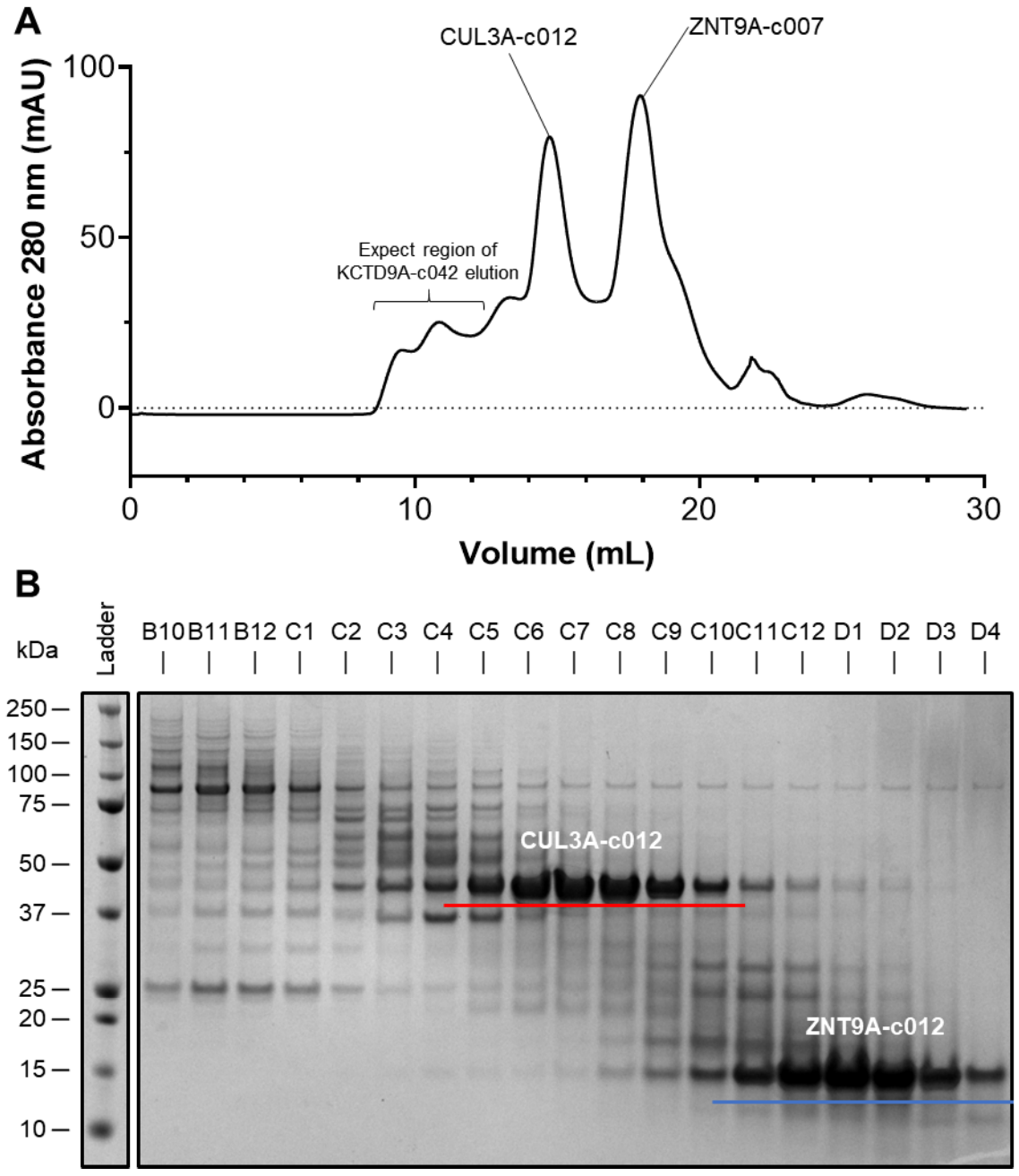


Figure 4.5: A) Size exclusion chromatography profile of KCTD9/CUL3/ZNT9 sample. Superdex 200 Increase 10/300 GL gel filtration column. B) SDS-PAGE of elution fractions (fraction size 0.5 mL). Eluted proteins underscored (CUL3: red, ZNT9: Blue).

4.2.2 Identification of Novel Interactors by MS Proteomics

4.2.2.1 Overview

Mass spectrometry (MS) proteomics encompasses a wide range of experiments that are able to survey the 'protein make-up' or proteome of a cellular environment at that condition.[89] Typically the interactors are identified by enriching the cellular environment with a modified target protein. This modification can simply be the addition of a highly selective purification tag (e.g. FLAG-tag), or the fusion of an enzymatic probe that labels endogenous proteins that are in close proximity to the target (e.g. BioID).[142] The sample is applied to an appropriate solid phase resin, and washed with buffer to remove non-binding proteins. Potential interactors are retained on the resin through either indirect binding to the tagged target protein (which itself is bound to the solid phase resin), or direct binding of the labelled interactor to resin.

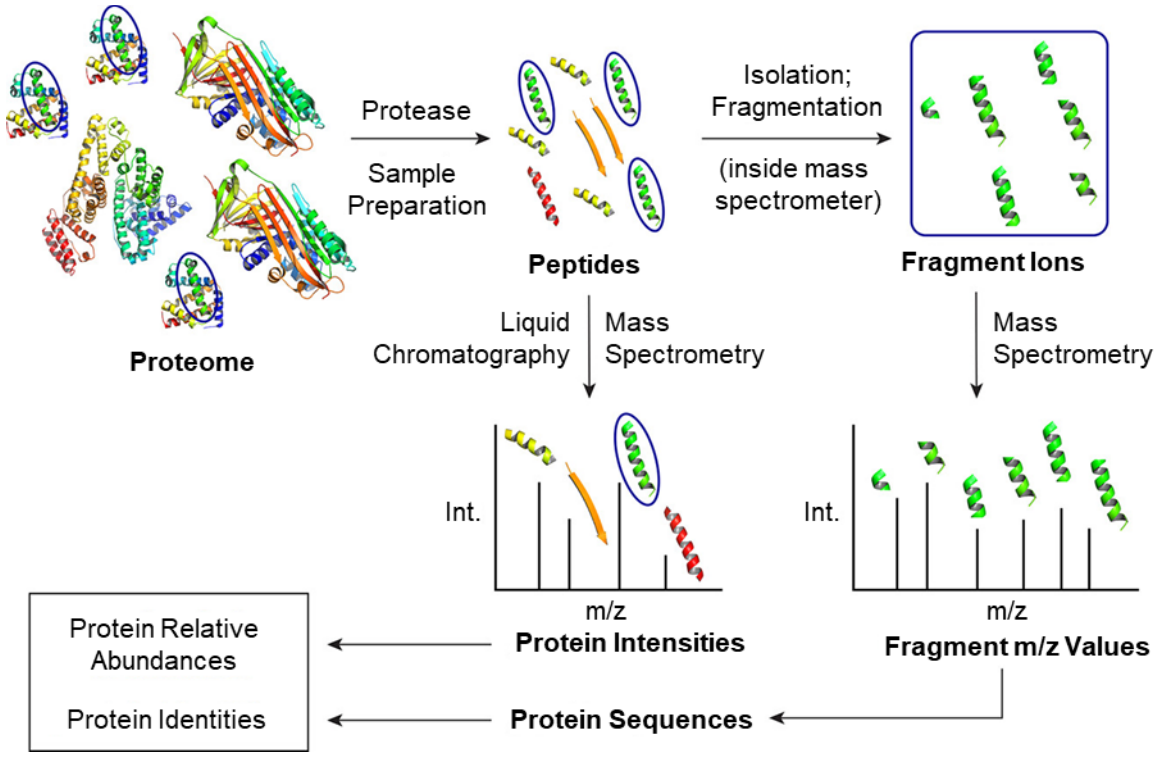


Figure 4.6: Overview of a generic mass spectrometry-based proteomic experiment. Figure sourced from a review by S. Shuken.[89]

The proteins are digested by a protease (such as trypsin) into smaller peptide fragments and the mixture is separated by liquid chromatography. The output is directly coupled to a mass spectrometer where those components are directly vaporised and ionised. The signal intensity and mass-to-charge (m/z) ratio of vaporised ions are measured. Further, tandem MS measurements will provide enough information to determine the many peptide sequences and intensities within the sample (Figure 4.6). The peptide sequences are cross-corelated with a proteome database to identify the potential interactor proteins retained during sample preparation.

4.2.2.2 Affinity Mass Spectrometry Proteomics in HEK293

I first performed a pilot experiment transiently expressing KCTD9A-c067 (3XFLAG-KCTD9$^{1-389}$, 48.9 kDa) in HEK293 cells (Figure 4.7A). The Human Protein Atlas suggests that KCTD9 is expressed across a broad range of tissue types, therefore HEK293 cells were initially selected because the cell line expresses a broad range of proteins that represents a good coverage of the proteome. Additionally, these cells were grown easily and could be transfected with using polyethylenimine.[143] The proteomics experiment was designed with a total four samples (Table: 4.2); each a different a condition and of which two were controls samples (n=1):

	Bait Protein	MLN4924 Treatment
Sample 1	Empty vector control	No treatment
Sample 2	Empty vector control	100 nM, 4h prior to harvest
Sample 3	KCTD9A-c067	No treatment
Sample 4	KCTD9A-c067	100 nM, 4h prior to harvest

Table 4.2: Summary of variables for each sample in HEK293 MS proteomics experiment.

For ease of discussion, the samples will be referred to as samples 1-4. Samples 1 and 2 were transfected with an empty control vector (a pCMV vector of similar size to KCTD9A-c067 vector, but where no protein is coded to be expressed) to establish a background level of proteins captured by nonspecific binding to the FLAG resin during immunoprecipitation (IP).

Additionally, as it is not known if KCTD9 facilitates degradative K48 ubiquitination or any other ubiquitination patterns, I treated samples 2 and 4 with the neddylation inhibitor MLN4924 (Figure 4.7B) prior to harvest. This compound serves as an AMP mimetic that covalently links to NEDD8 (N8) and prevents the NEDD8-activating enzyme from transferring the protein along a neddylation enzyme cascade. Ultimately, MLN4924 inhibits neddylation of CUL3, which prevents activation of the cullin-RING ligase (CRL) and substrate ubiquitination.[144] The deactivation of the CRL will prevent substrates from being ubiquitinated and degraded but still allows them to bind with KCTD9 and so be captured by immunoprecipitation. Comparison of protein levels in the samples treated with MLN4924 to those that were untreated may indicate which type of ubiquitination pattern KCTD9 facilitates.

Finally, FLAG-IP (rather than an endogenous tagging system) was chosen because there had been previous success in the Bullock group, where neo-substrates of mutant KBTBD4 (with low μM affinity) were identified using this protocol.[145] Potentially, fleeting protein/protein interactions could fail to be captured using this setup, but as ZNT9 was able to be immunoprecipitated with KCTD9 by FLAG-IP, I reasoned that other interactors with similar binding affinity and kinetics will be retained.[73]

The full-length KCTD9 gene was cloned into the pCMV-3Tag-1A vector by restrictive enzyme cloning. For each sample, ten large cell culture plates (Cellstar 82050-598, 145 cm^2) were cultured to 40% confluency and then transiently transfected with the appropriate vector using PEI. The cells were incubated for 48 hours and then harvested using a cell scraper. In the case of the MLN4924 treated samples (2 and 4), the cells were treated with the neddylation inhibitor (100 nM) 4 hours before cell harvesting.

The cell samples were lysed in lysis buffer containing a broad range of protease/phosphatase inhibitors and the clarified supernatant was applied to FLAG resin. The resins were washed with lysis buffer and the samples were submitted to the discovery proteomics facility at the Target Discovery Institute, University of Oxford. Western blotting of the samples showed that both Flag-tagged KCTD9 and endogenous CUL3 could be immunoprecipitated (Figure 4.7C), suggesting that the KCTD9/CUL3 complex could stay intact throughout sample preparation and that potentially other interactors could also be immunoprecipitated in complex with KCTD9. Additionally, the effectiveness of the MLN4924 treatment can be observed by the presence of a single CUL3 band for samples 3 and 4; whereas in samples 1 and 2, there is also a faint CUL3-N8 band.

Dr Sarah Flannery subjected the samples to on-bead trypsin digestion, and the peptide digests were analysed by LC-MS/MS using data independent acquisition methods. The raw data was searched against the UniProtKB human proteome database and the fold-change in protein intensity was log_2-transformed; missing values were imputed with random values generated from a downshifted normal distribution.

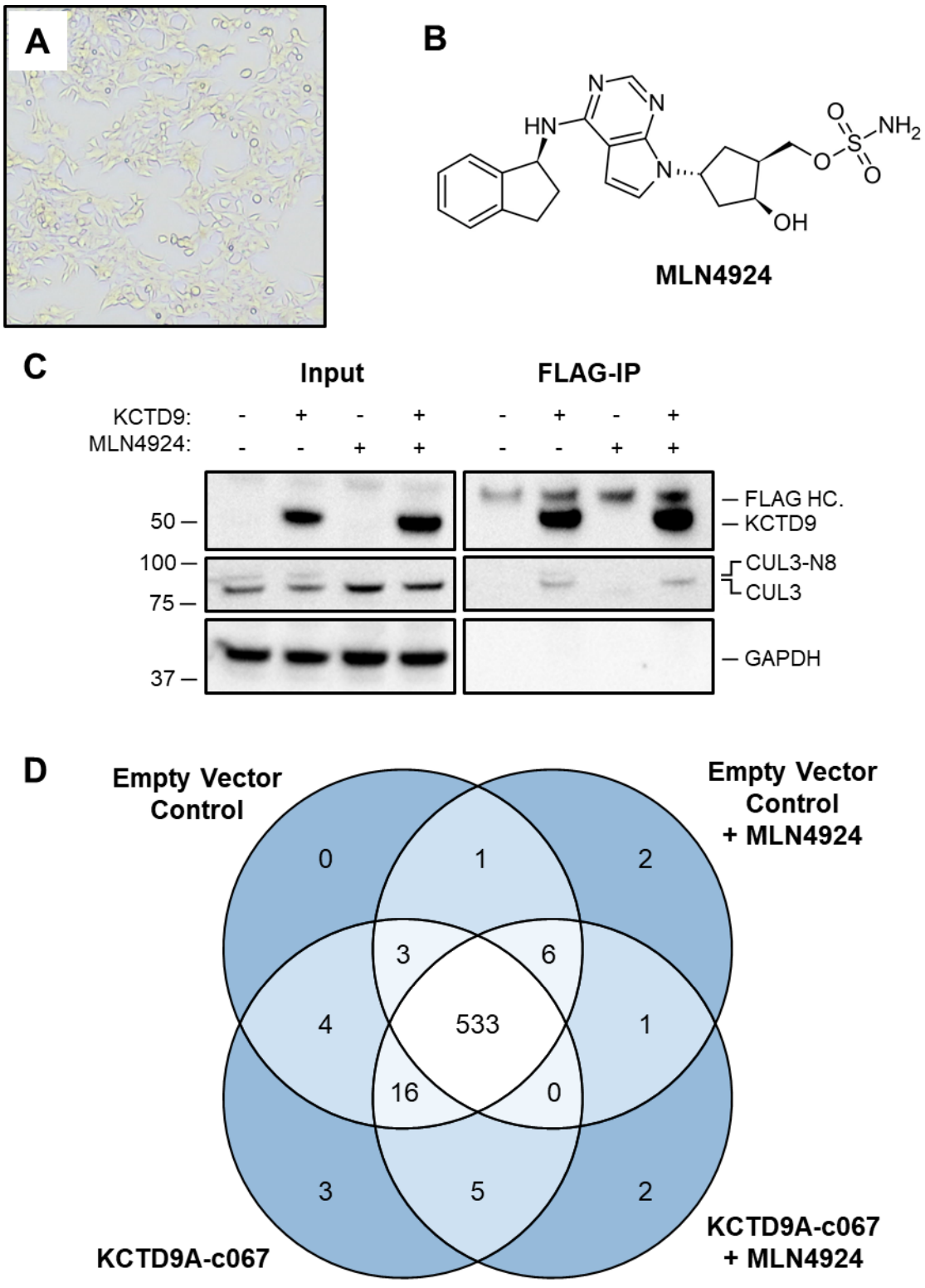


Figure 4.7: A) HEK293 prior to transient transfection (~40% confluency). B) Chemical structure of the neddylation inhibitor MLN4924. C) Western blot of HEK293 MS proteomics samples. HC: Heavy chain. D) Venn diagram of the number of proteins identified across all samples in HEK293 MS proteomics experiment.

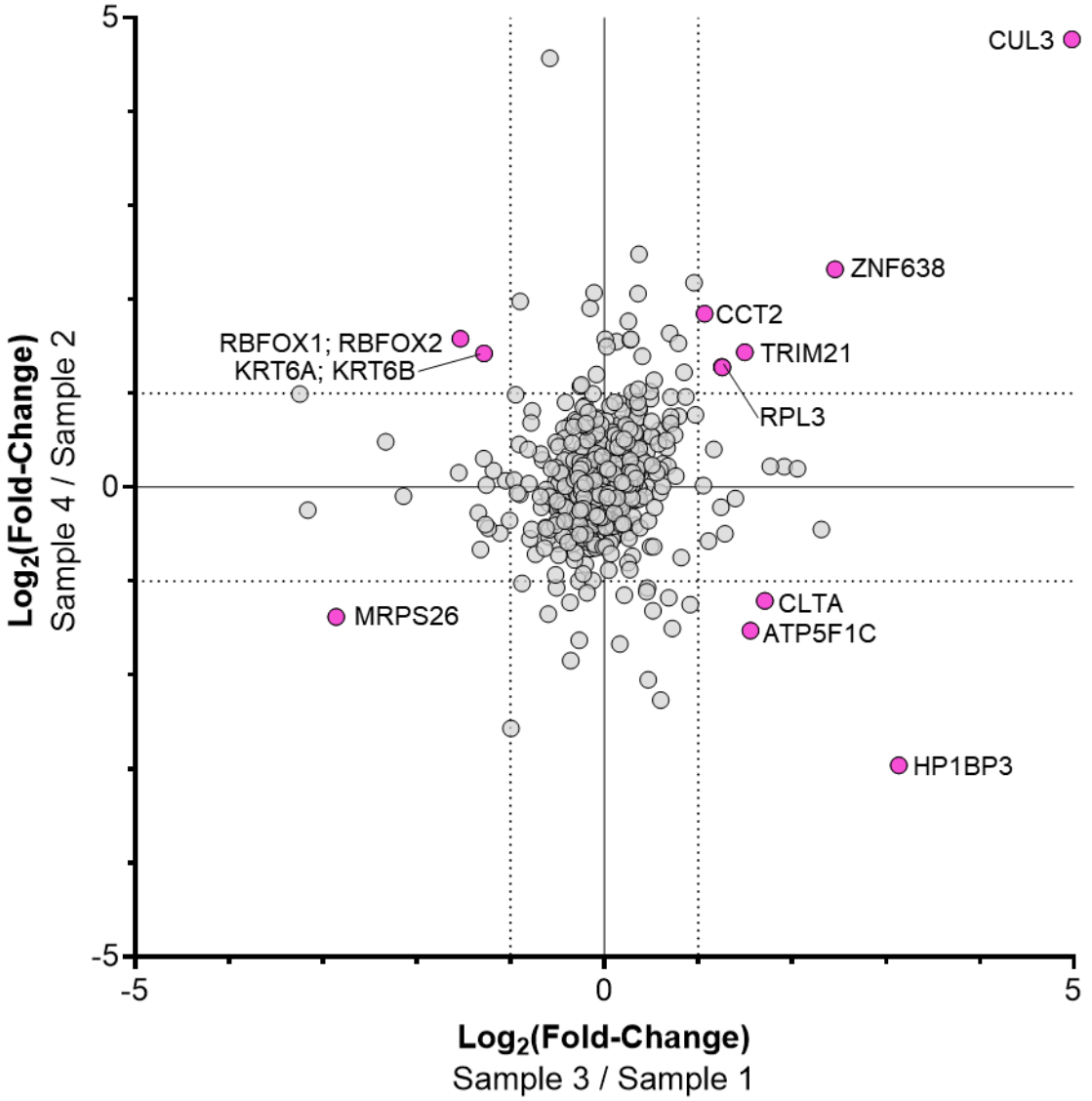


Figure 4.8: Log_2 fold change analysis of HEK293 MS proteomic samples 3 and 4 against their respective control sample (1 and 2). Dotted lines indicate the threshold where a two-fold-change is observed. Potential interactors with significant fold changes are highlighted in magenta. The KCTD9 data point is beyond the range of the chart.

576 proteins were identified across all the samples and a total of 10 proteins were only found in the samples where KCTD9 was overexpressed (Figure 4.7D). Fold-change analysis of conditions 3 and 4 against their respective control (1 and 2), indicated 12 potential interactors with greater than two-fold change in both comparisons (Figure 4.8, Table 4.3).

No.	Gene Name	Log_2(Fold-Change) Sample 3 / Sample 1	Log_2(Fold-Change) Sample 4 / Sample 2	Protein Intensity			
				Samp. 1	Samp. 2	Samp. 3	Samp. 4
1	KCTD9	9.09	11.08				
2	TRIM21	1.50	1.44				
3	UBA52; UBB; UBC	1.25	1.28				
4	KRT6A; KRT6B	-1.28	1.43				
5	HP1BP3	3.14	-2.96				
6	CUL3	4.98	4.77				
7	ZNF638	2.46	2.32				
8	RBFOX1; RBFOX2	-1.53	1.58				
9	ATP5F1C	1.56	-1.53				
10	MRPS26	-2.86	-1.38				
11	CCT2	1.07	1.85				
12	RPL3	1.26	1.28				
13	CLTA	1.71	-1.21				

Table 4.3: Proteins with significant fold-changes (> 2) in both comparisons of HEK293 MS proteomics samples. The relative protein intensity of each protein in each sample is coloured red-white-blue (high to low). Dark grey indicates where no protein was measured for that sample.

This experiment demonstrated that transiently expressed KCTD9A-c067 was able to retain E3 ligase machinery such as CUL3, and so could also potentially capture interactor proteins. However, there were only a limited number of proteins where a large fold change was observed and none of these have been linked to NK cell development. Therefore, none of these appeared as an obvious target for follow-up experiments.

While I was originally looking for a large change in protein intensity when I overexpressed KCTD9, it is possible experimental errors could hide a genuine interactor. To avoid this, I was advised to generate replicates of each sample to allow statistical analysis of the data and this would take experimental errors into account. Therefore, further proteomics experiments were planned and performed in triplicate.

4.2.2.3 Affinity Mass Spectrometry Proteomics in NK92mi

Some roles of E3 ligases could be cell type specific, and as studies by X. Zhang *et al.* have shown that KCTD9 is implicated in NK cell function, I thought that a follow up proteomics experiment in NK92mi cells would be more likely to yield genuine interactors related to this role.[72]

Unlike HEK293 cells, NK92 cells are not amenable to transient transfection, and other methods, such as electroporation, result in substantial cell death and/or a compromise in cell stability (which may affect the natural physiology of the cell). Moreover, lentivirus transduction would have required access to facilities that were not readily available and, on the scale needed for our experiment, required additional time and resources that I was unable to commit in the remainder of my studentship.

Therefore, I instead designed an experiment where KCTD9A-c067 was expressed in easily transfected Expi293F cells, immobilised on FLAG resin, and then incubated with clarified supernatant derived from NK92mi cells. These NK cells are derived from peripheral blood mononuclear cells and are reportedly interleukin-2 (IL-2) independent. The cells grow much slower than HEK293 cells and, as the cells are smaller, a larger number of cells were needed for these experiments. In total, the proteomics experiment in NK92mi cells had three conditions (Table 4.4) and each was performed in triplicate (n=3).

	FLAG Resin	**MLN4924 Treatment**
Condition 1	Empty vector control	100 nM, 4h prior to harvest
Condition 2	Pre-bound KCTD9A-c067	No treatment
Condition 3	Pre-bound KCTD9A-c067	100 nM, 4h prior to harvest

Table 4.4: Summary of variables for each condition in the NK92 MS proteomics experiment.

For ease of discussion, the conditions will be referred to as conditions 1-3. Condition 1 was incubated with the FLAG resin that was not preincubated with anything to establish a background level of proteins captured by non-specific binding to the FLAG resin during immunoprecipitation. Based on the pilot experiment, a second background condition which was not treated with MLN4924 was deemed unnecessary. As in the HEK293 experiment, comparison of samples treated with MLN4924 (condition 2 and 3) may identify degrative and non-degrative substrates of KCTD9. Thus, the conditions allow two comparison groups overall: (i) comparison of conditions 1 and 3 (+/- KCTD9) to identify those interactors specific to KCTD9; and (ii) comparison of conditions 2 and 3 (KCTD9 +/- MLN4924) to identify any KCTD9 interactors that are degraded via CRL-dependent K48-linked ubiquitination.

For each sample, 500 mL of NK92mi cells at a density of 0.4 x10^6 cells/mL (Figure 4.10A) were cultured in five T175 flasks for 48 hours and then harvested centrifugation. In the case of the MLN4924 treated conditions (1 and 3), the cells were treated with the neddylation inhibitor (100 nM) 4 hours before harvesting. Replicates 1, 2 and 3 were seeded at least 48 hours from one another, and the immunoprecipitation of all nine samples was conducted in parallel.

To obtain KCTD9A-c067, 1 L of Expi293F cells at a density of 0.4x10^6 cells/mL was transiently transfected with previously cloned pCMV vector using PEI. The cells were incubated for a further 48 hours and then harvested. The cell pellet was lysed using lysis buffer containing protease/phosphatase inhibitors and the clarified supernatant was applied to FLAG resin. The resin was washed thoroughly with lysis buffer and western blotting indicated that the resin was saturated with the KCTD9 construct (not shown).

The FLAG resin (pre-incubated with KCTD9A-c067 in conditions 2 and 3) was incubated with the clarified supernatant derived from the NK92mi cells for 16 hours at 4 °C. After the resins were washed with lysis buffer, the samples were submitted to the discovery proteomics facility at the Target Discovery Institute, University of Oxford. Dr Sarah Flannery processed them in a similar manner as the HEK293 samples, although now in triplicate. The observations of the western blots for these samples are similar to those in the HEK293 experiment (Figure 4.10B). Silver staining SDS-PAGE indicated that samples of replicate 1 had less protein present than the other two replicates and this was reflected in the protein intensity measured by the mass spectrometer (Figure 4.9).

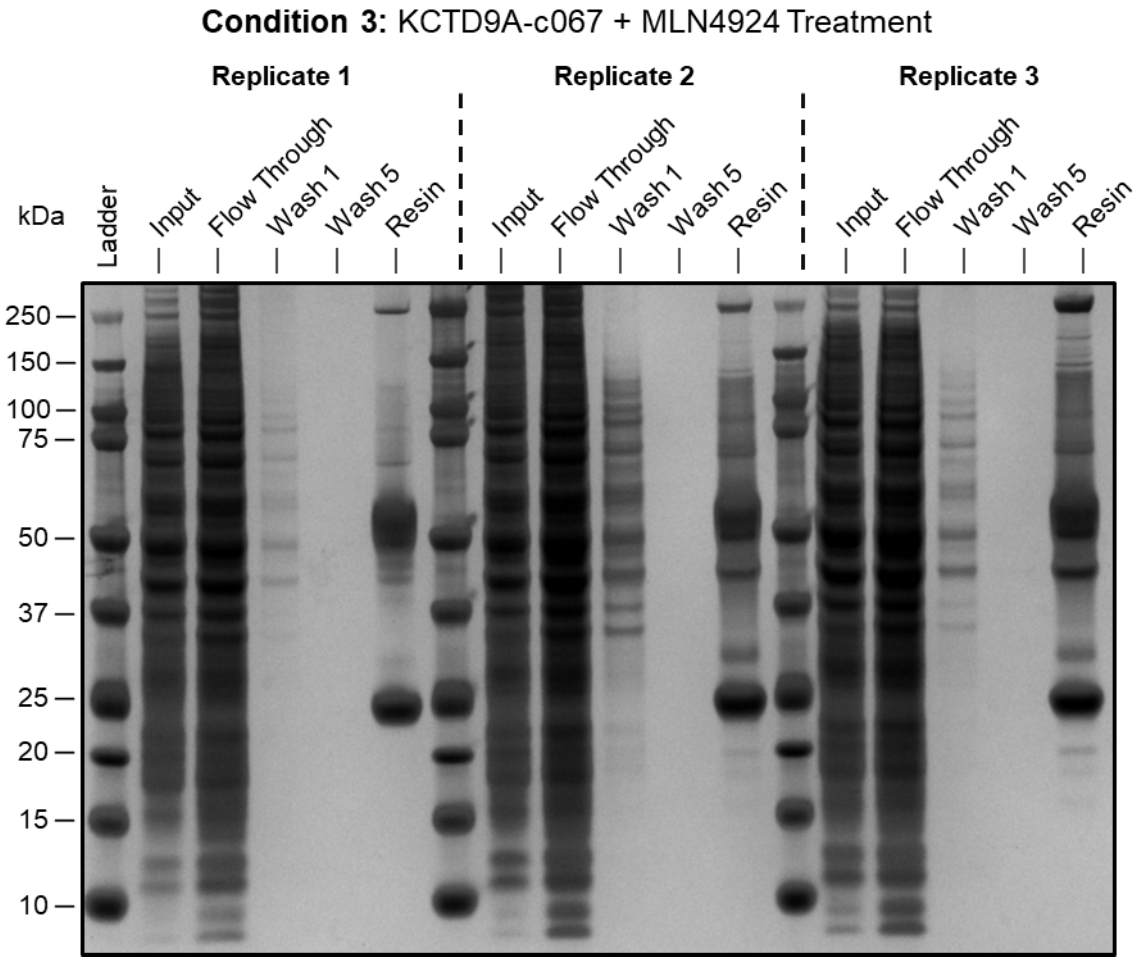


Figure 4.9: Silver stained SDS-PAGE of condition 3 replicates.

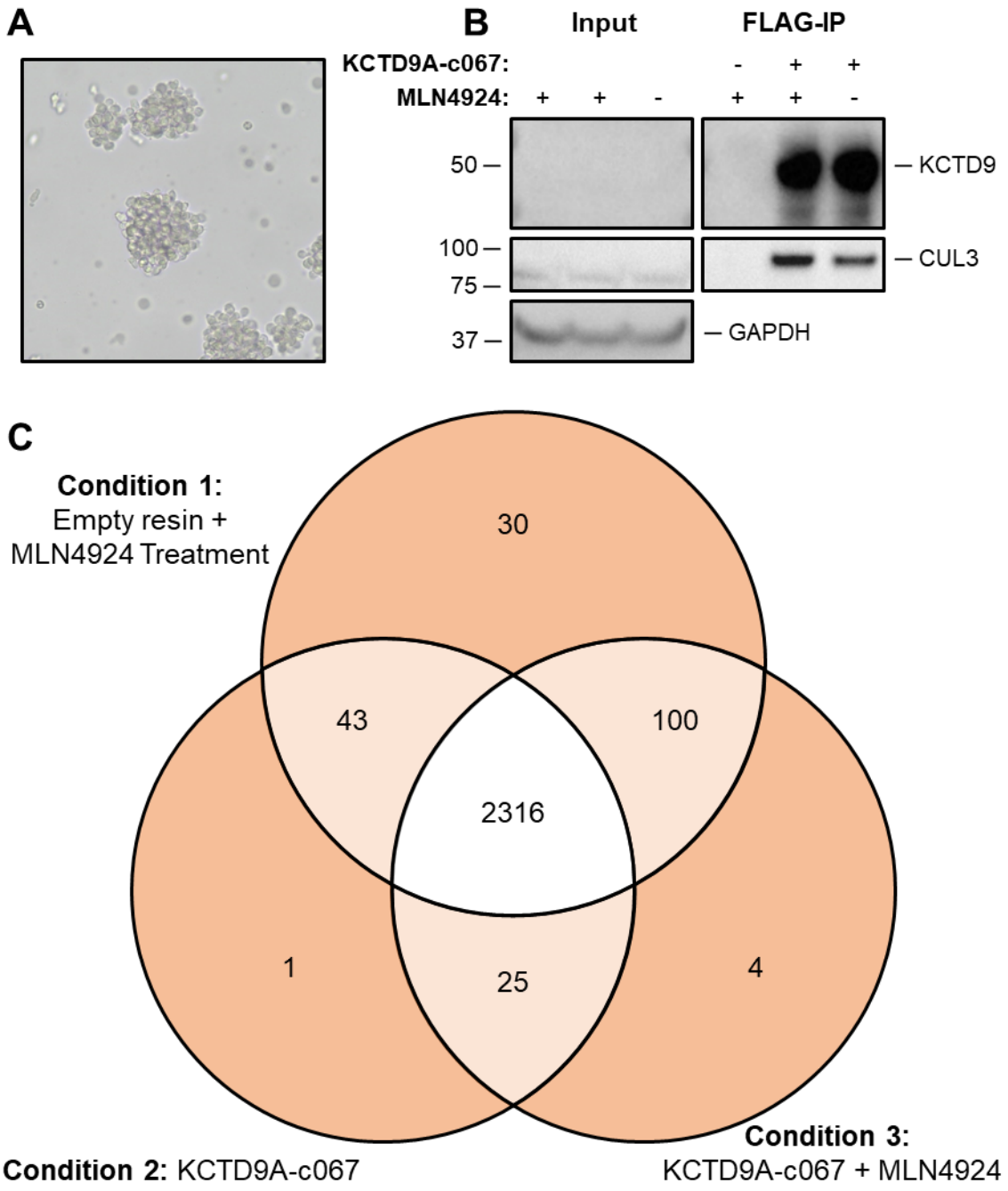


Figure 4.10: A) NK92mi after seeding (0.4x10^6 cells/mL). B) Western blot of NK92 MS proteomics samples. C) Venn diagram of the number of proteins identified across all conditions.

2,580 proteins were identified across all the samples and this number was filtered down to 2,519 to only include those identified by at least two replicates in at least one experimental condition. 30 proteins were only in the samples with the bound KCTD9, and 30 proteins were only found in the empty resin samples (Figure 4.10C).

While there were no significant proteins determined by Student's T-test, with truncation by 5% FDR (q <0.05), some proteins were significant by p-value (without FDR correction). I was advised to consider proteins with a p-value of <0.05, and a fold-change >2 as potential interactors with KCTD9. When comparing the fold-change between conditions 1 and 3, 63 proteins were above this threshold (Figure 4.11, Table 4.5, and Table 4.6).

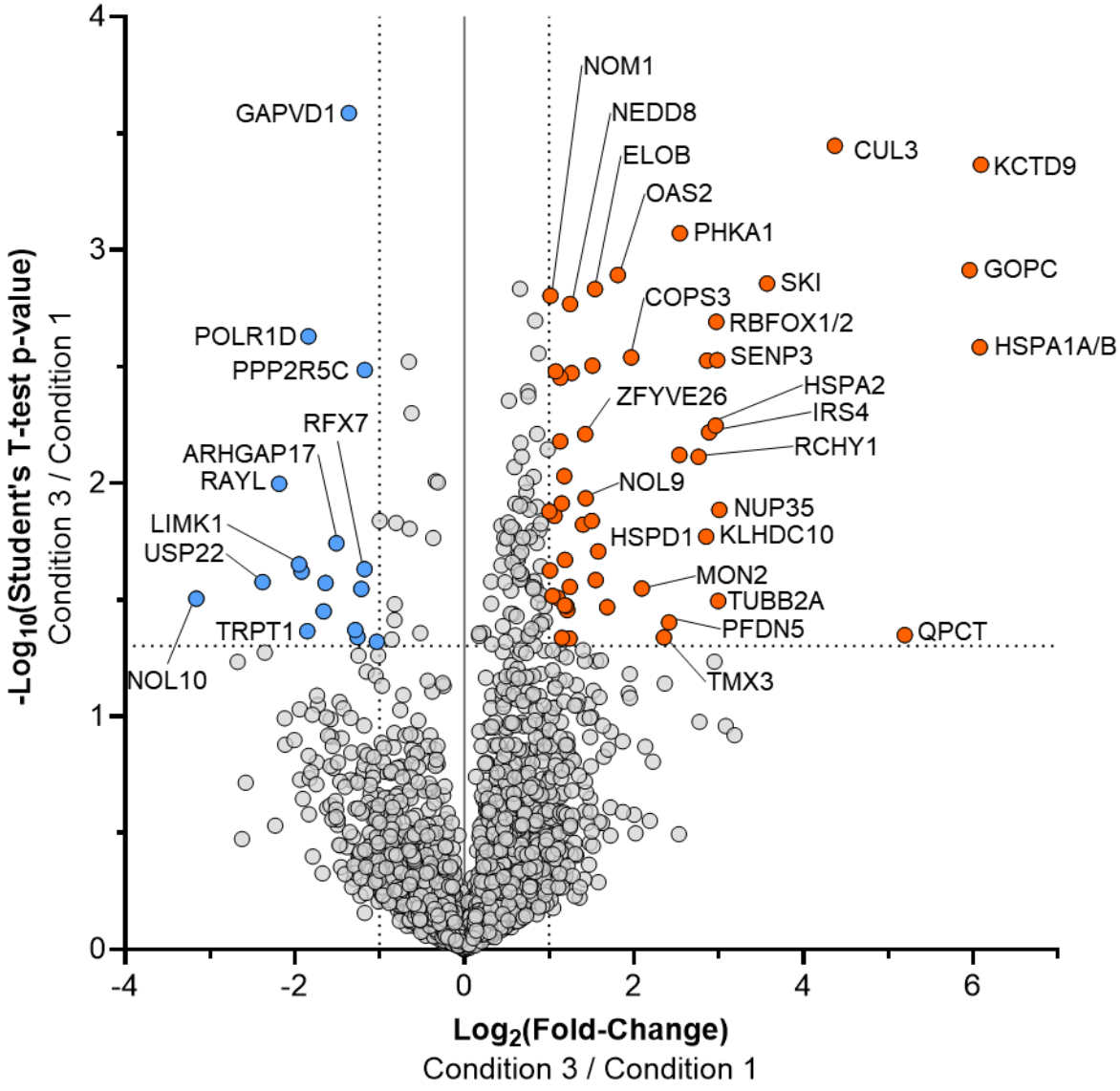


Figure 4.11: NK92 MS proteomics experiment. Student's T-test of condition 3 against condition 1. Dotted lines indicate the threshold where two-fold-change is observed (x-axis), and where p < 0.05 (y-axis). Potential interactors that were more abundant in condition 3 are highlighted in orange. Interactors that are more abundant in condition 1 are highlighted in blue.

No.	Gene Name	-Log$_{10}$(Student's T-test p-value)	Log$_2$(Fold-change)	Protein Intensity								
				Cond. 1			Cond. 2			Cond. 3		
				R1	R2	R3	R1	R2	R3	R1	R2	R3
1	KCTD9	3.37	6.09									
2	HSPA1A; HSPA1B	2.58	6.08									
3	CUL3	3.45	4.37									
4	HSPD1	1.71	1.58									
5	RUVBL2	1.52	1.04									
6	QPCT	1.35	5.19									
7	OAS2	2.89	1.81									
8	GOPC	2.91	5.96									
9	SLC25A6	2.45	1.13									
10	RUVBL1	1.48	1.18									
11	ELOB	2.83	1.54									
12	SRRM2	1.84	1.50									
13	ZNF638	2.53	2.86									
14	DNAJB6	2.47	1.27									
15	SKI	2.86	3.57									
16	ATP5ME	1.67	1.19									
17	NOL9	1.93	1.43									
18	RCHY1	2.11	2.76									
19	RCN2	2.18	1.13									
20	NEDD8	2.77	1.25									
21	NUP35	1.89	3.01									
22	KLHDC10	1.77	2.85									
23	GPS1	1.45	1.22									
24	COX5A	2.48	1.08									
25	IRS4	2.22	2.89									
26	HSPA2	2.25	2.96									
27	TUBB2A	1.49	2.99									
28	SRPRB	1.56	1.24									
29	PFDN6	1.40	2.42									
30	COPS2	1.82	1.40									
31	NOM1	2.80	1.02									
32	PHF21A	1.47	1.68									
33	SENP3	2.53	2.98									
34	RBFOX1; RBFOX2	2.69	2.97									
35	BCL11A; BCL11B	1.58	1.55									
36	NDUFA5	2.03	1.18									
37	NDUFV2	1.51	1.10									
38	TMX3	1.34	2.36									
39	DCAF1	1.34	1.15									
40	MON2	1.55	2.09									
41	PHKA1	3.07	2.54									
42	P4HA1	1.86	1.06									
43	COPS3	2.54	1.97									
44	DPM1	2.50	1.51									

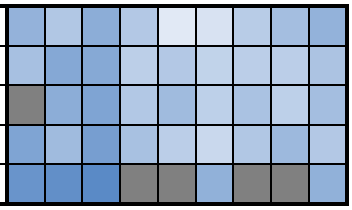
45	NCSTN	1.33	1.24	
46	EIF2B4	1.62	1.01	
47	VARS1	1.91	1.15	
48	FKBP8	1.47	1.20	
49	IQGAP2	1.88	1.00	

Table 4.5: NK92 MS proteomics experiment. Student's T-test analysis and fold-change analysis of conditions 1 and 3. Proteins that were more abundant in condition 3, Log_2(Fold-change) > 1, and deemed significant by Student's T-test, $-Log_{10}(p) > 1.3$. The relative protein intensity of each protein in each replicate is coloured red-white-blue (high to low). Dark grey indicates no protein was measured for that sample. Only genes with at least 5 datapoints are included.

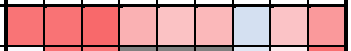
No.	Gene Name	$-Log_{10}$(Student's T-test p-value)	Log_2(Fold-change)	Protein Intensity								
				Cond. 1			Cond. 2			Cond. 3		
				R1	R2	R3	R1	R2	R3	R1	R2	R3
1	GAPVD1	3.59	-1.36									
2	NOL10	1.50	-3.16									
3	STAT5B	1.62	-1.92									
4	USP22	1.58	-2.38									
5	RALYL	2.00	-2.18									
6	LIMK1	1.65	-1.95									
7	RFX7	1.63	-1.18									
8	PLEKHF2	1.45	-1.66									
9	ARHGAP17	1.74	-1.51									
10	BCKDK	1.34	-1.26									
11	TBC1D4	1.57	-1.64									
12	PPP2R5C	2.48	-1.17									
13	PFKM	1.55	-1.22									
14	SPTY2D1	1.37	-1.29									
15	POLR1D	2.63	-1.84									

Table 4.6: NK92 MS proteomics experiment. Student's T-test analysis and fold-change analysis of conditions 1 and 3. Proteins that were more abundant in condition 1, Log_2(Fold-change) < -1, and deemed significant by Student's T-test, $-Log_{10}(p) > 1.3$. The relative protein intensity of each protein in each replicate is coloured red-white-blue (high to low). Dark grey indicates no protein was measured for that sample. Only genes with at least 3 datapoints are included.

4.2.2.3.4 DAVID Analysis of Potential Interactors

Genes that had greater than 2-fold-change, and a Student's T-test p-value less than 0.05 were submitted to the DAVID bioinformatics server for Gene Ontology analysis and clustering.[146] This platform grouped genes together by a number of factors including reported molecular functions, cell localisation, and/or biological processes. Using the default parameters (medium stringency), six clusters were reported (out of 15 in total) with p-values < 0.05 (Table 4.7).

Annotation Cluster 1 Enrichment Score: 3.48

Category	Term	Count	p-value	Genes
Cellular Component	Nucleoplasm	28	7.87×10^{-6}	OAS2, BCL11A, COPS2, DCAF1, NUP35 etc.
Cellular Component	Nucleus	32	8.36×10^{-4}	
Cellular Component	Nucleus	29	5.38×10^{-3}	

Annotation Cluster 2 Enrichment Score: 3.13

Category	Term	Count	p-value	Genes
Biological Process	Protein Neddylation	4	3.94×10^{-5}	COPS2, COPS3, CUL3, DCAF1, GPS1, NEDD8, RFX7
Cellular Component	COP9 Signalosome	4	1.92×10^{-4}	
Biological Process	Protein Deneddylation	3	4.91×10^{-4}	
Cellular Component	Signalosome	3	5.55×10^{-4}	
Biological Process	Regulation of Protein Neddylation	3	9.29×10^{-4}	
Interpro	Pci Domain	3	2.34×10^{-3}	

Annotation Cluster 3 Enrichment Score: 2.15

Category	Term	Count	p-value	Genes
Cellular Component	RPAP3/R2TP/Prefoldin-Like Complex	3	5.55×10^{-4}	RUVBL1, RUVBL2, PFDN6
Cellular Component	Protein Folding Chaperone Complex	3	3.77×10^{-3}	
Biological Process	Protein Stabilization	3	1.74×10^{-1}	

Annotation Cluster 4 Enrichment Score: 1.43

Category	Term	Count	p-value	Genes
Post-translational Modification	Isopeptide Bond	15	2.52×10^{-2}	BCL11A, FKBP8, LIMK1, NEDD8, PHF21A, etc.
Post-translational Modification	Ubl Conjugation	18	6.41×10^{-2}	

Annotation Cluster 5 Enrichment Score: 1.19

Category	Term	Count	p-value	Genes
KEGG Pathway	Alzheimer Disease	7	2.42 x10^{-3}	ATP5ME, DPM1, PFKM, P4HA1, NDUFV2, NDUFA5, COX5A, IRS4, NCSTN, SLC25A6, TBC1D4, TUBB2A, etc.
KEGG Pathway	Diabetic Cardiomyopathy	5	6.01 x10^{-3}	
KEGG Pathway	Oxidative Phosphorylation	4	1.32 x10^{-2}	
KEGG Pathway	Parkinson Disease	5	1.57 x10^{-2}	
Biological Process	Proton Transmembrane Transport	4	1.64 x10^{-2}	
Biological Process	Mitochondrial ATP Synthesis	3	1.67 x10^{-2}	
KEGG Pathway	Prion Disease	5	1.71 x10^{-2}	
KEGG Pathway	Huntington Disease	5	2.46 x10^{-2}	
Cellular Component	Mitochondrion Inner Membrane	5	2.72 x10^{-2}	
KEGG Pathway	Amyotrophic Lateral Sclerosis	5	4.31 x10^{-2}	
KEGG Pathway	Thermogenesis	4	5.19 x10^{-2}	
Cellular Component	Mitochondrial Inner Membrane	5	6.48 x10^{-2}	
KEGG Pathway	Pathways of Neurodegeneration	5	9.39 x10^{-2}	
KEGG Pathway	Non-Alcoholic Fatty Liver Disease	3	1.10 x10^{-1}	
Cellular Component	Mitochondrion	8	2.39 x10^{-1}	
Cellular Component	Mitochondrion	7	2.86 x10^{-1}	
KEGG Pathway	Metabolic Pathways	7	5.07 x10^{-1}	
Molecular Function	Oxidoreductase	3	6.28 x10^{-1}	

Annotation Cluster 6 Enrichment Score: 1.16

Category	Term	Count	p-value	Genes
Interpro	RNA Binding Domain	5	8.34 x10^{-3}	NOL9, NUP35, OAS2, RALYL, RBFOX1, RBFOX2, SRRM2, ZNF638, etc.
Interpro	Nucleotide Binding Domain	5	8.89 x10^{-3}	
Interpro	RNA Recognition Motif	4	3.32 x10^{-2}	
Molecular Function	RNA Binding	7	3.98 x10^{-2}	
Molecular Function	mRNA Binding	4	5.17 x10^{-2}	
Biological Process	RNA Splicing	3	1.29 x10^{-1}	
Biological Process	mRNA Processing	3	1.80 x10^{-1}	
Molecular Function	RNA Binding	8	1.96 x10^{-1}	
Molecular Function	Nucleic Acid Binding	3	2.29 x10^{-1}	
Biological Process	mRNA Splicing	3	2.80 x10^{-1}	
Biological Process	mRNA Processing	3	3.89 x10^{-1}	

Table 4.7: DAVID analysis of potential interactors identified by fold-change analysis of conditions 1 and 3 in the NK92 MS proteomics experiment.[146] Clusters 7-15 are omitted due to their high p-values.

Cluster 1 suggests that many of the proteins that were identified are thought to be localised within cell nucleus. According to The Human Protein Atlas, KCTD9 is also mainly localised in the nucleoplasm (Figure 4.12).[147] Interestingly, one member of this cluster, NUP35 (a component of the nuclear pore complex) has been previously identified as a potential interactor of KCTD9 in multiple high-throughput yeast two-hybrid (Y2H) assays (BioGrid).[148,149]

Cluster 2 and 4 indicated that many of the proteins were associated to neddylation and ubiquitination systems. Many components of the COP9 signalosome (GPS1, COPS2 and COPS3) were more abundant in the KCTD9 IP. This multimeric assembly deactivates CRLs and binds to deneddylated cullin.[150] As the samples were treated with MLN4924, which prevents neddylation of CUL3, the pulldown of these components are likely an artifact of this treatment. Several substrate receptors/adaptors such as DCAF1 (CRL4 substrate receptor), KLHDC10 (CRL2 substrate adaptor), and Elongin-B (CRL2/5 substrate adaptor) were identified. Additionally, increased quantities of a RING-finger containing ligase RCHY1 was detected in condition 3.

While cluster 5 appears to be broad category which encompasses a variety of pathways, cluster 6 specifically identifies a number of proteins which have a RNA binding domain, or are related RNA processing. NUP35, ZNF638, and RALYL are some examples of proteins which contain a RRM (RNA recognition motif) domain. This highly conserved domain has been shown to bind both RNA and other proteins for a wide variety of biological processes.[151] Furthermore, several DNA binding zinc finger domain containing proteins such as BCL11A/BCL11B and ZNF638 were clustered together.

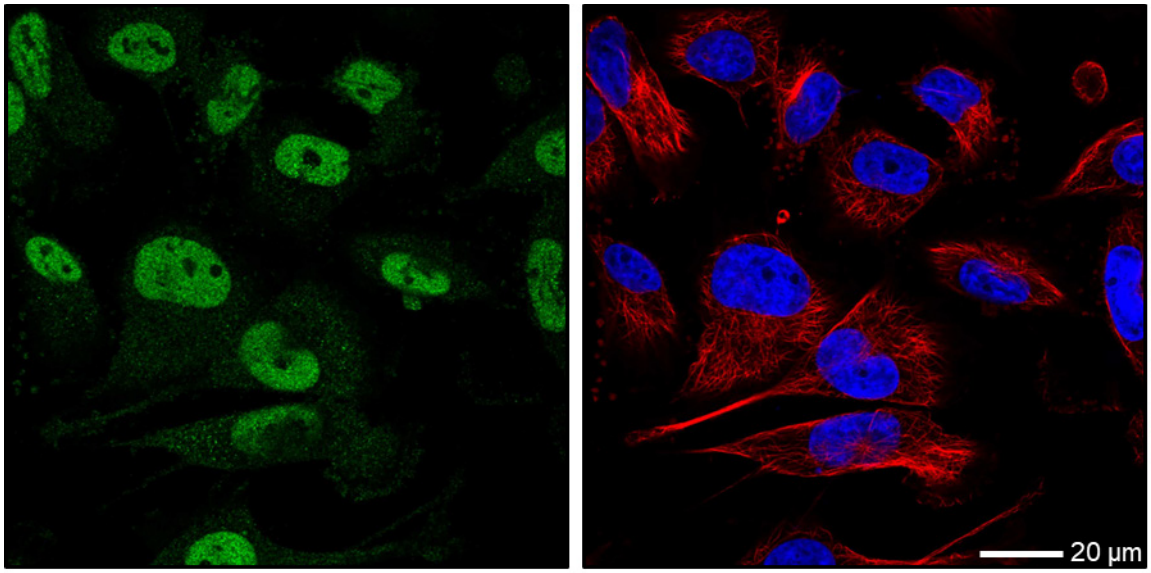


Figure 4.12: KCTD9 is localised to the nucleoplasm in U-251MG cells. KCTD9 (HPA042468, green), microtubules (red), cell nucleus (blue). Images sourced from the Human Protein Atlas.[149]

As a final method of analysis, I identified which proteins had been linked to NK cells and/or the innate immune response. OAS2 is a member of the 2-5 synthetase family and involved in the primary response to viral infection.[152] The enzyme is activated by binding of double stranded RNA and 2',5'-oligoadenylates are synthesised. These molecules activate latent RNaseL, which then degrades viral RNA and prevents virus replication. Transcription factors STAT5B and RFX7 were found more abundantly in the control sample (condition 1). STAT5B mediates the signalling of IL-2 (Interleukin-2), a cytokine that regulates many processes of NK cells and cytotoxic T cells.[153,154] RFX7 has been shown to limit the metabolism of NK cells and its knockdown limited the number of NK cells present.[155] Finally, BCL11A expression was shown to be upregulated in NKTL (natural killer/T-cell lymphoma) and, when it was silenced, reduced cell proliferation and increased apoptosis.[156]

4.2.2.3.5 Additional Fold-Change Analysis

Fold-change analysis between conditions 2 and 3 did not identify many proteins with significant abundance in one condition over the other (Figure 4.13 and Table 4.8). This indicates that the ubiquitination activity of KCTD9 might be for a process other than protein degradation. Comparison of conditions 1 and 2 revealed a high number of proteins with significant changes in protein abundance (Figure 4.14, Table 4.9, and Table 4.10). This is to be expected as between the conditions both the MLN4924 treatment and presence of KCTD9 are different. The data is only included for completeness. I prioritised proteins identified by fold-change analysis between conditions 1 and 3 (Table 4.5 and 4.6) to determine which ones should be evaluated further.

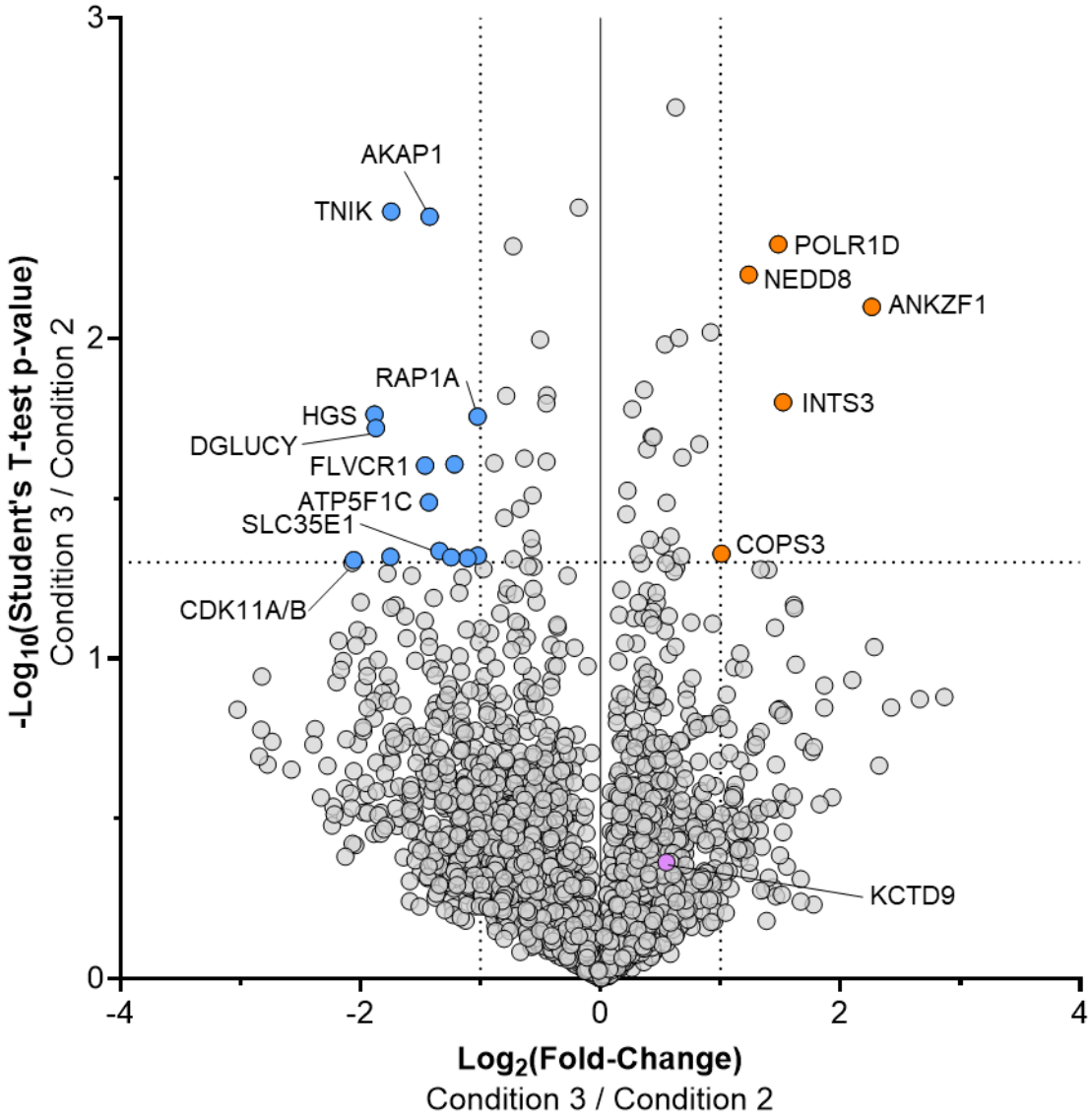


Figure 4.13: NK92 MS proteomics experiment. Student's T-test of condition 3 against condition 2. Dotted lines indicate the threshold where two-fold change is observed (x-axis), and where $p < 0.05$ (y-axis). Potential interactors that were more abundant in condition 3 are highlighted in orange. Interactors that are more abundant in condition 2 are highlighted in blue. KCTD9 datapoint is highlighted in purple.

No.	Gene Name	-Log_{10}(Student's T-test p-value)	Log_2(Fold-change)	Protein Intensity								
				Cond. 1			Cond. 2			Cond. 3		
				R1	R2	R3	R1	R2	R3	R1	R2	R3
1	GFAP;KRT4;KRT7	3.41	4.22									
2	NSD2	1.32	-1.02									
3	NEDD8	2.20	1.24									
4	RAP1A	1.76	-1.02									
5	ANKZF1	2.10	2.26									
6	CDK11A;CDK11B	1.31	-2.06									
7	ATP5F1C	1.49	-1.43									
8	RPRD1B	1.61	-1.22									
9	COPS3	1.33	1.01									
10	HGS	1.76	-1.88									
11	INTS3	1.80	1.52									
12	DGLUCY	1.72	-1.87									
13	AKAP1	2.38	-1.42									
14	TNIK	2.40	-1.75									
15	SLC35E1	1.32	-1.75									
16	PDHA1;PDHA2	1.34	-1.34									
17	TBC1D10B	1.31	-1.11									
18	FLVCR1	1.60	-1.46									
19	POLR1D	2.29	1.48									

Table 4.8: NK92 MS proteomics experiment. Student's T-test analysis and fold-change analysis of conditions 2 and 3. Proteins that had a large change abundance between the two samples, Log_2(Fold-change) > 1 or < -1, and deemed significant by Student's T-test, -Log_{10}(p) > 1.3. The relative protein intensity of each protein in each replicate is coloured red-white-blue (high to low). Dark grey indicates no protein was measured for that sample. Only genes with at least 3 datapoints are included.

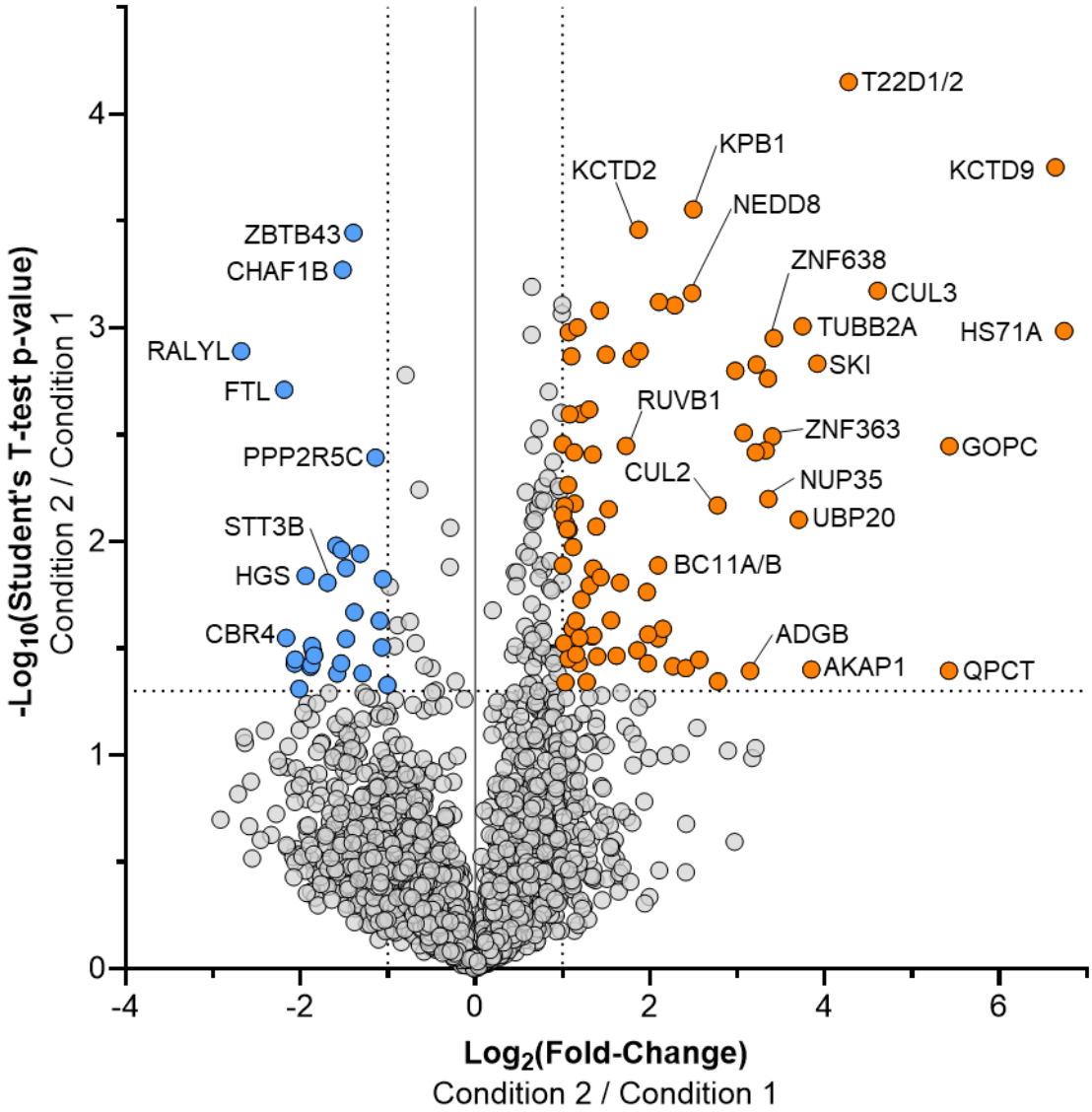


Figure 4.14: NK92 MS proteomics experiment. Student's T-test of condition 2 against condition 1. Dotted lines indicate the threshold where two-fold change is observed (x-axis), and where $p < 0.05$ (y-axis). Potential interactors that were more abundant in condition 2 are highlighted in orange. Interactors that are more abundant in condition 1 are highlighted in blue.

No.	Gene Name	-Log10(Student's T-test p-value)	Log2(Fold-change)	Protein Intensity								
				Cond. 1			Cond. 2			Cond. 3		
				R1	R2	R3	R1	R2	R3	R1	R2	R3
1	KCTD9	3.75	6.64									
2	HSPA1A; HSPA1B	2.98	6.74									
3	TUBB	2.6	1.21									
4	HSPA8	2.62	1.31									
5	TUBA4A	2.18	1.14									
6	RPS27A	1.55	2.10									
7	CUL3	3.17	4.61									
8	HSPD1	1.76	1.96									
9	HSPA9	2.15	1.53									
10	RUVBL2	2.07	1.38									
11	QPCT	1.40	5.42									
12	OAS2	3.10	2.28									
13	HSPA5	2.08	1.03									
14	GOPC	2.45	5.43									
15	NAP1L1	1.49	1.86									
16	RUVBL1	2.45	1.73									
17	ELOB	2.86	1.79									
18	SQOR	1.56	1.34									
19	ZNF638	2.95	3.42									
20	ELOC	1.81	1.66									
21	DNAJA1	2.26	1.06									
22	USP20	2.10	3.70									
23	DNAJB6	3.08	1.43									
24	SERPINB9	1.34	1.03									
25	NT5DC2	1.40	3.85									
26	ADGB	1.39	3.15									
27	ICAM1	1.34	1.28									
28	TSC22D1/2	4.15	4.28									
29	NEDD8	3.16	2.48									
30	SKI	2.83	3.92									
31	NOL9	1.73	1.22									
32	NUP35	2.20	3.36									
33	RCHY1	2.49	3.40									
34	AFG2A	2.05	1.07									
35	RUNX3	1.52	1.02									
36	COPS5	1.34	2.78									
37	KLHDC10	2.43	3.32									
38	GPS1	1.63	1.55									
39	IRS4	2.51	3.07									
40	RUNX1	2.87	1.10									
41	PSMC2	2.59	1.08									
42	LAS1L	1.42	2.26									
43	SRPRB	1.46	1.4									
44	HSPA2	2.76	3.35									

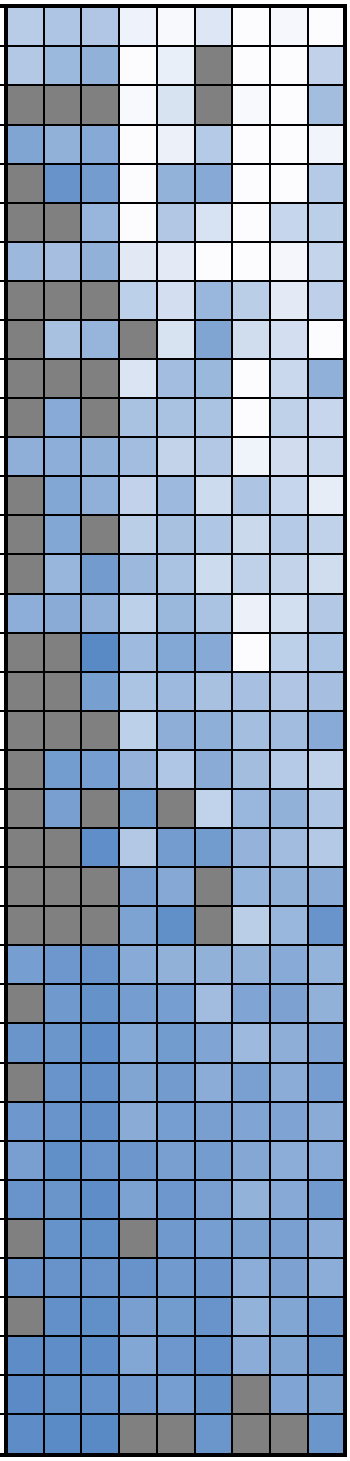
45	MYO9B	2.46	1.00	
46	NCAPD3	1.43	1.19	
47	TUBB2A	3.01	3.75	
48	COPS2	2.89	1.88	
49	CUL2	2.17	2.77	
50	PFDN6	1.41	2.41	
51	NOM1	1.89	1.00	
52	SENP3	2.83	3.22	
53	OBI1	1.59	1.11	
54	RBFOX1; RBFOX2	2.42	3.21	
55	BCL11A; BCL11B	1.89	2.09	
56	PANK4	2.98	1.07	
57	DCAF1	1.56	1.35	
58	TMX3	1.45	2.56	
59	LTN1	1.47	1.62	
60	ATP6V1H	2.17	1.02	
61	COPS3	2.80	2.98	
62	MON2	1.59	2.15	
63	PHKA1	3.55	2.49	
64	P4HA1	2.87	1.5	
65	SERPINH1	1.57	1.98	
66	COPS7B	3.12	2.10	
67	KCTD2	3.46	1.87	
68	COPS6	1.43	1.98	
69	ETV3	2.12	1.01	
70	ZNF281	1.63	1.15	
71	AMBRA1	1.87	1.35	
72	VARS1	1.97	1.12	
73	SLK	2.06	1.05	
74	ADRM1	1.45	1.07	
75	HYOU1	1.47	1.16	
76	METAP2	2.41	1.34	
77	MLH1	3.00	1.17	
78	PFDN1	1.79	1.31	
79	ARL4C	1.83	1.43	
80	TBC1D2B	1.55	1.19	
81	IQGAP2	2.42	1.13	

Table 4.9: NK92 MS proteomics experiment. Student's T-test analysis and fold-change analysis of conditions 1 and 2. Proteins that were more abundant in condition 2, Log_2(Fold-change) > 1, and deemed significant by Student's T-test, Log_{10}(p) > 1.3. The relative protein intensity of each protein in each replicate is coloured red-white-blue (high to low). Dark grey indicates no protein was measured for that sample. Only genes with at least 5 datapoints are included.

No.	Gene Name	-Log_{10}(Student's T-test p-value)	Log_2(Fold-change)	Protein Intensity								
				Cond. 1			Cond. 2			Cond. 3		
				R1	R2	R3	R1	R2	R3	R1	R2	R3
1	H1-5	1.33	-1.01									
2	DCTPP1	1.50	-1.07									
3	RBM3	1.98	-1.59									
4	CSTF2	1.31	-2.01									
5	IRF3	1.43	-2.07									
6	FTL	2.71	-2.19									
7	CBR4	1.55	-2.16									
8	RPRD1B	1.96	-1.53									
9	STT3B	1.81	-1.69									
10	RALYL	2.89	-2.68									
11	PTMA	1.48	-1.86									
12	BMP2K	1.43	-1.54									
13	WDR6	1.42	-1.88									
14	HGS	1.84	-1.94									
15	ATG3	1.47	-1.84									
16	CHAF1B	3.27	-1.52									
17	DOK1	1.45	-2.06									
18	PWP1	1.51	-1.87									
19	GAK	1.38	-1.58									
20	ZBTB43	3.44	-1.39									
21	ORC3	1.67	-1.38									
22	PPP2R5C	2.39	-1.14									
23	NSD2	1.38	-1.29									
24	ELMO1	1.54	-1.48									
25	CEP41	1.63	-1.09									
26	PFKM	1.94	-1.32									

Table 4.10: NK92 MS proteomics experiment. Student's T-test analysis and fold-change analysis of conditions 1 and 2. Proteins that were more abundant in condition 1, Log_2(Fold-change) < -1, and deemed significant by Student's T-test, $Log_{10}(p) > 1.3$. The relative protein intensity of each protein in each replicate is coloured red-white-blue (high to low). Dark grey indicates no protein was measured for that sample. Only genes with at least 3 datapoints are included.

4.2.2.4 Shortlisting Potential Interactors Using Alphafold3

The proteomics dataset is still under analysis and potential interactors will need to be validated further by co-IP, western blotting, and ubiquitination assays. To assist in narrowing the list of proteins that will be followed up, I utilised the Alphafold3 server to predict potential interactions with KCTD9.[107] It is not known if a single interactor will associate with a KCTD9 pentamer, or if five interactors will associate with the assembly. For the first round of modelling using Alphafold3, I tried to model the latter ratio (unless prevented by the size limit of 5,000 total residues per prediction). Decameric KCTD9 was not modelled because Alphafold3 could not correctly predict the structure observed in the cryo-EM data.

The first interactor I tried to model with KCTD9 was GOPC (Golgi-associated PDZ and coiled-coil motif-containing protein) as it was only found in conditions containing KCTD9. As the name implies, this protein contains an N-terminal coiled-coil region and C-terminal PDZ domain. It has been shown to mediate ubiquitination of CTFR (cystic fibrosis transmembrane conductance regulator) using a different E3 ligase, MARCH2, which ultimately results in lysosomal degradation of CTFR.[157]

The PDZ domain generally recognises one of three short motifs (type I-III) at the C-termini of the interactor proteins. GOPC specifically accommodates the type I motif (X-S/T-X-Φ*, X: any amino acid, Φ: hydrophobic amino acid) present in MARCH2 and CTFR. J. Amacher *et al.* generated several $GOPC^{PDZ}$ co-crystal structures with 10-mer peptides; including the high-affinity iCAL36 peptide (K_i = 22.6 ± 8.0 uM, PDB: 4E34, Figure 4.15A), which can be used to compare with KCTD9/GOPC interactions. [158]

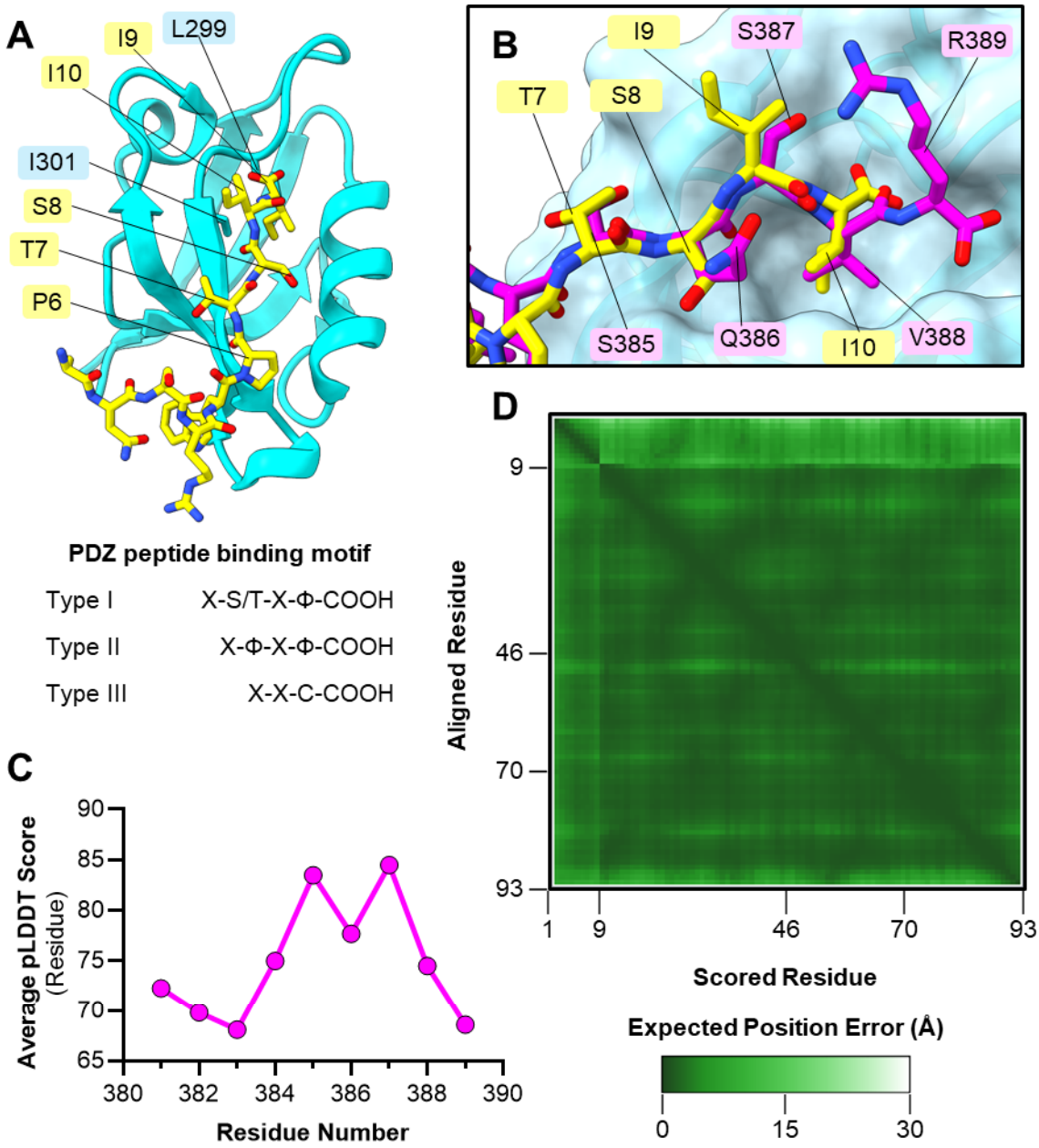


Figure 4.15: A) Structure of GOPC PDZ domain (cyan) with iCAL36 (yellow) bound (PDB:4E34).[158] Type I, II, and III PDZ binding motifs. B) Overlay of iCAL36 (yellow) and the predicted interactions of KCTD9$^{381\text{-}389}$ (magenta) in the binding site of GOPC^{PDZ} (blue). C) Average pLDDT plot (residue) of the KCTD9 peptide for KCTD9$^{381\text{-}389}$/GOPC^{PDZ} Alphafold3 prediction. D) 2D PAE plot of the KCTD9$^{380\text{-}389}$/GOPC^{PDZ} complex. Coloured by PAE score.

The C-terminal sequence of full-length KCTD9 does not match any class of PDZ binding motifs. The final residue of full length KCTD9 is a charged arginine (R389), and not the typical hydrophobic residue seen for other interactors. Indeed, when the last 9 residues of KCTD9 (381-PLHMSQSVR-389) are modelled with the GOPC^{PDZ}, the peptide does not adopt the same binding mode

as iCAL36: The terminal carboxylate of KCTD9 is not positioned at the same location as bona fide GOPC interactor peptides, and the hydrophobic pocket of the PDZ domain is occupied by V388 of KCTD9 instead (Figure 4.15B). Despite this, the pLDDT and PAE scores for this model are high (Figure 4.15C), indicating there is some confidence in this prediction.

CoCoPred (a coiled-coil protein prediction server) indicated that GOPC is likely to form a homodimer.[159] When two copies of GOPC were modelled with a pentameric KCTD9, a dimeric GOPC was positioned alongside the PPR domains of KCTD9. In this model, however, the coiled-coil domain and not the PDZ domain of GOPC is interfacing with KCTD9 instead (Figure 4.16A). The overall pLDDT and PAE scores suggest that there is little confidence in this prediction of the KCTD9/GOPC complex (Figure 4.16B). Indeed, when KCTD9A-c067 was transiently expressed in HEK293 cells and immobilised on FLAG resin, endogenous GOPC was not immunoprecipitated alongside KCTD9/CUL3 (Figure 4.16C). Treatment of these cells with MLN4924, did not prompt capture of GOPC, nor did it alter protein levels in the input sample either. Overall, these observations suggest that GOPC is unlikely to be an interactor of KCTD9.

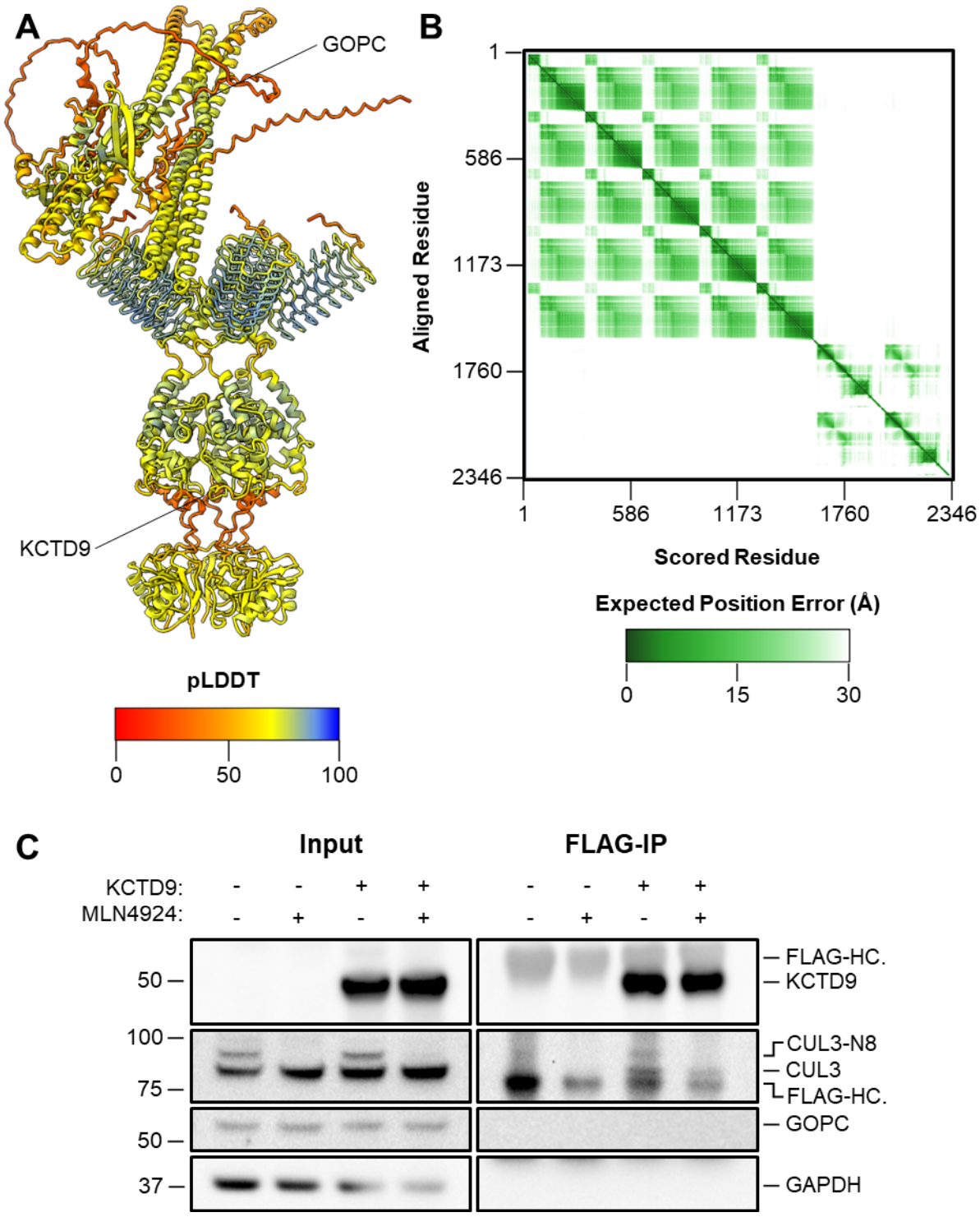


Figure 4.16: A) Alphafold3 of dimeric GOPC with pentameric KCTD9. Coloured by pLDDT score. B) 2D PAE plot of Alphafold3 model. Coloured by PAE score. C) FLAG-IP western blot. KCTD9A-c067 was transiently expressed in HEK293 cells and immunoprecipitated using FLAG resin. Samples were immunoblotted for FLAG, CUL3, GOPC, and GADPH.

The PPR domain is uncommon to eukaryotes and potentially examining other examples of this domain could provide some indications about the nature of a KCTD9 interactor. MfpA is a bacterial pentapeptide repeat protein that mimics T-segment DNA in order to bind DNA gyrase and provide resistance to Fluoroquinolones, a class of antibiotics.[68] I postulated that the PPR domain of KCTD9 could act as DNA/RNA mimic and that a DNA/RNA binding protein could be an interactor. A review by H. Wang *et al.* noted that DNA mimic proteins have diverse range of structures and no overall census motif.[160] While this makes predicting new DNA mimics more challenging, these proteins do have a regular arrangement of negatively charged residues (aspartic acid and glutamic acid) at the surface; which mimics the charge distribution of B-form DNA. Examining the PPR domain of KCTD9, while the outward faces are not as negatively charged as MfpA, the placement of negatively charged residues on faces 2 and 3 somewhat mimic B-form DNA (Figure 4.17).

DAVID analysis (Table 4.7) identified a group of proteins which contain an RRM domain or are implicated in RNA binding (cluster 6). NUP35 stood out of this cluster as potential interactor because it had been previously identified as a potential interactor of KCTD9 in multiple Y2H assays (BioGrid).[161] This protein is a key component in the biogenesis of the nuclear pore complex (NPC), a 120 MDa assembly that regulates the movement of molecules between the cytoplasm and nucleus (Figure 4.18A).[162] Other nucleoporin family members were also captured in MS proteomics experiments, but the fold-change of NUP35 between conditions 1 and 3 is much greater than the rest (Table 4.11), indicating that there could be a specific interaction between NUP35 and KCTD9.

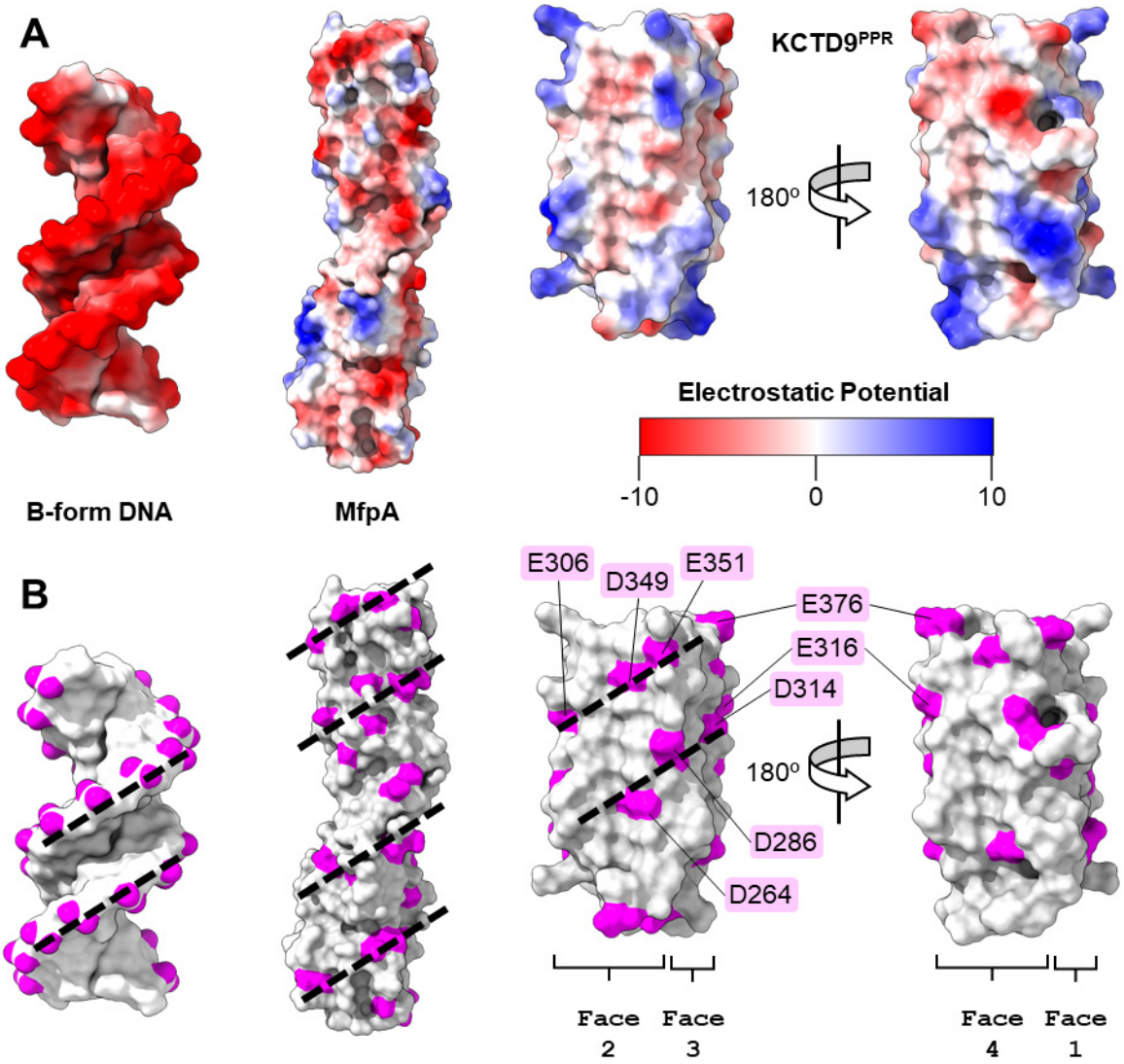


Figure 4.17: A) B-form DNA, MfpA, and KCTD9^{PPR} (Left to right) surfaces coloured by electrostatic potential. B) External facing oxygen atoms of B-form DNA and the negative residues (Asp and Glu) of MfpA and KCTD9 are coloured magenta. Diagonal arrangement of negative residues mimic that of B-form DNA are shown by dotted line.

No.	Gene Name	-Log Student's T-test p-value	Student's T-test Difference	Protein Intensity								
				Cond. 1			Cond. 2			Cond. 3		
				R1	R2	R3	R1	R2	R3	R1	R2	R3
1	NUP205	1.10	0.58									
2	NUP35	1.89	**3.01**									
3	NUP93	1.12	0.47									
4	NUP98	0.06	0.27									
5	NUP43	0.01	-0.05									
6	NUP155	1.14	-0.25									

Table 4.11: All the nucleoporin members detected in the NK92 MS proteomics experiments. Student's T-test analysis and fold-change analysis of conditions 1 and 3. The relative protein intensity of each protein in each replicate is coloured red-white-blue (high to low). Dark grey indicates no protein was measured for that sample.

Cryo-ET structures have resolved the NPC, in a constricted and dilated state, and NUP35 forms a dimer with a NUP35 unit in an adjacent subcomplex.[163] Several NUP35/KCTD9 models were predicted with Alphafold3 using different sequence lengths, assembly ratios, and with additional nucleoporin components (NUP93 and/or NUP155). Across all these models, various NUP35/KCTD9 interactions were predicted and it is difficult to identify any single model over the others; as the confidence metrics (pLDDT and PAE scores) are equally low throughout. This exemplified by Alphafold3 predictions of two $NUP35^{RRM}$ with a pentameric KCTD9 complex (Figure 4.18B,C). Both monomeric and dimeric states of the RRM domain (H170-E250) are predicted, and the position of these domains alongside KCTD9 are varied. Additionally, there is no predicted interaction of the RRM domains and the negatively charged surface of the PPR domain. Despite the lack of promising predictive models, I believe there is enough other experimental evidence to consider evaluating NUP35 as a potential interactor by conducting follow-up immunoblotting experiments.

Another protein with an RRM domain that seemed interesting was ZNF638 (Zinc Finger Protein 638). This large protein (~220 kDa) has a Matrin-type zinc finger domain, two RRM domains (S676-K751 and C905-E979), and several MH (mad homology) domains. It is able to bind cytidine-rich double stranded DNA and has been implicated in transcription regulation during adipogenesis.[164] ZNF638 is localised to the nucleoplasm and it was also identified in the pilot MS proteomics experiment using HEK293 cells.

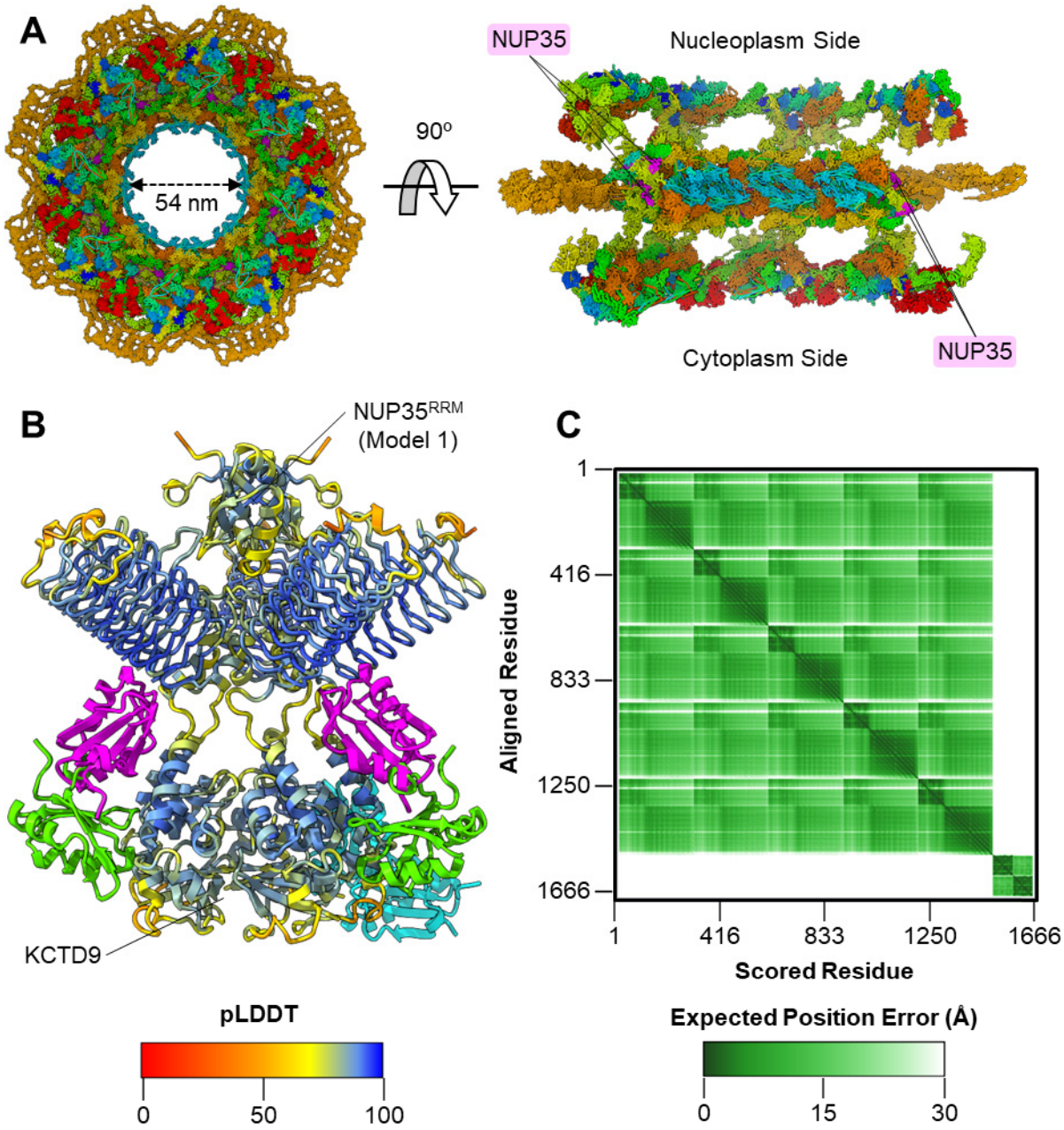


Figure 4.18: A) Nucleoporin complex in the dilated state (PDB: 7R5K).[163] Unique polymers are coloured by rainbow. B) Superimposition of dimeric NUP35^{RRM} and pentameric KCTD9 models. Model 1 is coloured by pLDDT score. NUP35^{RRM} chains in models 2,3 and 4 are coloured blue, cyan, and magenta respectively. C) 2D PAE graph of KCTD9/NUP35^{RRM} Alphafold3 model. Coloured by PAE score.

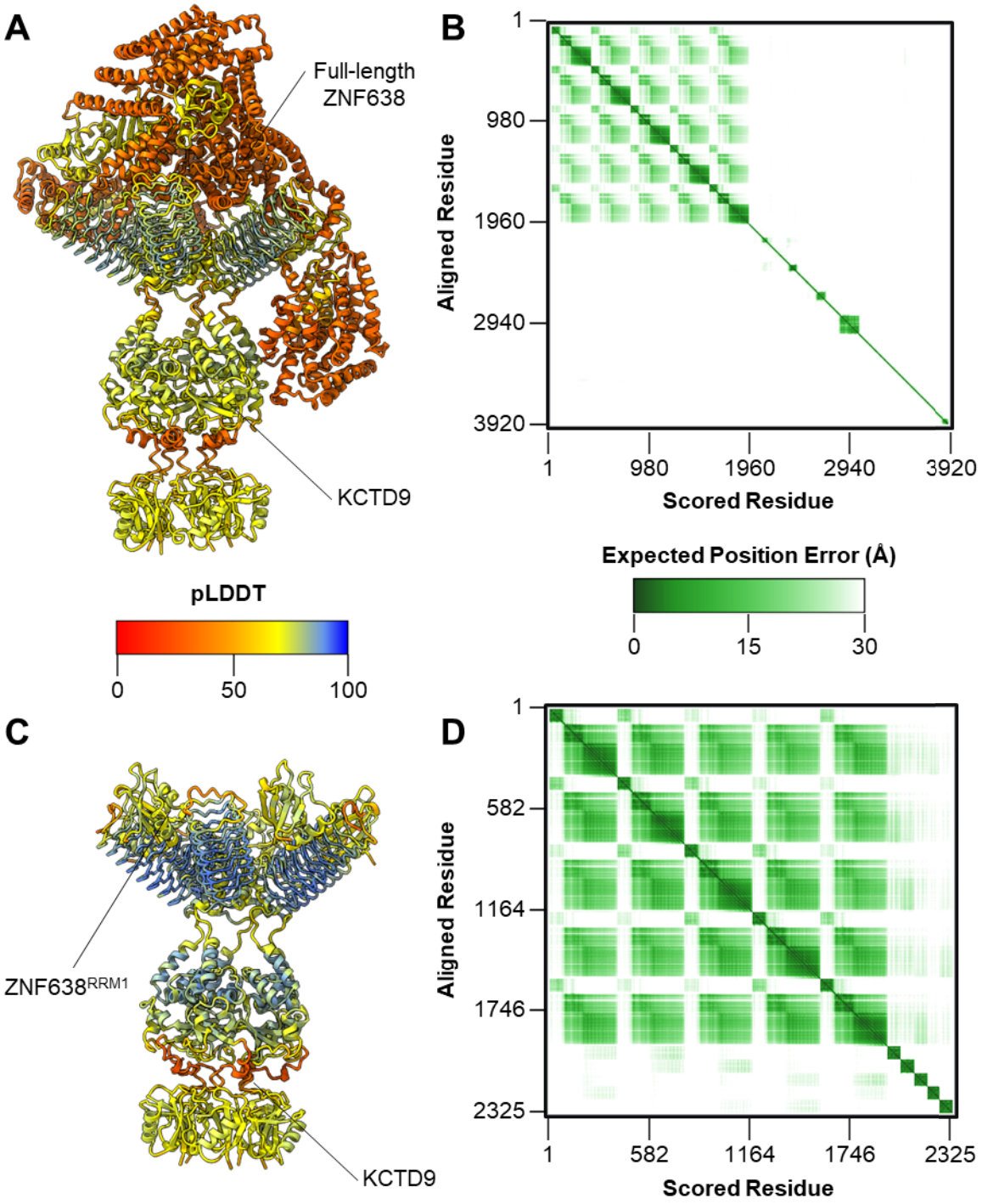


Figure 4.19: Alphafold3 predictions of KCTD9/ZNF638 complexes A) Monomeric ZNF638 with pentameric KCTD9 model. Coloured by pLDDT score. B) 2D PAE plot of KCTD9/ZNF638 complex. Coloured by PAE score. C) Pentameric ZNF638^{RRM1} with pentameric KCTD9 model. Coloured by pLDDT score. B) 2D PAE plot of KCTD9/ZNF638^{RRM1} complex. Coloured by PAE score.

Due to the large size of ZNF638, the full-length protein can only be modelled in a monomeric state with pentameric KCTD9 (Figure 4.19A,B). The resulting model has very low pLDDT and PAE scores; which is likely due to the large amount of intrinsically disordered sections being modelled. Instead the ordered RRM1, RRM2, and the zinc finger domains were modelled together, separately, and in various oligomeric states. Like in the case for NUP35 modelling, there was a variety of arrangements of the ZNF638 domains with respect to KCTD9 but there was a propensity for Alphafold3 to predict $KCTD9^{PPR}$/$ZNF638^{RRM1}$ interactions and this yielded some slightly higher PAE scores than was seen for other models (Figure 4.19C,D). This model is to be treated sceptically, but I would suggest including ZNF638 in follow-up immunoblotting experiments alongside NUP35.

4.3 Discussion

Identifying novel interactors and characterising the specific interactions with KCTD9 would illuminate the roles of this unusual cullin-RING substrate adaptor. To that end, I first tried recombinant pulldown experiments of TRIM proteins with KCTD9A-c047. None of the TRIM RING domain constructs tested were shown to interact with KCTD9 but this result was not that surprising, given that high-throughput Y2H assays are known to have many false positives.[140] Potentially the interaction may occur at a different region of the TRIM protein than was tested, but this would be a different interaction to that was observed between SopA and TRIM56.

ZNT9 is (so far) the only interactor of KCTD9 to be reported. Unusually, KCTD9 is shown to suppress Wnt signalling by destabilising β-catenin and promoting its degradation, yet KCTD9 does not ubiquitinate ZNT9 or β-catenin.[73] I tried to

generate recombinant ZNT9 constructs to verify the reported interaction and potentially generate KCTD9/ZNT9 cryo-EM datasets. The constructs expressed in *Sf9* cells produced some soluble protein, but the yield varied between expression batches. One attempt to co-express KCTD9A-c042 with ZNT9A-c007, only yielded the ZNT9 construct and no KCTD9/ZNT9 complex was observed. Future attempts to co-expression ZNT9 with another, more optimal KCTD9 construct (such as KCTD9A-c047) may yield a complex that could then be further characterised.

Novel protein interactors of KCTD9 were sought by conducting affinity mass spectrometry proteomics in HEK293 cells. While I was able transiently express KCTD9A-c067 and show that the CRL complex can survive immunoprecipitation on FLAG resin, there were few potential interactors and no obvious candidates that related to NK cell development. I was advised to conduct this experiment in NK92mi cells with more biological replicates so that statistical analysis (Student's T-test) could be performed.

The second MS proteomics dataset identified many proteins that may require further evaluation (Table 4.5 and Table 4.6) and this list is currently under analysis. KCTD9/interactor models were generated using Alphafold3 to filter the potential interactor list and identify which targets to follow-up with immunoblotting experiments. Initially, GOPC stood out from the dataset as it was only present in conditions 2 and 3 (where KCTD9A-c067 was also immunoprecipitated). However, the Alphafold3 predictions of a KCTD9/GOPC complex did not indicate that there were any specific interactions and no interaction was observed in immunoblotting experiments.

As the PPR domain of MfpA has been shown to mimic DNA, I speculated that KCTD9 could interact with DNA/RNA binding proteins by mimicking the structure of a polynucleotide. I tried to model KCTD9 with NUP35 and ZNF638, DAVID analysis of the dataset groups these proteins into cluster 1 (nucleoplasm associated proteins), and cluster 6 (DNA/RNA binding proteins). Additionally, NUP35 is reported to interact with KCTD9 in several high-throughput Y2H assays, and ZNF638 was one of the few proteins identified in the HEK293 MS proteomics dataset. Although no convincing Alphafold3 models with these proteins have been generated, I believe there is enough experimental evidence to evaluate these two targets further; even if it is just to rule them out as non-interactors.

One interesting observation is that ZNT9 was not detected in either of the two MS proteomic datasets. Potentially, this interactor is less abundant in HEK293 and NK92 cells than HCT-116 cells. Future immunoprecipitation experiments should include ZNT9 as target to verify it as an interactor, and then it could serve as a positive control for other targets.

There are plenty of proteins identified from the NK92 MS proteomics dataset that require evaluation, particularly proteins which are implicated in NK cell development (such as OAS2 and BCL11A), and I am optimistic that a genuine interactor can be identified from this experiment. The challenge is finding a method to effectively filter the list in order to conduct efficient follow-up immunoblotting experiments, which are time consuming and costly. It remains to be seen if Alphafold3 modelling can aid in this process, and if it is able to predict a novel KCTD9/interactor complex before it is validated by a physical experiment.

Chapter 5: Structural Insights into KLHL12 Interactor Recognition

5.0 Structural Insights into KLHL12 Interactor Recognition

5.1 Introduction

KLHL12 is a dimeric cullin-3 associated E3 ligase that recognises a variety of interactors through a common recognition motif. These interactors have distinct biological roles and as a result, KLHL12 in a range of biological processes. KLHL12 is one regulates of Wnt signalling by ubiquitinating Dishevelled proteins (DVL1-3).[81] Here, the dishevelled proteins are acting as KLHL12 substrates and are ultimately degraded by the 26S proteasome, which results in reduced β-catenin stabilisation. PEF1, on the other hand, acts as a KLHL12 co-adaptor to cooperatively recruit SEC31. This transport protein is only monoubiquitinated by KLHL12 to promote COPII complex formation and PEF1 remains unmodified.[84] Finally, PLEKHA4 serves as an antagonist to the E3 ligase by sequestering it to the plasma membrane. The substrate recognition site is occupied by PLEHKA4 and this prevents the KLHL12 from recruiting other interactors including dishevelled proteins (Figure 1.11).[86]

Previous work in the Bullock group identified a consensus PGXPP recognition motif within disordered regions of these interactor partners.[78] The group obtained a KLHL12^{KELCH}/DVL1$^{650-664}$ peptide co-crystal structure (PDB: 6TTK) in which the peptide adopted a U-turn conformation within the kelch pocket to facilitate several hydrophobic and hydrogen bond interactions (Figure 5.1A). Simultaneously, a KLHL12^{KELCH}/DVL3$^{679-688}$ peptide co-crystal structure (PDB: 6V7O) was deposited by Zhao *et al.* In this structure, the truncated peptide appears in an alternative binding mode (Figure 5.1B).[165]

As the 10-mer DVL3 peptide lacks the residues that precede the recognition motif, it is unclear if this alternative conformation is a structural artifact, or instead representative of a binding mode that some KLHL12 interactors may adopt. In this chapter, additional structural datasets using PEF1 and PLEKHA4 interactor peptides in complex with KLHL12 were generated to investigate if there is a consensus binding mode. Additionally, the relationship between the interactor peptide sequence and KLHL12 binding affinity was examined by conducting biophysical assays.

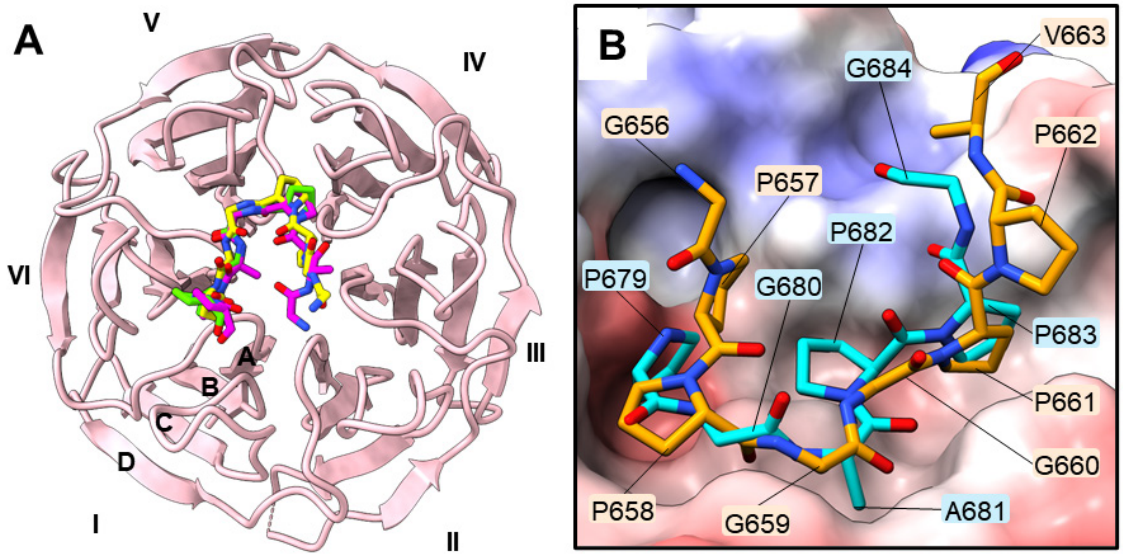


Figure 5.1: A) The KLHL12 kelch domain is made up of six blades (I-VI) in a β-propeller fold. The four β-strands of Blade I are labelled A-D. DVL1 (yellow), PEF1 (magenta), and PLEKHA4 (green) bind to the centre of this domain. B) Overlay of DVL1 peptide (orange, PDB: 6TTK) and DVL3 peptide (cyan, PDB: 6V7O) at the KLHL12 binding site.[78,165]

5.2 Generation of New KLHL12/Interactor Co-crystal Datasets

KLHL12A-c012 (His-TEV-KLHL12$^{268-567}$, 32.9 kDa) was expressed in *E. coli* and captured using Ni^{2+}-sepharose resin. The N-terminal his-tag was cleaved using TEV protease and purified by size exclusion chromatography. The sample was concentrated to ~10 mg/mL and incubated at room temperature with an 11-mer PEF1 peptide (UniProt: Q9UBV8, residues 15-25, GQA**PGAPP**GSY) at a final concentration of 2.4 mM for 1 hour. The protein-peptide mixture was applied to coarse crystal screens (Table 5.2) and the best-diffracting crystals of the KLHL12/PEF1 complex was obtained from a 150 nL sitting drop containing a 1:1 mix of the protein sample and reservoir solution (0.1 M Bis-Tris pH 5.5, 200 mM NaCl, and 30% PEG4K).

Peptide	Sequence	Use
DVL1 20-mer	AYTVVGGP**PGGPP**VRELAAV	Biophysical assays
DVL1 15-mer	AYTVVGGP**PGGPP**VR	Co-crystallisation (PDB: 6TTK)
DVL1 11-mer	GGP**PGGPP**VRE	Biophysical assays
DVL2 20-mer	PVPPAVQP**PGAPP**VRDLGSV	Biophysical assays
DVL3 20-mer	PPPAAMGP**PGAPP**GRDLASV	Biophysical assays
DVL3 10-mer	**PGAPP**GRDLA	Co-crystallisation (PDB: 6V7O)
PEF1 11-mer	GQA**PGAPP**GSY	Co-crystallisation (PDB: 6RBH) and biophysical assays
PLEKHA4 11-mer	PGG**PGGPP**EVS	Co-crystallisation (PDB: 8OIO) and biophysical assays
SEC31 11-mer	HGG**PGAPP**SSS	Biophysical assays
DVL1^{P657G} 11-mer	GGG**PGGPP**VRE	Biophysical assays
DVL1^{G660A} 11-mer	GGP**PGAPP**VRE	Biophysical assays

Table 5.1: Summary of interactor peptides referenced in this chapter (PGXPP recognition motif in bold, DVL1 point mutations coloured red).

Construct ID (Scarab)	Purification Result	Crystal Screen	Crystallisation Result	Diffraction Dataset
KLHL12A-c012	9.8 mg/mL	LFS, 20 ºC	None	-
		JCSG, 20 ºC	Several	1.88 Å, solved
		HIN3, 4 ºC	None	-
	9.7 mg/mL	HCS, 4 ºC	Several	None mounted
		JCSG C3d, 20 ºC	Several	None mounted
		JCSG H8a, 20 ºC	Several	None mounted
		JCSG H10a, 20 ºC	Several	None mounted
		HCS A9, 4 ºC	Several	None mounted
		HCS A9, 20 ºC	Several	>4 Å, not processed

LFS: Ligand-Friendly Screen HIN3: Hampton Research Index Screen
HCS: Hampton Crystal Screen
JCSG: Joint Center for Structural Genomics Screen

Table 5.2: KLHL12 purification and crystallisation summary table. Experiments shown are only those conducted by author. Best dataset is described in diffraction dataset column. Italics indicate fine crystal screen.

A KLHL12/PLEKHA4 (UniProt: Q9H4M7, residues 174-184, PGG**PGGPP**EVS) co-crystal was prepared by E. Dalietou using the same protocol. A diffracting crystal was only obtained after seeding KLHL12/PEF1 crystals (1:10) into a 150 nL sitting drop containing a 2:1 of the protein sample and reservoir solution (0.1 M citrate pH 5.5, 400 mM NaCl, and 27.1% PEG8K). The crystals were cryoprotected by addition of the reservoir solution (1 µL) containing 25 % ethylene glycol to the drop and vitrified in liquid nitrogen.

KLHL12/PEF1 and KLHL12/PLEKHA4 co-crystal datasets were obtained, at a resolution of 1.9 and 2.0 Å respectively, by unattended data collection on the I04 beamline at Diamond Light Source, Didcot, UK. In the instance of the KLHL12/PEF1 dataset, two sweeps were merged and pre-processed by the xia2-dials automated pipeline at the beamline. While R-merge was higher between batches 3000-3300 (Figure 5.2C), generally radiation damage of the crystal did not appear to significantly affect the quality of the dataset. When the

dataset was reprocessed with omission of batches where the R-merge values were slightly higher, there were only minor improvements refinement statistics. The KLHL12/PLEHKA4 dataset also did not suffer serious radiation damage and was processed by the autoPROC automated pipeline (Figure 5.2D). For both datasets, the initial outputs from these automated pipelines were used for phasing and structure refinement. These pipelines generated datasets with suitable quality for fitting of the interactor peptide within the KLHL12 kelch binding pocket.

Molecular replacement was performed with Phenix.phaser-MR using a $KLHL12^{KELCH}$ structure (PDB: 2VPJ) as the search model. I manually built the PEF1 chain using COOT and refined the co-crystal structure with Phenix.refine and PDB validation tools (Table 5.3).[116,117] Although the clash score was a little high for this structure, I deemed it passable as most of the clashes arose from the $KLHL12^{KELCH}$ domain (which has been resolved in previous datasets), and the structure of the PEF1 peptide is the key information of this study. E. Dalietou built and refined the KLHL12/PLEKHA4 structure in a similar manner.

The KLHL12/PEF1 structure contained two complexes in the asymmetric unit (ASU), and the PEF1 peptide was found in the same binding region of KLHL12 as the DVL1 peptide in both chains. The KLHL12/PLEKHA4 structure contained four complexes in the ASU (average RMSD of 0.325 Å) and all the KLHL12 binding sites were occupied by the peptide. The structures were uploaded to the PISA server for buried surface area (BSA) calculations.[134,166]

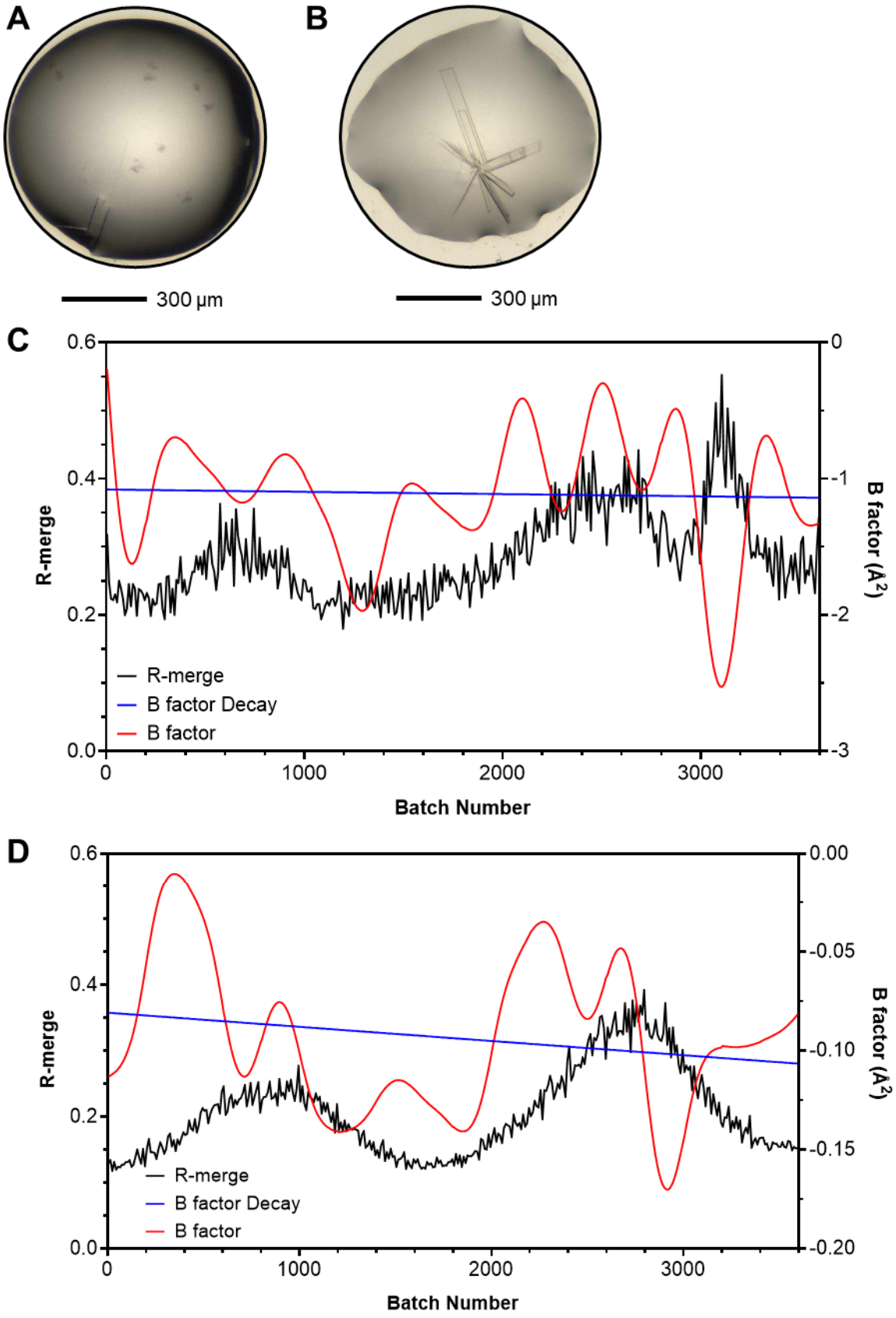


Figure 5.2: A) KLHL12/PEF1 co-crystals in a 150 nL drop, 7 days after dispensing. Formulatrix Rock Imager. B) KLHL12/PLEKHA4 co-crystals in a 150 nL drop, 14 days after dispensing. C) R-merge (black), B factor, and B factor decay (red and blue) statistics of the merged sweeps for KLHL12/PEF1 dataset (3600 images total). D) Statistics for the KLHL12/PLEKHA4 dataset (3600 images total).

	KLHL12/PEF1 PDB: 8RBH	KLHL12/PLEKHA4 PDB: 8OIO
Data collection		
Space group	*P 1 2_1 1*	*P 1 2_1 1*
Unit cell		
a, b, c, (Å)	45.0, 78.4, 73.8	80.5, 73.1, 103.7
α, β, γ, (°)	90, 99.2, 90	90, 98.8, 90
Resolution range (Å)	72.83 - 1.88 (1.92 - 1.88)	79.52 - 1.95 (1.99 - 1.95)
R-merge	0.273 (3.250)	0.188 (2.416)
R-meas.	0.322 (3.842)	0.222 (2.879)
CC1/2	0.991 (0.370)	0.996 (0.320)
Completeness (%)	97.20 (94.30)	100 (99.9)
Mean I/sigma (I)	6.2 (0.6)	7.3 (0.8)
Redundancy	6.9 (7.0)	7.0 (6.7)
Refinement		
Resolution	1.88	1.95
Reflections used in refinement	40,015 (3858)	85,940 (8317)
Reflections used for R-free	1986 (171)	4317 (404)
R-work	0.209 (0.347)	0.199 (0.342)
R-free	0.254 (0.364)	0.250 (0.389)
Number of non-hydrogen atoms	4799	9175
Macromolecules	4443	8747
Ligands	14	58
Solvent	348	390
R.M.S. deviations		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	1.01	0.99
Ramachandran favoured (%)	95.37	96.46
Ramachandran allowed (%)	4.46	3.45
Ramachandran outliers (%)	0.17	0.09
Rotamer outliers (%)	0.43	0.21
Clash score	15.02	11.04
Average B-factor (Å2)	31.14	35.77
Macromolecules	30.68	35.65
Ligands	43.32	45.62
Solvent	36.66	37.51

Table 5.3: Data collection and refinement statistics for KLHL12^{KELCH}/peptide co-crystal structures. Statistics for the highest-resolution shell are shown in parentheses. Data collected at beamlines I03 and I04, Diamond, Didcot (Proposals MX19301 and M28172). λ = 0.979 Å.

5.2.1 Structural Observations

The Kelch domain structures in the KLHL12/PEF1 structure were in close agreement with those in the KLHL12/DVL1 structure (average RMSD: 0.603 Å – all atoms). The electron density for residues G15-P22 of chain C could be resolved and this showed the PEF1 peptide adopting the same U-turn conformation that was observed with the DVL1 peptide (Figure 5.3). The side chain of Q16 extended into the solvent and lacks electron density to complete model building for this residue. Additionally, the final proline of the PGXPP motif appears to be more flexible as the electron density was only partially resolved for this residue in both the PEF1 and DVL1 structures. Contrastingly, in chain D the side chain of A17 is rotated nearly 180º to what is observed in chain C, and the N-terminus of the peptide extends into a small cleft of the binding pocket instead. Here, G15 lacks resolvable density to be modelled and the peptide occupies a more open confirmation than is observed in any other KLHL12/interactor peptide structure.

The KLHL12/PLEKHA4 structure appeared to be more dynamic than the other structures as a loop region (Y351-D356) of the kelch domain lacked density to be modelled in all chains (Figure 5.4). Furthermore, only the 117-PGGPP-181 residues of the PLEKHA4 peptide could be confidently resolved and neither the position of G175 and G176 could be assigned to the nearby electron density. Despite this, the residues that were modelled did overlay well with the other interactor peptide structures and this suggests that the PLEKHA4 peptide also adopts a U-turn conformation but that it wasn’t able to be definitively resolved in this dataset.

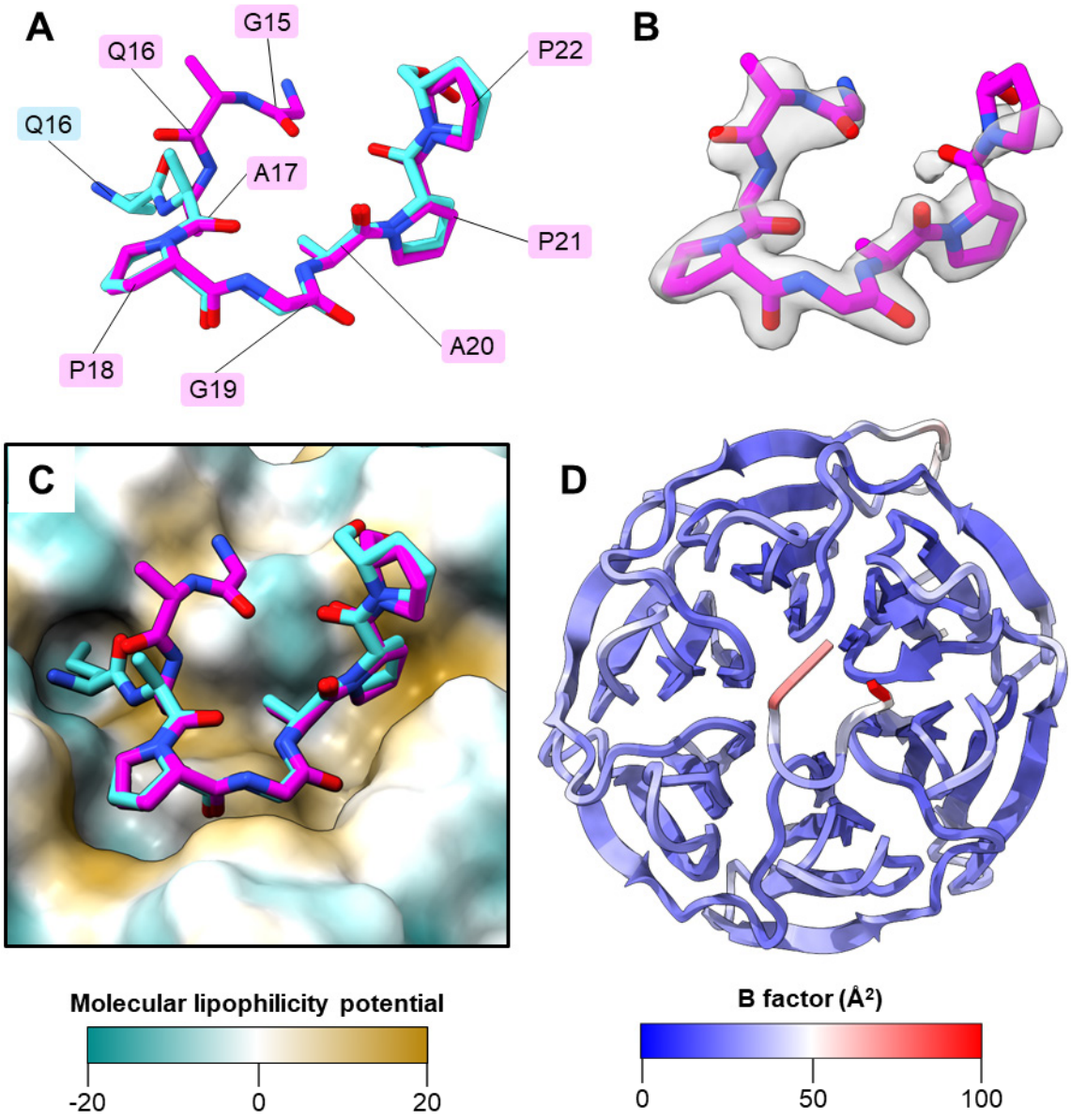


Figure 5.3: A) Overlay of PEF1 peptide from chains C and D of the KLHL12 co-crystal structure. B) Electron density of chain C cutoff at 1σ. C) PEF1 peptide within the KLHL12 binding pocket. KLHL12 surface coloured by the molecular lipophilicity potential. D) KLHL12/PEF1 (chain A and C respectively). Coloured by B factor.

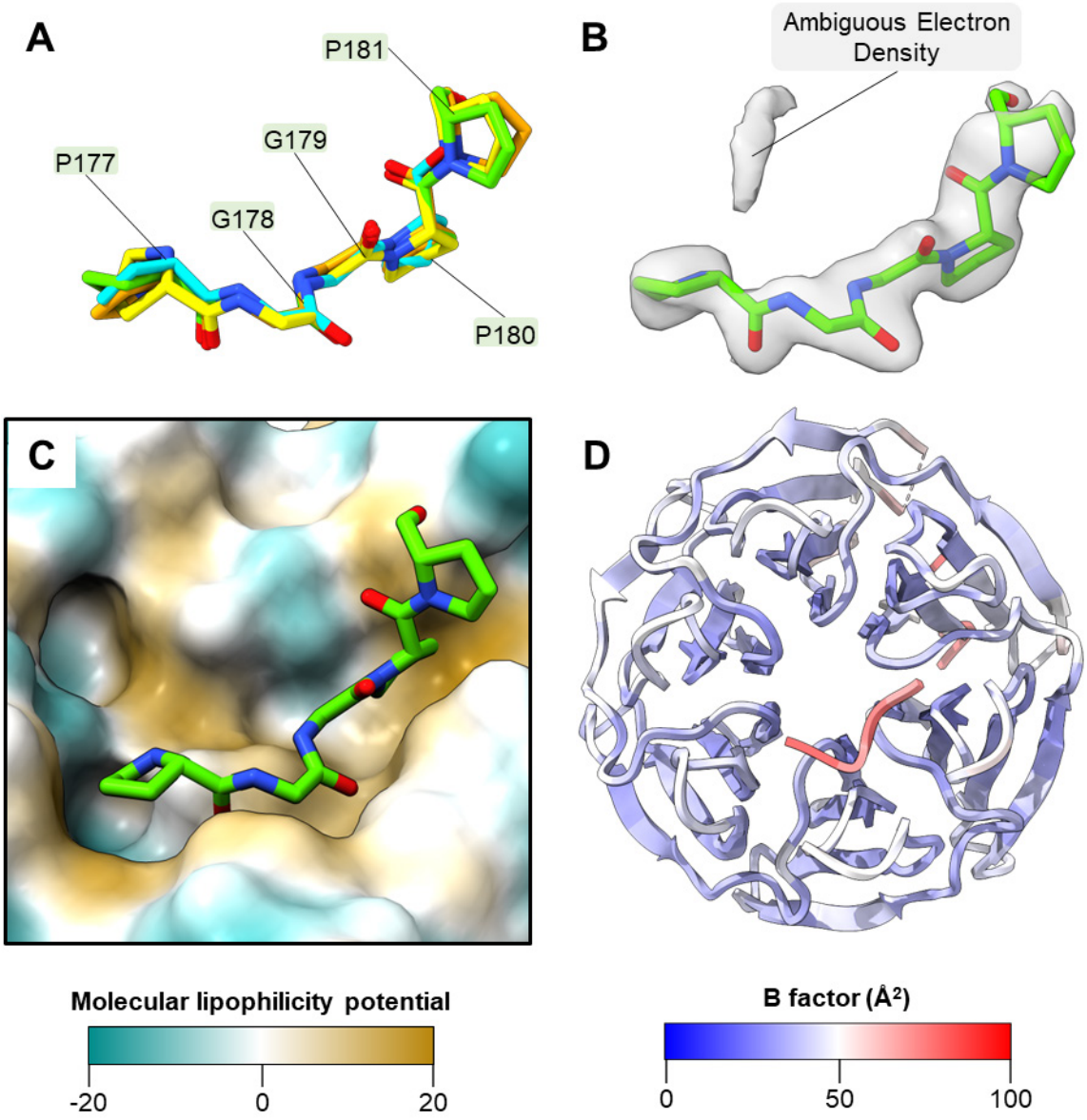


Figure 5.4: A) Overlay of PLEKHA4 peptide from chains C and D of the KLHL12 co-crystal structure. B) Electron density of chain C cutoff at 1σ. C) PLEHKA4 peptide within the KLHL12 binding pocket. KLHL12 surface coloured by the molecular lipophilicity potential. D) KLHL12/PLEKHA4 (chain A and E respectively). Coloured by B factor.

Comparison of the KLHL12/DVL1 co-crystal structure (PDB: 6TTK), PEF1 co-crystal structure (PDB: 8RBH), and PLEKHA4 co-crystal structure (PDB: 8OIO) reveal a common mix of hydrophobic and hydrogen bonding interactions between the PGXPP region of the peptides and the kelch domain (Figures 5.5, 5.6, and 5.7). In all three structures, there is a hydrogen bond between the phenolic hydroxyl group of KLHL12 Y512, situated beneath the peptide, and the carbonyl of the first proline in the PGXPP motif (P18, P177, and P658 for PEF1, PLEKHA4,

and DVL1 respectively). Additionally, the PEF1 and PLEKHA4 peptides form a second interaction between the first glycine of the motif and H486. A third hydrogen bond interaction is observed between Q293 of the kelch domain and the final proline of the two peptides. Finally, in the case of the PEF1 and DVL1 peptides, an intramolecular hydrogen bond is present between the nitrogen of the variable group (alanine or glycine) and the residue preceding the PGXPP motif (A17 or P657 respectively). While the hydrogen bond acceptor residue is not resolved in the PLEKHA4 structure, I postulate that this interaction is likely to be present. This notion is supported by Alphafold3 predictions of a KLHL12/PLEKHA4 complex, where P176 of PLEKHA4 is in the same position of P657 of DVL1 (not shown). The ambiguous density could be an indicator of the PLEKHA4 peptide adopting a U-turn confirmation, but the poorer quality of this dataset (when compared to the other structures) leaves the intramolecular hydrogen bond unresolved. Potentially, a higher quality dataset generated in alternative crystallisation conditions could resolve this region of the peptide and confirm this interaction.

The kelch pocket complements this confirmation and facilitates several hydrophobic interactions. Most notably, there are π-stacking interactions between the second proline (P21, P180, and P661 for PLEKHA4, PEF1, and DVL1 respectively) and the aromatic residue Y528 of KLHL12. This proline residue is also flanked by P289 and, to a lesser degree, R320. The position of this proline appears to be critical as the sharp turn enabled by this residue positions the c-terminal of the peptide upwards and away from the first blade of the kelch domain; avoiding potential steric clashes. At the start of the PGXPP motif, the first two residues sit parallel to the aromatic residues Y434 and F481 of KLHL12. Lastly,

in the PEF1 and DVL1 structures, Y321 of KLHL12 is situated beneath the residue preceding the PGXPP motif (G17 or P657 respectively) and mediates some van der Waals interactions.

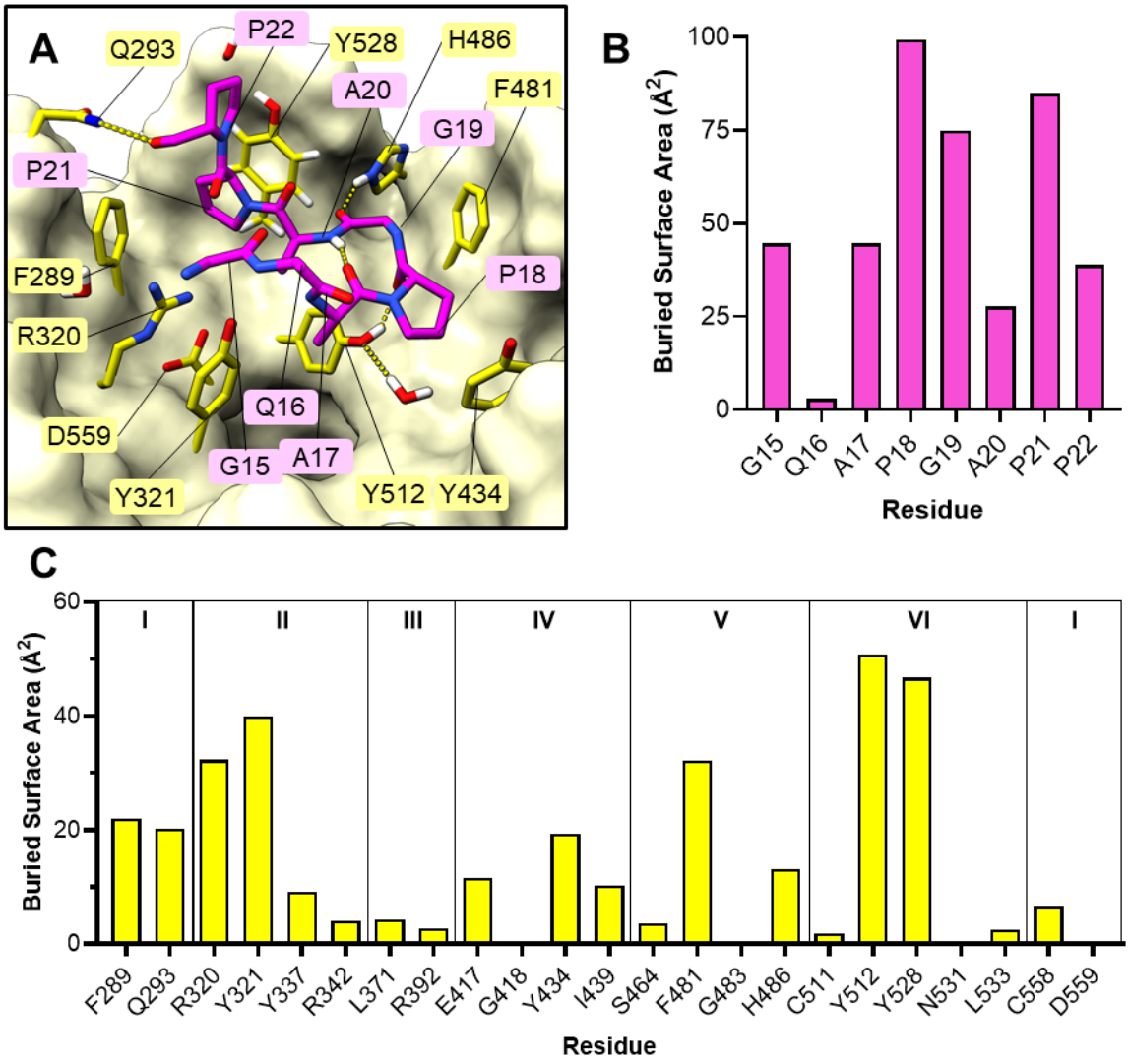


Figure 5.5: A) PEF1 peptide chain C (magenta) and contact KLHL12 residues (yellow). B) Buried surface area of PEF1 peptide chain C with KLHL12 chain A. C) Buried surface area of KLHL12 chain A with PEF1 peptide chain C. Only KLHL12 residues which have had a buried surface area with at least one of the of interactor peptides (DVL1, PEF1, or PLEKHA4) are shown.

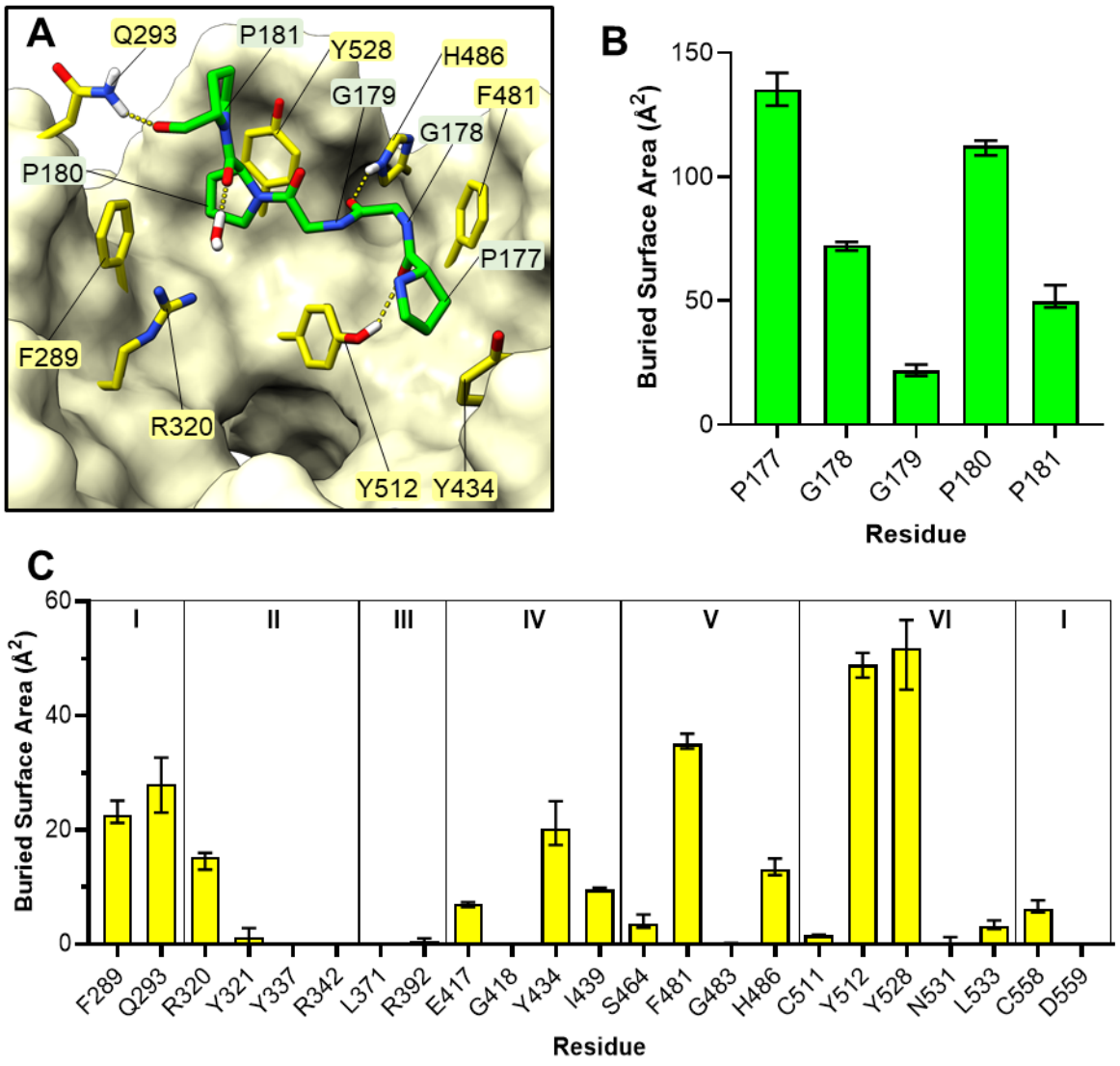


Figure 5.6: A) PLEKHA4 peptide chain G (lime) and contact KLHL12 residues (yellow). B) Average buried surface area of PLEKHA4 peptide chain E-H with KLHL12. Error bars represent the range in values between the chains. C) Average buried surface area of KLHL12 chain A-D with PLEKHA4. Only KLHL12 residues which have had a buried surface area with at least of the of interactor peptides (DVL1, PEF1, or PLEKHA4) are shown.

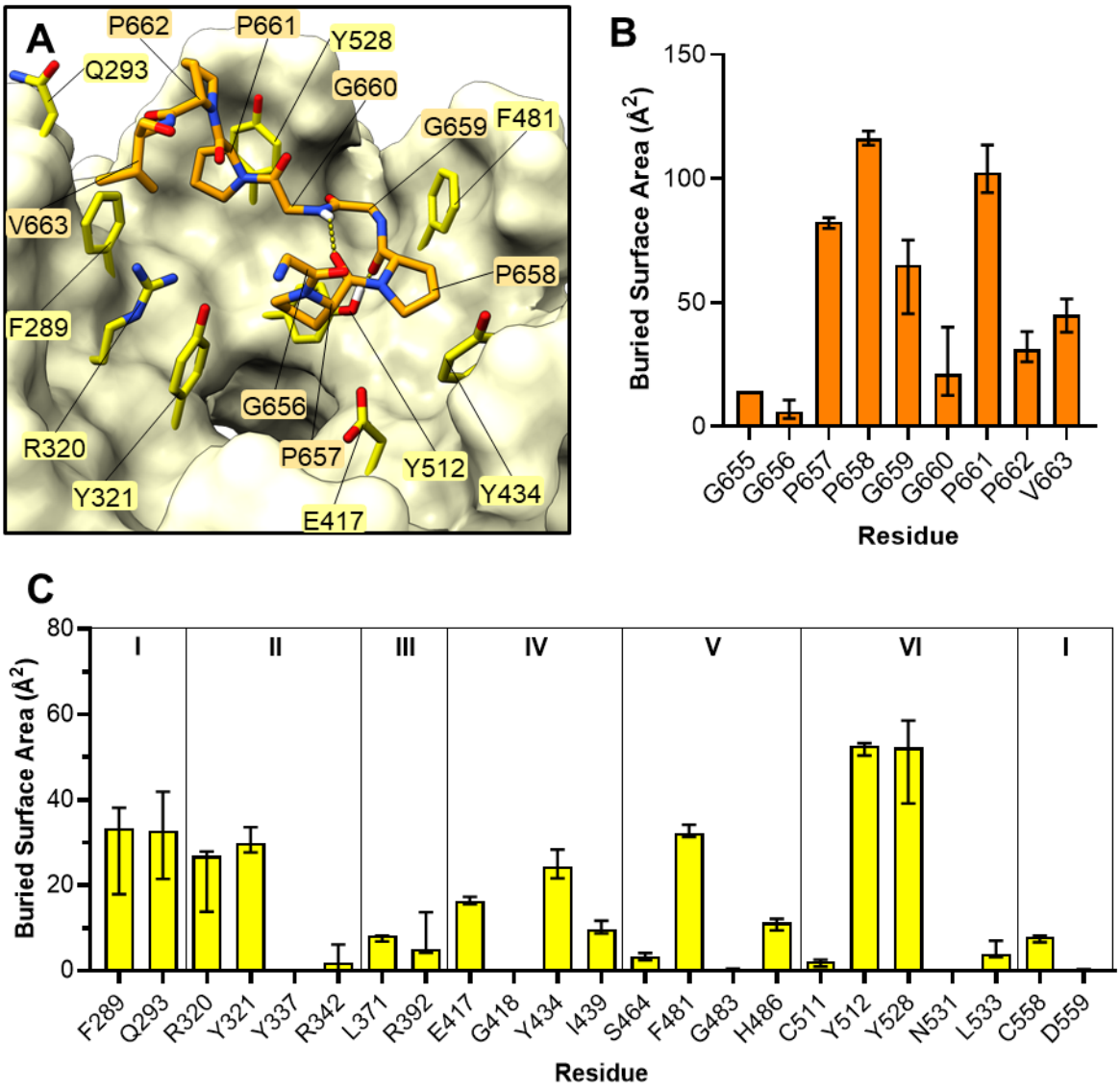


Figure 5.7: A) DVL1 peptide chain E (orange) and contact KLHL12 residues (yellow). B) Average buried surface area of the DVL1 peptide chain E-H with KLHL12. Error bars represent the range in values between the chains. C) Average buried surface area of KLHL12 chain A-D with DVL1. Only KLHL12 residues which have had a buried surface area with at least of the of interactor peptides (DVL1, PEF1, or PLEKHA4) are shown.

The buried surface areas (BSA) of the structures were calculated using the EMBL-PISA server. In all of the complexes, the interactor peptides have contacts with all six blades of the kelch domain (Figure 5.5, 5.6, and 5.7). In the case of the KLHL12/PLEKHA4 structure, the minimal buried surface area for kelch blades II-III is a result of the unmodelled residues prior to the PGXPP motif. The calculations for the kelch residues correlate to interactions noted previously and the residues are situated beneath the peptides, Y512 and Y528, have the highest BSA.

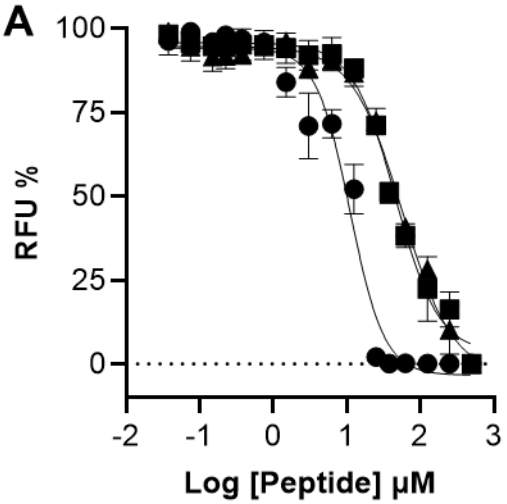
These calculations for the peptides identify the critical residues of the consensus PGXPP motif. In all cases, the first and second prolines in the motif have the greatest BSA; indicative of their key hydrophobic interactions with KLHL12. In contrast, the variable residues of the motif (alanine for PEF1, glycine for DVL1 and PLEKHA4) have a much smaller BSA. The interactions of this middle residue with the kelch domain are minimal and indicates how a variation in this residue does not impact the protein/peptide binding. Given how the side chain of the alanine is directed into the centre of the U-turn conformation and above the kelch domain centre, I hypothesis that only smaller residues (glycine or alanine) are tolerated. Side chains from larger or charged residues could impinge on the N-terminal region prior to the PGXPP motif, and prevent the peptide from adopting the U-turn conformation that is critical for binding.

5.2.2 Biophysical Measurements of Interactor Peptides

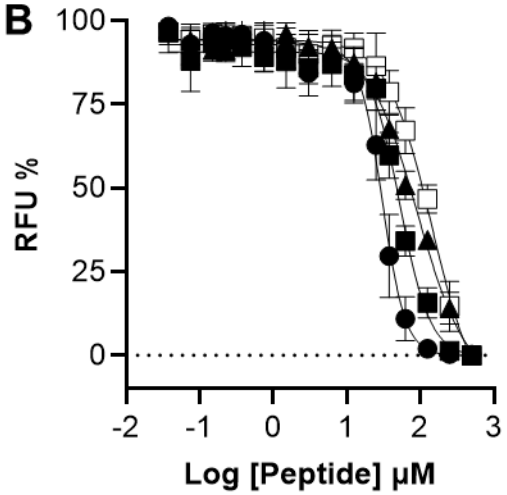
I wanted to examine the relationship between the interactor peptide sequence and KLHL12 binding affinity and so several interactor peptides containing the PGXPP motif were measured by Karly Buchan using an Alphascreen peptide

displacement assays (Figure 5.8). Their initial assays with 20-mer peptides measured the DVL1 peptide to have micromolar affinity (IC_{50} = 11 μM) to KLHL12, while the alanine containing DVL2 and 3 peptides were 4 to 6-fold lower (IC_{50} values of 45.8 μM and 56.6 μM, respectively. Figure 5.8A). Guided by the peptide sequences resolved in the structures, further measurements were performed using shorter 11-mer peptides. Surprisingly, the shorter DVL1 peptide had a lower apparent affinity (IC_{50} = 30.2 μM) than was observed in the first assay and I reason that there are some non-specific hydrophobic interactions between the longer peptide and kelch domain. The calculated Grand average of hydropathy (GRAVY) values of the longer peptides indicated that these were more hydrophobic than the shorter peptides.[167]

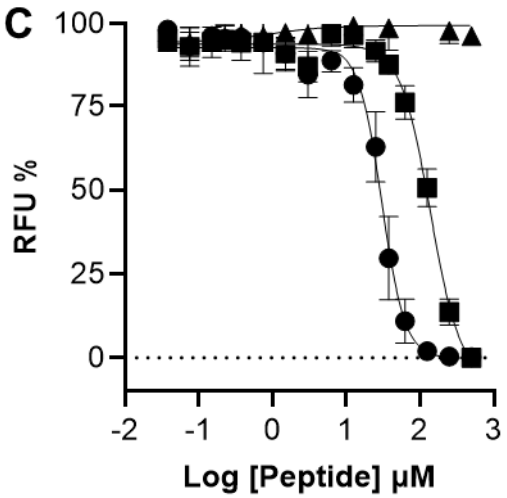
The PLEKHA4 peptide bound to KLHL12 with less affinity (IC_{50} = 52.6 μM) than DVL1, despite containing the same PGGPP binding motif sequence. I thought that the residue preceding the motif (P657 for DVL1, G176 for PLEKAH4) could impact the affinity and so a $DVL1^{P657G}$ mutant was tested. This peptide bound had a much lower affinity (IC_{50} = 138 μM) to KLHL12 than expected and it was a near 3-fold weaker than the PLEKHA4 peptide. The presence of an alanine instead of a second glycine in the PGXPP motif again appeared to reduce the affinity. The SEC31 and PEF1 peptides had IC_{50} values of 90.7 μM and 143 μM respectively, while a $DVL1^{G660A}$ mutant peptide showed no binding whatsoever. Interestingly, the PEF1 peptide had lower binding affinity to KLHL12 than PLEKHA4, despite appearing much clearer in the crystallography dataset. This could reflect that the KLHL12/PLEKHA4 dataset being more dynamic and less resolved, or that the crystallography conditions impeded the binding of the PLEKHA4 peptide.



Peptide	Sequence	IC_{50} (µM)	GRAVY
DVL1	20-mer	11.1	0.4
DVL2	20-mer	45.9	0.0
DVL3	20-mer	56.6	-0.2



Peptide	Sequence	IC_{50} (µM)	GRAVY
DVL1	GGP**PGGPP**VRE	30.2	-1.0
PLEKH4	PGG**PGGPP**EVS	52.6	-0.7
SEC31	HGG**PGAPP**SSS	90.7	-0.9
PEF1	GQA**PGAPP**GSY	143	-0.7



Peptide	Sequence	IC_{50} (µM)	GRAVY
DVL1	GGP**PGGPP**VRE	30.2	-1.1
P657G	GG**G****PGGPP**VRE	138	-1.0
G660A	GGP**PG****A****PP**VRE	n.b.	-0.9

Figure 5.8: Alphascreen biophysical assay results. SD error bars are plotted, n = 3. A) 20-mer DVL1-3 peptides (Table 5.1). B) 11-mer KLHL12 interactor peptides. C) 11-mer DVL1 mutants. Grand average of hydropathy (GRAVY) values for each peptide was calculated by the Expasy ProtParam server.[167]

To understand how the DVL1 mutations were affecting binding to KLHL12, I modelled several 20-mer DVL1 peptides with KLHL12^{KELCH} using Alphafold3 (Figure 5.9). In almost all models generated, including the DVL1^{G660A}, the peptides adopted the U-turn conformation seen in the KLHL12/DVL1 structure and had confident pLDDT and PAE scores. The only exception to this was the KLHL12/DVL1^{G660P} model, where due to the cyclised structure of the proline residue, the peptide is unable to adopt this binding mode. Instead, the 20-mer occupies an alternative conformation (not shown) with much lower confidence scores; indicating a DVL1^{G660P} mutant is unlikely to bind to KLHL12.

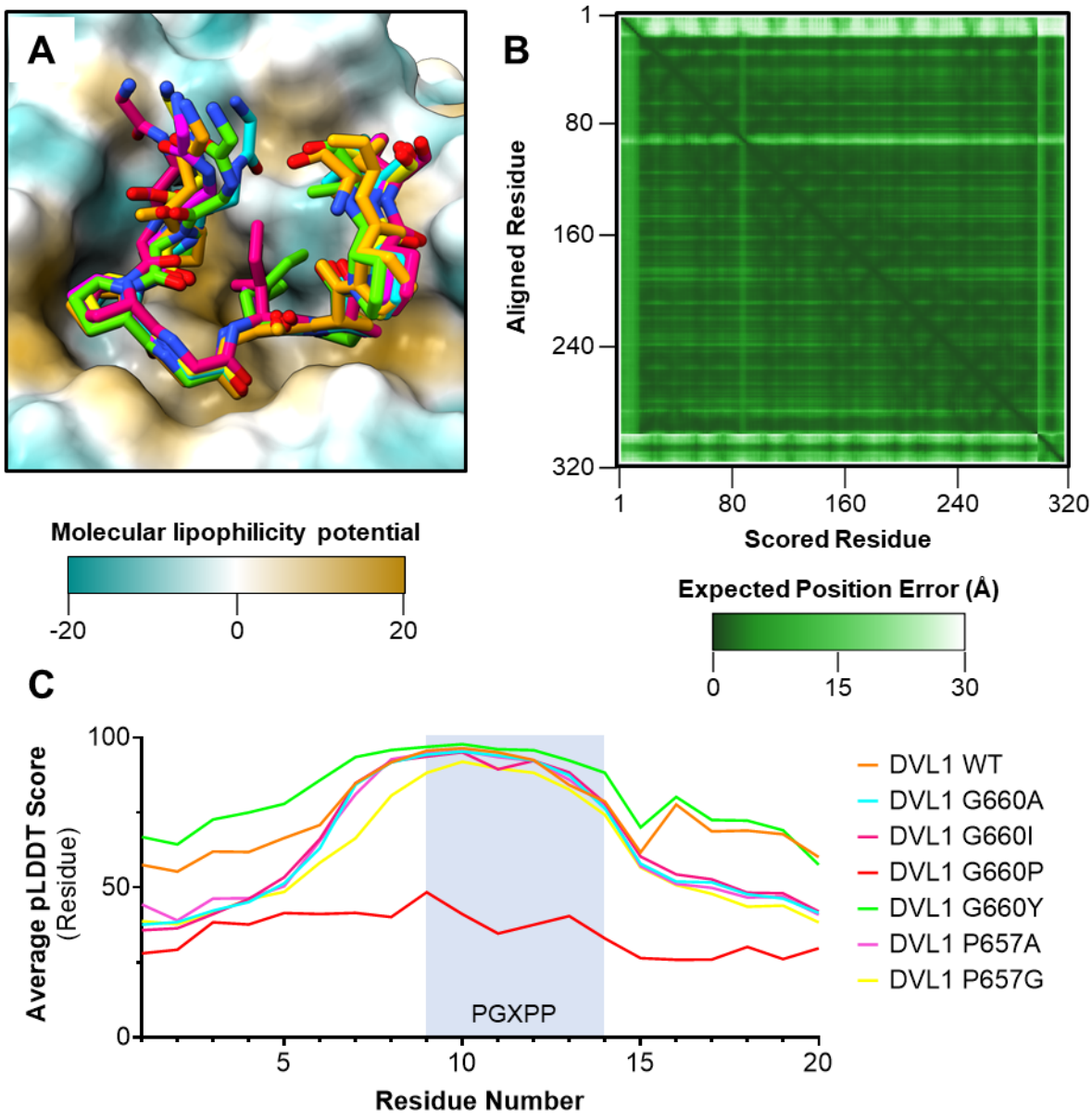


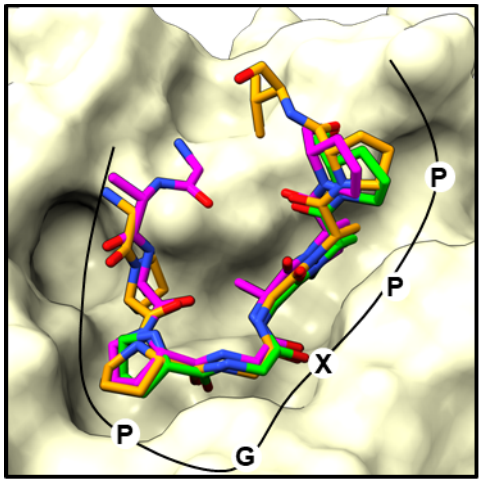
Figure 5.9: A) Overlay of the KLHL12^{KELCH}/DVL1 20-mers Alphafold3 models. Models superimposed by the kelch domain. Kelch surface is coloured by molecular lipophilicity potential. Residues 1-6, and 15-20 of the DVL1 peptides are omitted for clarity. DVL1 peptide from deposited KLHL12^{KELCH}/DVL1 (PDB: 6TTK) is coloured in orange. B) Representative 2D PAE graph of KLHL12^{KELCH}/DVL1^{WT} 20-mer model. Coloured by PAE score. C) Average pLDDT score (residue) of the DVL1 20-mers, residues 9-13 is the PGXPP motif region (shaded pale blue).

5.3 Discussion

The two structures presented in this chapter demonstrate that the PGXPP regions of PEF1 and PLEKHA4 partake in the same protein/protein interactions that was observed in the KLHL12^{KELCH}/DVL1 structure. The first and penultimate proline residues of the motif are involved in π-stacking hydrophobic interactions, whereas the middle variable residue has very little contact surface to the kelch pocket; indicating how variability in this position is tolerated. In the PEF1 and DVL1 structures the peptide adopted the same U-turn conformation, and (although there lacked resolvable electron density in the dataset to confirm this) Alphafold3 modelling suggests that PLEKHA4 can also (Figure 5.10).

The observed binding mode is different to the one reported in the KLHL12^{KELCH}/DVL3$^{679-688}$ structure (PDB: 6V7O). I hypothesise that the shorter 10-mer DVL3 peptide used in that study was unable to form a key intramolecular hydrogen bond between P678 (which it lacked) and A681. This interaction was observed in the both the PEF1 (A17↔A20) and DVL1 (P657↔G660) co-crystal structures, where it appears to stabilise the U-turn conformation. Without this interaction, the truncated DVL3 peptide adopted an alternative binding mode that is not representative of the KLHL12 interactor recognition.

The biophysical data identified some the relationships between the interactor peptide sequence and KLHL12 binding affinity. Firstly, the sequences the variable residue of the motif is substituted with an alanine (PG**A**PP) had a reduced the affinity to the kelch domain (DVL2, DVL3, SEC31, and PEF1). Secondly, sequences where the recognition motif is preceded by a proline (**P**PGXPP) (DVL1, DVL2, DVL3), displayed greater binding affinity than those preceded by a glycine (PLEKHA4, SEC31), or an alanine (PEF1).



Interactor	Sequence	Relationship
DVL1	657-P**PG**G**PP**-662	Substrate
PLEKHA4	176-G**PG**G**PP**-181	Antagonist
PEF1	17-A**PG**A**PP**-22	Co-adaptor
Consensus	**PGXPP**	

Figure 5.10: Overlay of PEF1, PLEKHA4, and DVL1 peptide structures adopting a U-turn conformation within the kelch binding pocket.

Following these observations, the affinity of two DVL1 mutants were measured (P657A and G660A) to confirm these relationships. Surprisingly, in the case of the DVL1^{G660A} mutant, the binding was completely abolished. This is somewhat an outlier when considering the fact that both the DVL2 and DVL3 peptides bound to KLHL12 with only slightly reduction less affinity to that of the wild-type DVL1 sequence. Furthermore, the P657G mutant showed a much greater reduction in affinity than was expected. At the recognition motif, this mutant shares the same six residues as PLEKHA4 (GPGGPP), but the wild-type antagonist peptide had an almost 3-fold stronger binding affinity. Alphafold3 models of DVL1 mutants suggest that many similar sequences could adopt the same U-turn conformation seen in the structural datasets, but these predictions do not give any insights into the kinetics of the transition between the relaxed-state and the U-turn conformation.

Considering all of the biophysical data, the relationship between the peptide sequence and binding affinity to KLHL12 appears to more complicated than it first appeared. While the PGXPP motif is critical for KLHL12 binding, the variation in the measured binding affinity between sequences where this motif is identical

(DVL1^{G660A} vs DVL2), indicates that the residues surrounding this region must also impact the ability of the peptide to adopt the U-turn conformation. Examining the sequences of KLHL12 interactors identified by the ELIM resource (ELME000536), the PGXPP motif is conserved and adjacent residues have much more variability (Figure 5.11). I hypothesise that the ability of the peptide to adopt the U-turn conformation is an important factor when considering KLHL12 binding affinity and that molecular dynamic simulations of the transition between the relaxed and U-turn conformational states may identify the key determinants of this.

A

B

DVL1:	650-	AYTVV GGP**PG** **GPP**VR ELAAV	-669
DVL2:	691-	PVPPA VQP**PG** **APP**VR DLGSV	-710
DVL3:	671-	PPPAA MGP**PG** **APP**GR DLASV	-690
PLEKHA4:	169-	QPGEG PGG**PG** **GPP**EV SRGEE	-188
SEC31A:	950-	GASFQ HGG**PG** **APP**SS SAYAL	-969
PEF1:	10-	CPGAA GQA**PG** **APP**GS YYPGP	-29
LNPK:	191-	GPPKD SSA**PG** **GPP**ER TVTPA	-210
CPSF6:	215-	DRFPG PAG**PG** **GPP**PP FPAGQ	-234
CPSF6:	340-	PPHLP GPP**PG** **APP**PA PHVNP	-359
DAAM1:	564-	PPPPP PLP**PG** **GPP**PP PGPPP	-583
DAAM2:	561-	PPGGP PTP**PG** **APP**CL GMGLP	-580

Figure 5.11: A) KLHL12 PGXPP motif region sequence logo (residues coloured by amino acid chemistry). Generated by WebLogo3 server.[70] B) PGXP motif regions of KLHL12 interactors identified in the ELIM resource (ELME000536).[168] PGXPP motif in bold.

KLHL12 is a rare example of an E3 ligase can identify a variety of interactors, each with different roles and cellular activity, through a shared recognition motif and binding mode. In cellular conditions, KLHL12 and these interactors are a part of a complicated system to regulate its ubiquitination activity: While dishevelled DVL1-3 proteins are simple substrates for KLHL12 mediated degradation, the antagonist PLEKHA4 is one mechanism to limit its ubiquitination activity.[86] As PLEKHA4 clusters are located at the plasma membrane, this mechanism also has the effect of raising the local concentration of KLHL12. If the E3 ligase is able to disassociate from PLEKHA4 and not defuse across the cytoplasm, KLHL12 activity would be increased for substrates that are also localised to this area of the cell.

Lastly, dimeric KLHL12 utilises the co-adaptor PEF1 (in complex with ALG2) to cooperatively recruit SEC31 and enable monoubiquitination.[84] The presence of a co-adaptor intermediate enables the cell to fine tune the level of monoubiquitination of SEC31, via the concentration of PEF1: At low concentrations of PEF1, there may be insufficient co-adaptor protein to recruit SEC31 for mono-ubiquitination. Whereas at too high concentrations, the PEF1 antagonise and outcompete SEC31, reducing the ubiquitination activity. Thus, the levels of mono-ubiquitinated SEC31, and so COPII complex formation, would be co-regulated by the concentration of PEF1.

These ideas raise to immediate three follow-up questions to consider: 1) What regulates the levels of the non-substrate interactors (PEF1, PLEKHA4) in the cell? 2) How do these interactors avoid being ubiquitinated by KLHL12 themselves? While the regulator of these interactors has not been identified, I hypothesise that these interactors escape ubiquitination activity by deliberately

missing a surface lysine residue at the appropriate position for KLHL12 to be able to transfer ubiquitin. If the ubiquitination sites of KLHL12 substrates (DVL1-3) were mapped, these sites could be confirmed through structural data or modelling. 3) Are there any other E3 ligases that are also regulated in a similar manner? Most E3 ligases can ubiquitinate a variety of targets, for a range of different purposes, KLHL12 is a rare example where a common recognition motif is shared between interactors implicated in different roles (substrate, co-factor, antagonist). In other E3 ligase systems, how similar would a shared recognition motif be, and what range of processes would these systems be involved in?

Chapter 6: Conclusions and Future Directions

6.0 Conclusions and Future Directions

I was drawn to study E3 ligases for my DPhil studies because of their key relevance in targeted protein degradation (TPD) modalities.[27] Even while just focusing specifically on cullin-RING ligases, I now appreciate just how varied these multimeric E3 ligases can be. Members of the KCTD family form a range of oligomeric states (dimers, tetramers, and pentamers) and have wildly variable substrate recognition domains. In contrast, kelch family members (including KLHL12), contain a more conserved substrate binding domain, where many of their substrate recognition interactions are already well understood. Furthermore, F-box E3 ligases require a separate SKP1 substrate adaptor to recruit CUL1, and they do not form any high-order oligomeric states.[169] Overall, the proteins I investigated represent a slice of the diversity in E3 ligases systems. As more detail into the structural mechanisms of these ligases are uncovered, there will be increased potential to utilise them in TPD modalities.

Many of the KCTD family were first not well characterised and at the start of my DPhil studies there was limited structural data for KCTD9.[55] I utilised a high-throughput cloning method to generate several KCTD9 constructs for expression in *Sf9* insect cells. I identified that fusion of a C-terminal FLAG-tag to KCTD9 improved the expression yield and stability of the protein. Samples were used to generate two high resolution cryo-EM datasets: one of KCTD9A-c047 alone, and the other of KCTD9A-c047 in complex with CUL3A-c012. These datasets resolved the proposed substrate recognition PPR domain, which adopts a unique β-solenoid structure that has been likened to B-form DNA. While this fold is commonly observed in cyanobacteria, this is the known example within *Homo sapiens*.[66] I submitted the PPR domain (T214-E376) from my structure model to

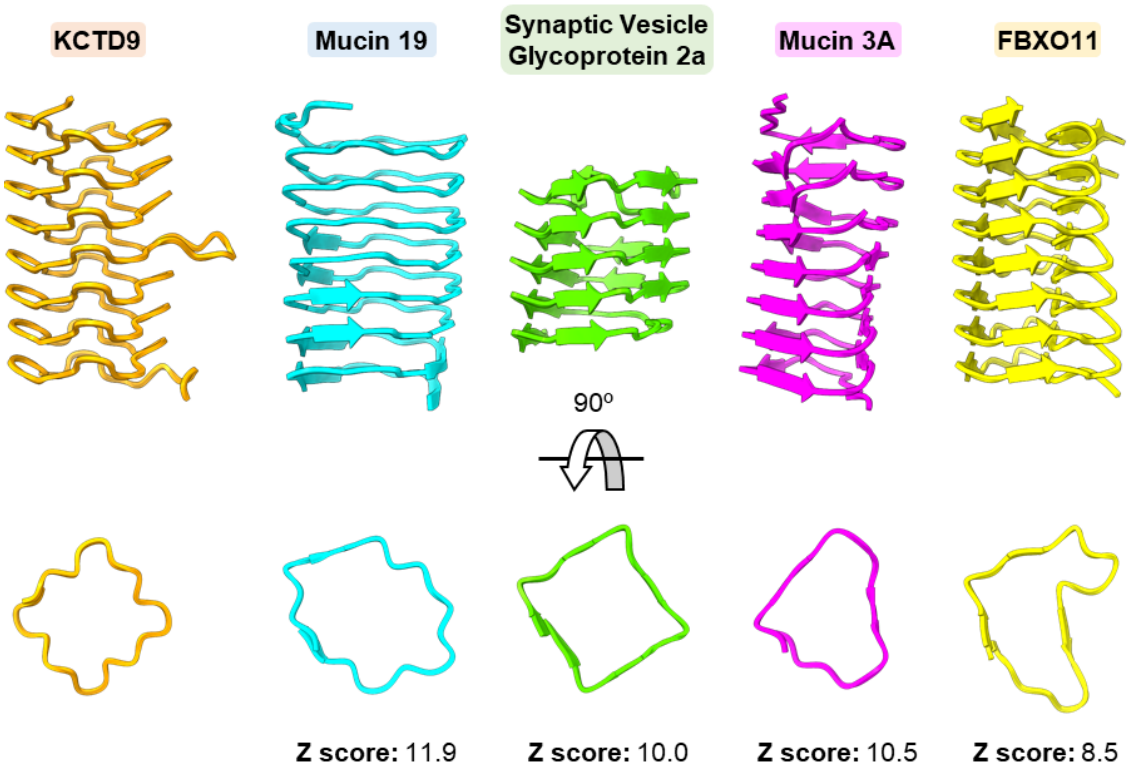


Figure 6.1: Comparison of models, within the human Alphafold V2 database, that were identified by the DALI server to be most similar to KCTD9 PPR domain (T214-E376). Structures are limited to the KCTD9 PPR region and are overlaid using the untransformed coordinates from DALI analysis. A single representative coil of each model and Z score (similarity score) are shown beneath.

the DALI server for comparison against the human Alphafold V2 database.[47,49] While the models with the highest similarity had solenoid-like folds, their structures did not resemble the four-sided coil that KCTD9 displayed (Figure 6.1).

Unexpectedly, KCTD9 was observed in a homo-decameric state, with interactions mediated through the C-termini of the PPR domains. No other member of the KCTD family has been observed to assemble to such a high-order state, and it is unclear if this representative of a biologically relevant storage state, or if it is structural artifact. Considering that the PPR is the proposed substrate recognition domain, how this observed state relates to interactor recognition requires further investigation.

Overall, this project required efforts to generate new samples of KCTD9 and identify suitable conditions for cryo-EM. These results can be used to aid future structural studies of this E3 ligase, especially when in complex with interaction partners that will inevitably be identified.

Several studies have implicated KCTD9 in NK cell development and effector function.[63,71,170] Yet despite the many investigations, no interactor of KCTD9 for this role has been identified. I generated affinity MS proteomics datasets, first in HEK293 and the in NK92mi cells, to identify potential interactors of this E3 ligase and elucidate its potential role. This data is still currently under analysis, whereby KCTD9/interactor complex models are being predicted using Alphafold3.[107] Whether this will aid in shortlisting plausible KCTD9 interactors remains to be seen.

After validating a bona fide interactor, through co-IP and immunoblotting experiments, more exciting investigations can commence. Cryo-EM studies of KCTD9/interactor complexes will provide insights into the structural mechanisms that govern interactor recognition. Additionally, ubiquitination assays will determine the type of cellular event mediated by KCTD9; whether it functions as a degrader, or if ubiquitination triggers an alternative process such as signalling. Most interestingly, if KCTD9's role in NK cell development is understood, it could prompt investigation into potential targets for immunotherapy.[88,171]

DAVID analysis of the NK92mi proteomics dataset identified a group of genes that were DNA or RNA-binding proteins.[146,151] Given that the PPR domain of bacterial MpfA mimics DNA for DNA-gyrase binding, potentially the PPR of KCTD9 may interact with DNA/RNA binding proteins such as ZNF638.[68] This zinc

finger protein, that is also localised to the nucleoplasm, has been show to regulate innate antiviral immune response.[172] I postulate KCTD9 could regulate ZNF638 levels through K48-linked ubiquitination, and that its degradation could stimulate NK cell development.

Another mystery the remains with KCTD9 is the role of its Ub-like domain, for which I resolved a 2.2 Å crystal structure. This domain is another unique feature of this non-canonical KCTD family member, and its biological relevance is still not clear. The RING E3 ligase UHRF1 also possesses an N-terminal Ub-like domain, which promotes ubiquitination activity by stabilising the E2/E3 complex.[173] It is possible that the Ub-like domain in KCTD9 functions similarly, or perhaps it facilitates the localization of KCTD9 to a specific cellular region. The precise role of this domain will likely be clarified once an interactor identified from the MS proteomics data is validated.

When I was exploring the structural mechanism of KLHL12, I generated a KLHL12/PEF1 peptide co-crystal dataset (PDB: 8RBH), which adopted the same U-turn confirmation seen by the DVL1 peptide (PDB: 6TTK), and PLEKHA4 (PDB: 8OIO). In these structures, a common PGXPP motif allows KLHL12 to recognise multiple interactors by the same binding mechanism.

KLHL12 is a rare example of an E3 ligase that is able to recruit a number of different partners that partake in a number of roles through a shared PGXPP recognition motif. DVL1-3 are typical substrates, whose degradation is mediated by this E3 ligase, while PEF1 serves as a co-adaptor to enable SEC31 monoubiquitination, and PLEKHA4 sequesters KLHL12 to regulate its ubiquitination activity.[82,84,86] The newly generated KLHL12/PEF1 and

KLHL12/PLEKHA4 peptide co-crystal datasets (PDB: 8RBH and 8OIO, respectively) reveal that these interactor peptides adopt the same U-turn conformation previously observed in the KLHL12/DVL1 structure (PDB: 6TTK).[78] Biophysical affinity measurements of different PGXPP-containing peptides identified a general relationship between motif sequence and binding affinity. However, deviations from this trend suggest that additional factors, such as the conformational dynamics involved to adopt the required U-turn binding mode, may influence binding affinity. Molecular dynamic simulations may provide further insights into this substrate recognition system.

Finally, the rise in interest of E3 ligases and ubiquitination biology corresponds with the growth of the TPD research field.[174] Characterising E3 ligase systems is critical for the advancement of this field, as it enables the discovery of novel E3 ligase binders to allow the selection of ligases that are best suited for degrading the given POI. However, these characterisation studies are often challenging, and often require a range of biological research expertise. These studies need to be thorough and high-quality; else, mischaracterisation of substrate binding interactions can misdirect a project, ultimately leading to wasted research efforts and resources, as demonstrated in my investigation of FBXO31 (see appendix chapter).

Characterisation studies must consider all the potential interaction partners of an E3 ligase (KLHL12, FBXO31, and most likely KCTD9, all have multiple interactors), and the range of diverse cellular processes they are implicated in. This aspect is largely overlooked in the TPD field when hijacking the UPS, but the consequences of re-localising an E3 ligase, substituting one ubiquitination event for another, or altering the balance of free ubiquitin within the cell could

have severe effects on cell homeostasis; which may hinder the development of TPD therapeutics. As this field and ubiquitin biology continue to advance, the research presented in this thesis will contribute our deeper understanding of E3 ligase functions.

Appendix: FBXO31 Chemical Probe Development

7.0 Appendix: FBXO31 Chemical Probe Development

7.1 Introduction

This brief appendix chapter details attempts to develop a chemical probe for FBXO31, a SCF (SKP1-CUL1-F-box) cullin-RING E3 ligase. FBXO31 is one of four F-box proteins (the others being SKP2, FBXW8, and FBXO4) that ubiquitinates and degrades cyclin D1 (CD1) in order to regulate G1 phase progression.[175-177] Specifically, FBXO31 is activated following DNA damage, triggering cyclin D1 degradation and resulting in a G1 arrest, allowing the cell time to recover before proceeding with the cell cycle.[178] Developing a chemical probe not only serves as a tool to further investigate the roles of FBXO31, it can also serve as a starting compound for the development of an novel E3 ligase binder that could be use in TPD (targeted protein degradation) modalities.

FBXO31 was selected as a target to develop such a tool because there was a 2.7Å FBXO31/SKP1/CD1 peptide co-crystal structure that had been reported by B. Hao *et al.* (Figure 7.1, PDB: 5VZU).[91] In this structure, the final four residues of the CD1 peptide (292-295) are deep within the second β-barrel region of the F-box protein. The C-terminal carboxylate group of CD1 participates in key hydrogen bonding interactions with K330 and D334, and water-mediated interactions with T343 and L472.

Peptide	Sequence	K_d	GRAVY Score
Cyclin D1	279- EEVDLAC**TPTDV**RD**V**D**I** -295	2.7 ± 0.4 μM	-0.194
Cyclin D2	273- DELDQAS**TPTDV**RD**I**D**L** -289	3.0 ± 0.5 μM	-0.865
Cyclin D3	276- QGPSQTS**TPTDV**TA**I**H**L** -292	1.9 ± 0.5 μM	-0.435

Table 7.1: 17-mer Cyclin D1-3 peptides, their reported binding affinities to FBXO31 (K_d), and calculated GRAVY (Grand average of hydropathicity) scores (Protparam).[91]

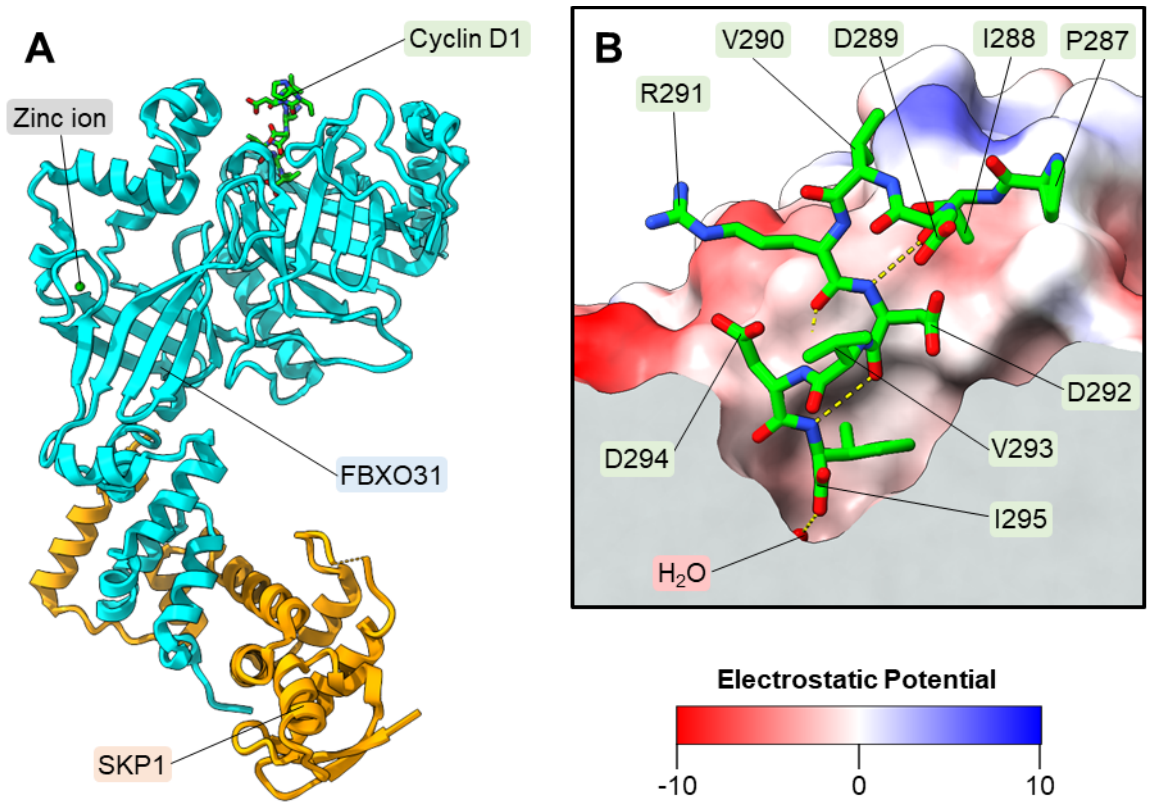


Figure 7.1: A) FBXO31/SKP1/CD1 peptide complex structure (PDB: 5VZU). FBXO31 (cyan), SKP1 (orange), Cyclin D1 peptide (lime).[91] B) CD1 peptide within the β-barrel pocket of FBXO31. FBXO31 surface coloured by electrostatic potential.

This structural dataset can be directly used in virtual ligand screening (VLS) of compounds to generate a refined list of compounds to start biophysical assays; potentially avoiding huge resources that experimental screening of a large ligand panel would require. Additionally, if a reproduceable, high-resolution crystal ligand-crystal soaking system was established, it would enable structure guided optimisation of the chemical probe. Furthermore, the researchers quantified the binding affinities of cyclin D1-3 peptides to FBXO31 by ITC (Table 7.1), raising the possibility of developing a peptide-displacement assay in order to validate chemical binders and quantify their binding affinities to FBXO31.

7.2 Expression and Purification of the FBXO31/SKP1 Complex

I designed new FBXO31 and SKP1 constructs for expression in bacterial cells using codon-optimised entry clones (Table 7.1). Many F-box proteins aggregate when expressed on their own but they can be stabilised when co-expressed with the substrate adaptor SKP1 (Figure 7.2A). Therefore, constructs SKP1AB-c003 and SKP1AB-c005 were inserted into the pCDF-LIC DUET vector. This vector has different replicons and drug resistance markers to that of the pNIC28-Bsa4 vector that FBXO31 is cloned into and thus allows both vectors to co-transformed into *E. coli* for the co-expression of the FBXO31/SKP1 complex.[179]

Construct ID (Scarab)	Construct Boundaries	N-terminal Tag	C-terminal Tag	Test Purification Yield
FBXO31A-c001	A76-S559	His-tag	-	*Medium*
FBXO31A-c002	P82-S559	His-tag	-	*Medium*
FBXO31A-c003	A76-S559	StrepII-tag	-	*Medium*
FBXO31A-c004	P82-S559	StrepII-tag	-	*Medium*
FBXO31A-c005	A76-S559	His-tag	Avi-tag	*Low*
FBXO31A-c006	P82-S559	His-tag	Avi-tag	*Low*
SKP1AB-c003*	M21-K169	His-tag	-	High
SKP1AB-c004	M21-K169	His-tag	-	High
SKP1AB-c005*	M21-K169	-	-	High
SKP1AB-c008	M21-K169	-	His/FLAG-tag	Not Tested
XX02FBXO31A-c001	*Bicistronic*	StrepII	-	Low

*SKP1 construct cloned into Duet vector.

Table 7.2: FBXO31 and SKP1 *E. coli* constructs summary table. Italics: Yield when co-expressed with SKP1.

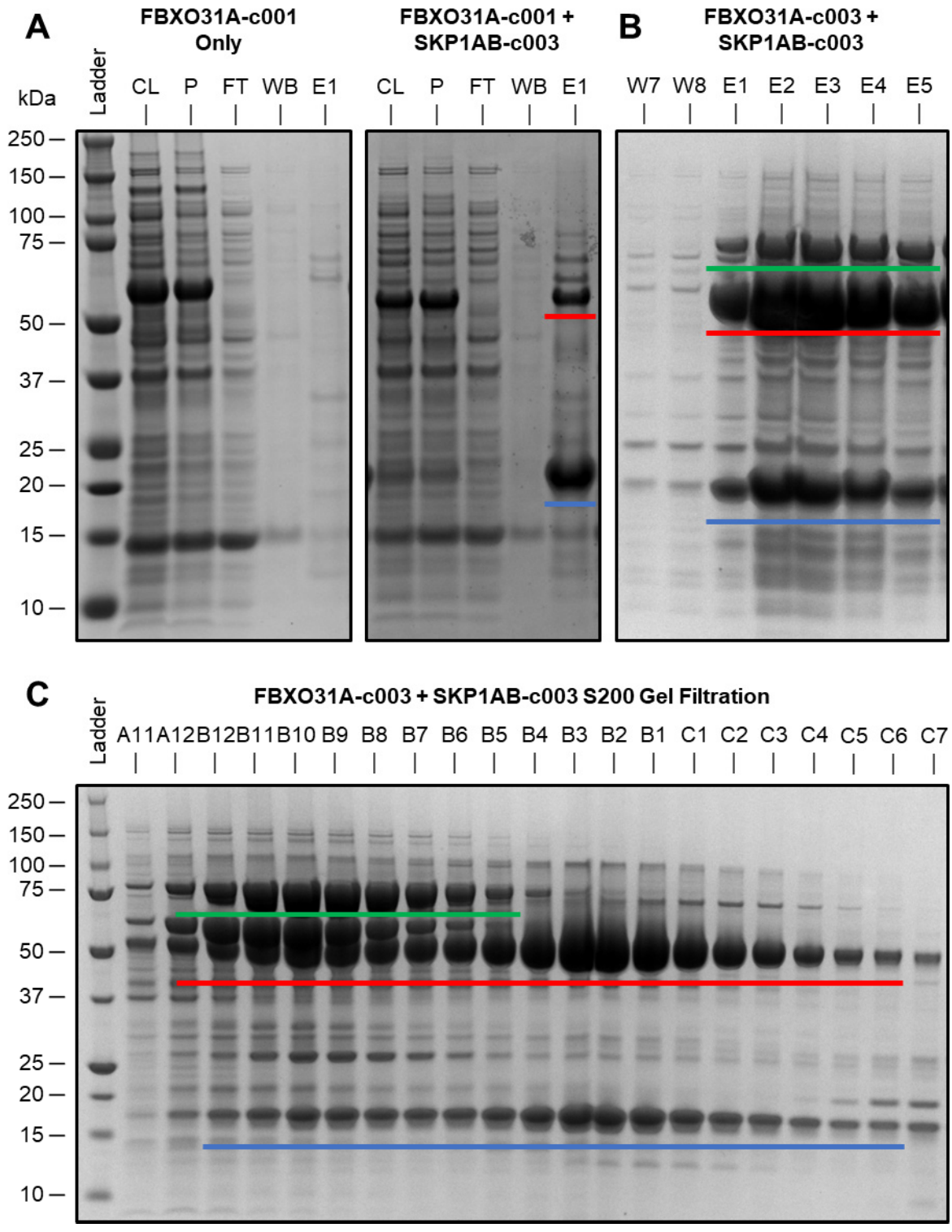


Figure 7.2: SDS-PAGE of FBXO31/SKP1 purification fractions A) Ni-Sepharose test purification of FBXO31A-c001, with and without co-expression of SKP1AB-c003. CL: cell lysate, P: pellet, FT: flow-through, WB: wash buffer 30, E1: Elution buffer (250 mM imidazole). B) Large-scale StrepTactin-XT capture of FBXO31A-c003/SKP1AB-c003. Final washes and initial elution fractions shown. C) S200 gel filtration elution fractions (0.5 mL fraction size). Soluble FBXO31 bands are underlined in red, SKP1 in blue, and HSP contaminant in green.

While the SKP1 component was high yielding, FBXO31 constructs were less so. Additionally, upon larger scale expression, a large heat-shock protein (HSP) that had affinity to the FBXO31 component was a common containment. Potentially this HSP is either DnaK or GrpE, two common bacterial HSPs, but the of the identity of contaminant was not confirmed.[180] Neither ATP-$MgCl_2$ washes, anion exchange steps, nor gel filtration was able to remove the HSP contaminant without sacrificing a large fraction of the yield (Figure 7.2B,C).

After these initial purification results and the crystallisation attempts which had not yielded any high-resolution data, I tried expressing the FBXO31/SKP1 complex using a bicisitronic vector (XX02FBXO31A-c001, StrepII-FBXO31/SKP1). This single vector encodes both FBXO31 and SKP1, but they are separated by a second ribosomal binding site.[181] The amino acid sequence of this construct exactly matched the construct used by B. Hao *et al* for their studies, and I thought this could have better purification/crystallisation results. However, upon large scale purification, I found that the HSP protein was still present and while the relative levels of FBXO31 and SKP1 were more balanced, the overall yield was lower.

I postulated that the presence of HSP, and its apparent affinity to FBXO31, might be a result of misfolded protein that was being generated during bacterial expression. FBXO31 contains a CHCH-type zinc-binding region, and it was reported that the role of the zinc ion is required to facilitate proper folding of the protein. Furthermore, B. Hao *et al.* found that FBXO31 constructs with mutations at this region became insoluble aggregates. Potentially, under the rapid synthesis of protein, there could be a depletion in the local concentration of zinc ions within the cell; although this reported to be difficult to achieve.[182] Nevertheless, I

conducted test expression and purifications of FBXO31A-c003 + SKP1AB-c003 with and without the addition of $ZnCl_2$ (final concentration of 50 µM) to the expression cultures. In both samples, the relative levels of FBXO31/SKP1 and the HSP containment appeared to be the same, and this alteration did not have the desired effect.

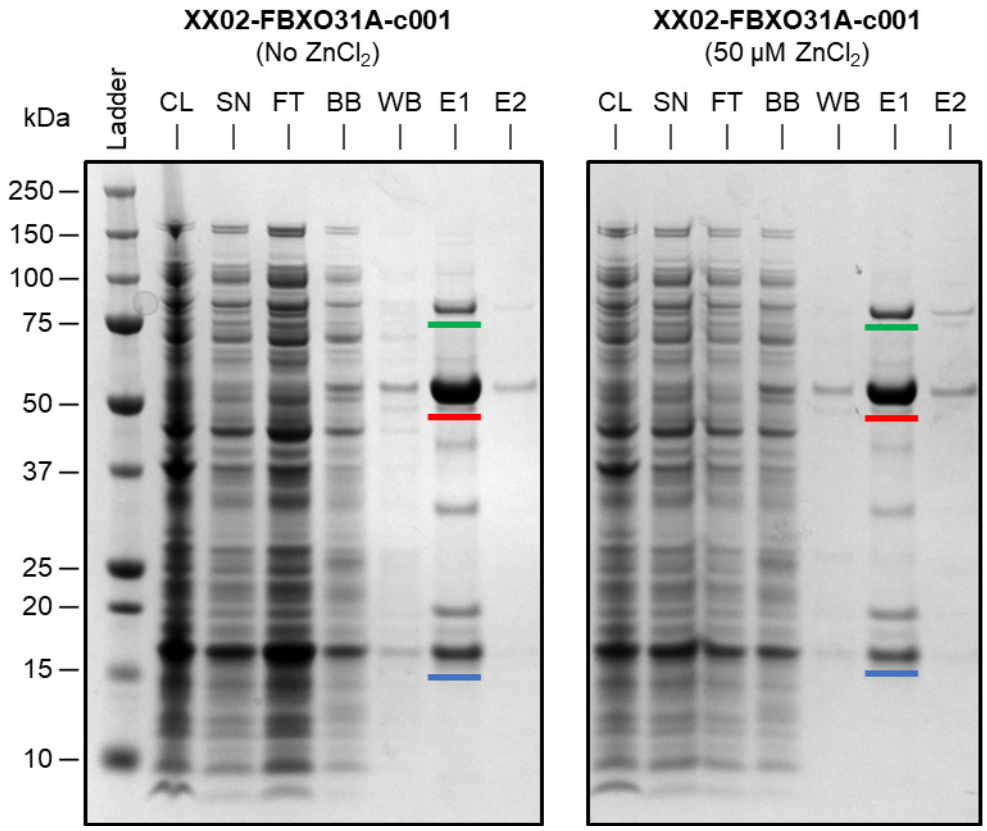


Figure 7.3: Test expression of XX02-FBXO31A-c001 in cultures treated with and without $ZnCl_2$. 1L test expression cultures.

7.3 Crystallisation of the FBXO31/SKP1 Complex

Despite the challenges discussed, purified samples of FBXO31/SKP1 were obtained by capturing using StrepTactinXT resin, and then further purification by S200 gel filtration. These samples were concentrated to 10 mg/mL and used applied to many crystallisation screens (Table 7.3).

Construct ID (Scarab)	Protein Concentration (mg/mL)	Crystal Screen	Mounted Crystals	Best Diffraction Result
FBXO31A-c003 + SKP1AB-c003	10	JCSG, 20 ºC	None	-
		JCSG, 4 ºC	2	No Diffraction
		MIDAS, 20 ºC	None	-
		MIDAS, 4 ºC	None	-
		LFS, 20 ºC	None	-
		LFS, 4 ºC	None	-
		HIN3, 20 ºC	None	-
		HIN3, 4 ºC	None	-
		HCS, 20 ºC	None	-
		HCS, 4 ºC	None	-
		JCSG H4d, 4 ºC	16	4.3 Å, large unit cell
		Literature	None	-
		Additive Screen	9	4.8 Å, large unit cell
		Silver Bullet	2	-
	5	*JCSG H4d, 4 ºC*	None	-
FBXO31A-c004 + SKP1AB-c003	10	Morpheus, 4 ºC	25	5.5 Å, large unit cell
		HIN3, 4 ºC	None	-
		HCS, 4 ºC	6	No Diffraction
		MIDAS, 4 ºC	None	-
		LFS, 4 ºC	None	-
		JCSG, 4 ºC	1	No Diffraction
		JCSG H4d, 4 ºC	16	2.7 Å, low completeness
		Literature	None	-
XX02FBXO31A-c001	10	JCSG, 20 ºC	None	-
		JCSG, 4 ºC	16	No Diffraction
		LFS, 20 ºC	None	-
		LFS, 4 ºC	None	-
		HIN3, 20 ºC	None	-
		HIN3, 4 ºC	None	-
		HCS, 20 ºC	None	-
		HCS, 4 ºC	None	-

LFS: Ligand-Friendly Screen HIN3: Hampton Research Index Screen
HCS: Hampton Crystal Screen
JCSG: Joint Centre for Structural Genomics Screen
Literature: Fine screen designed from literature condition (PDB: 5VZU).

Table 7.3: FBXO31 crystallisation summary table. Best dataset is described in diffraction result column. Fine screens are in italics.

Of the initial coarse crystal screens made using FBXO31A-c003/SKP1AB-c003, only one well (H4d) of the JCSG screen at 4°C yielded any crystals; although this crystal did not diffract. The crystal conditions of this well had elements that resembled the conditions reported in FBXO31/SKP1 crystal structure (Table 7.4).

Condition Name	Condition
JCSG H4d	45% MPD, 200 mM $CaCl_2$, 100 mM Bis-Tris (pH 5.5), 4 ºC
Literature	13-17% MPD, 20 mM $CaCl_2$, 270 mM ammonium acetate,100 mM Sodium Citrate (pH 5.3), 4 ºC

Table 7.4: JCSG H4d well conditions, and reported crystal condition for the FBX31/SKP1 complex structure (PDB: 5VZU).

Fine screens of both the JCSG H4d hit condition and the literature condition were established using the FBXO31A-c003/SKP1AB-c003. Although the literature fine screen did not yield any crystals, the JCSG H4d fine screen produced crystals in many conditions with varying morphologies, all of which fluoresced upon UV excitation (Figure 7.4). Most of these crystals generated datasets between 6-8 Å and the best resolution was 4.3 Å. Further crystallisation attempts were made using alternative screens and use of FBXO31A-c004/SKP1AB-c003, yet despite the many of crystals obtained from the Morpheus crystal screen, no dataset of sufficient resolution (1-2Å) and completeness for ligand soaking was obtained. XX02FBXO31A-c001 was also applied coarse crystal screens, but very few crystals were obtained and none of these diffracted.

It is worth noting, that in the datasets that did diffract, the unit cell was far larger than the reported FBXO31/SKP1/CD1 dataset (Table 7.5). Molecular replacement using phenix.Phaser-MR, with the FBXO31/SKP1 structure as a search model, was unable to phase the structure.

	Overall	Inner Shell	Outer Shell
Low Resolution Limit	71.47	71.47	4.65
High Resolution Limit	4.3	11.38	4.3
Rmerge (within I+/I-)	0.418	0.082	16.39
Rmerge (all I+ and I-)	0.425	0.084	18.071
Rmeas (within I+/I-)	0.429	0.084	16.809
Rmeas (all I+ & I-)	0.431	0.085	18.305
Rpim (within I+/I-)	0.093	0.018	3.723
Rpim (all I+ & I-)	0.067	0.014	2.902
Rmerge In Top Intensity Bin	0.082	-	-
Total Number of Observations	926676	50392	181786
Total Number Unique	22630	1310	4611
Mean ((I)/sd(I))	6.9	37.9	0.3
Mn (I) Half-Set Correlation $CC_{1/2}$	0.999	0.999	0.446
Completeness	100	99.4	100
Multiplicity	40.9	38.5	39.4
Mean (Chi^2)	1	0.91	0.9
Anomalous Completeness	99.9	99.5	99.8
Anomalous Multiplicity	21.3	21.4	20.2
DelAnom Correlation Between Half-Sets	0.138	0.323	-0.003
Average Unit Cell			
a, b, c (Å)		214.40, 214.40, 214.40	
α, β, λ (°)		90.00, 90.00, 90.00	
Space Group		$P\ 2_1\ 3$	

Table 7.5: Aimless data reduction statistics table for best resolution FBXO31A-c003/SKP1AB-c003.

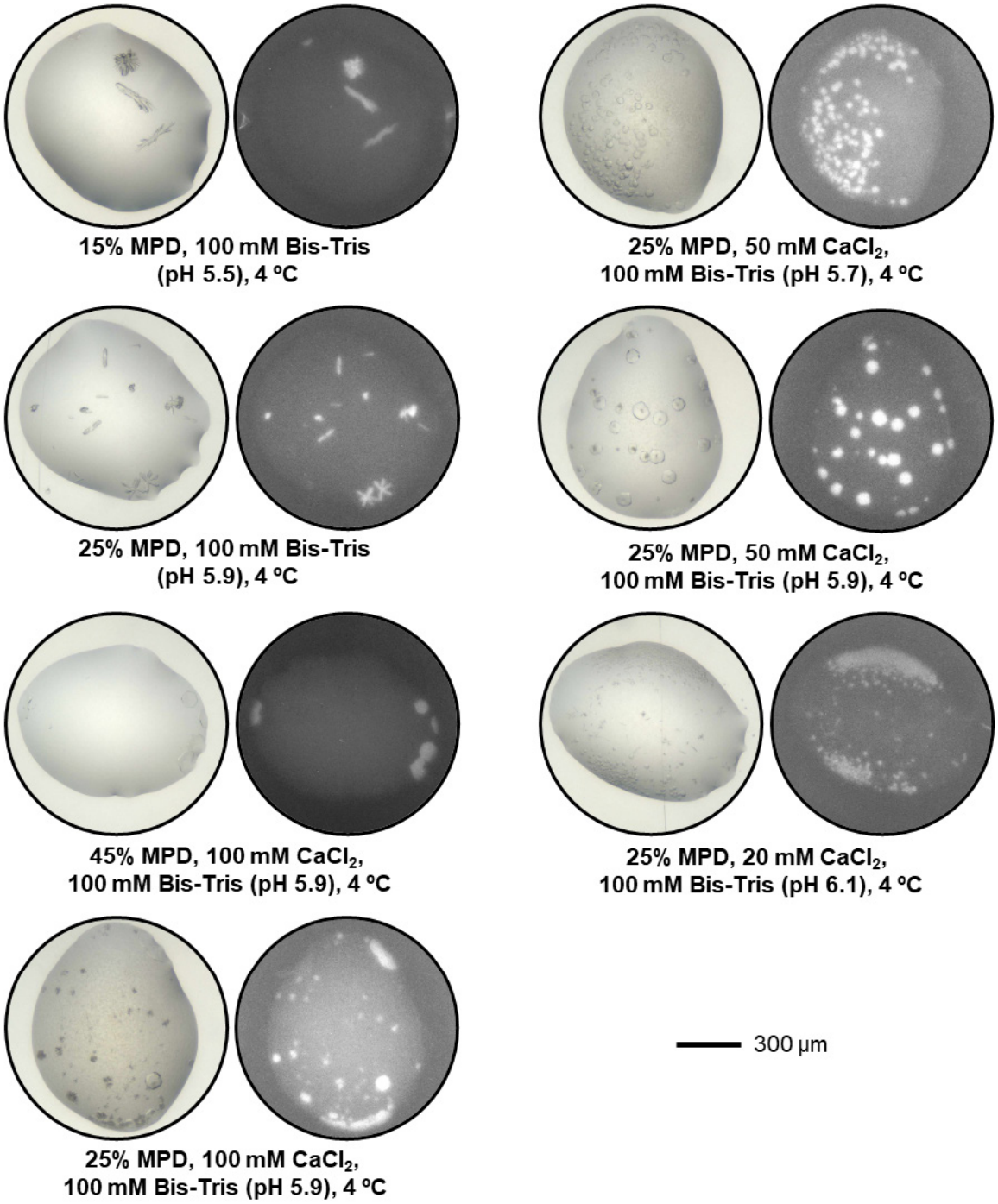
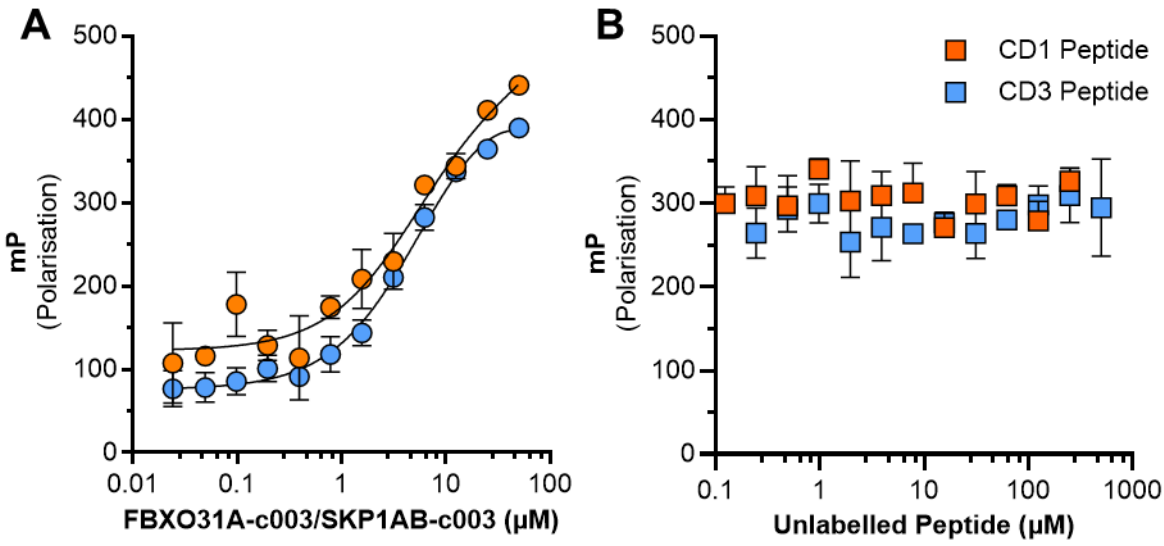


Figure 7.4: A selection of FBXO31A-c003/SKP1AB-c003 crystals that were produced from the JCSG H4d fine screen. Crystals imaged after 7 days using visible light and UV excitation.

7.4 Preliminary Biophysical Assay Trails

I aimed to establish a fluorescent polarisation (FP) peptide displacement assay in order to screen and validate binders for FBXO31.[183] Preliminary tests identified the optimal concentrations of RhodamineB labelled CD1 (RhoB-GG-CD1$^{285-295}$) and CD3 (RhoB-GG-CD3$^{282-292}$) peptides as 400 nM, with FBXO31A-c003/SKP1AB-c003 at approximately 6 μM (Figure 7.5A). The K_d of the labelled peptides to FBXO31 was 5.24 and 6.00 μM (CD1 and CD3 respectively), however no displacement of the labelled peptide by their unlabelled counterparts could be observed (Figure 7.5B,C). Potentially, the RhodamineB label may engage FBXO31 through non-specific hydrophobic interactions, and it is not in competition with the unlabelled peptide at the expected binding site.



C

Peptide	K_d	B_{max}	R^2
Cyclin D1$^{285-295}$	5.24 μM	326.5	0.93
Cyclin D3$^{282-292}$	6.00 μM	396.8	0.99

Figure 7.5: Preliminary FP assays. Ext: 562 nm, Emi 610 nm. 40 μL cell volume. SD error bars. A) Titration of FBXO31/SKP1 complex against RhodamineB labelled cyclin D peptides (400 nM) B) Titration of unlabelled cyclin D peptides against FBXO31/SKP1 complex (6 μM) and cyclin D peptide (400 nM). C) Non-linear fit to data within Figure 7.5A using GraphPad.

Due to the lack of a positive control in this assay, I performed preliminary experiments using complementary biophysical methods, including Grating-Coupled Interferometry (GCI), Surface Plasmon Resonance (SPR), and Homogeneous Time-Resolved Fluorescence (HTRF).[184-187] These assays used either biotinylated FBXO31A-c005 (His-FBXO31$^{76-559}$-Avi) or a biotinylated CD1 peptide (data not shown). Unfortunately, none of these techniques were able to demonstrate a reliable positive control, preventing the screening of compounds against FBXO31.

7.5 Virtual Ligand Screening Against FBXO31

Compounds were virtually screened against FBXO31 using the Glide package within Schrodinger Maestro.[188] This method aims to generate a refined list of compounds that have increased chances of binding to FBXO31; thus, limiting costly large-scale compound screening. The FBXO31/SKP1/CD1 structural dataset aided virtual ligand screening (VLS), as the CD1 peptide could be used define the docking site.

I wanted to identify a small starting compound that formed a strong hydrogen bond with K330, as seen with CD1. Therefore, Enamine's compound library (~2.3 million total ligands, accessed 2020) was filtered to select ligands that had a single carboxylic acid moiety, and a molecular weight between 200-350 g/mol (Figure 7.6A).[189] Using KNIME, compounds which contained a PAINS (Pan-Assay Interference Compounds) and REOS (Rapid Elimination of Swill) motif were also filtered to avoid problematic hit compounds.[190,191] 124,436 ligands passed these filters and used in standard precision (SP) docking. The top 10,000 compounds, based on Glide docking scores, were then subjected to extra

precision (XP) docking. The similarity in the predicted binding poses of the ligands and the cyclin D1 peptide (D292-I295) were quantified by calculating their SuCOS (Susan Combined Overlap Score).[192] The compounds were sorted in 555 clusters based on Tanimoto similarity clustering (performed in ICM).[193] After manual inspection of the predicted binding poses, 95 compounds which displayed strong interactions with K330 (like the cyclin D1 peptide) were selected for purchasing and biophysical validation (Figure 7.7, Table 7.6). As the Glide docking score and SuCOS scores did not show clear correlation, both compounds with either high docking scores or high SuCOS were selected (Figure 7.6C). These compounds are yet to be screened against FBXO31 because a biophysical assay was not established.

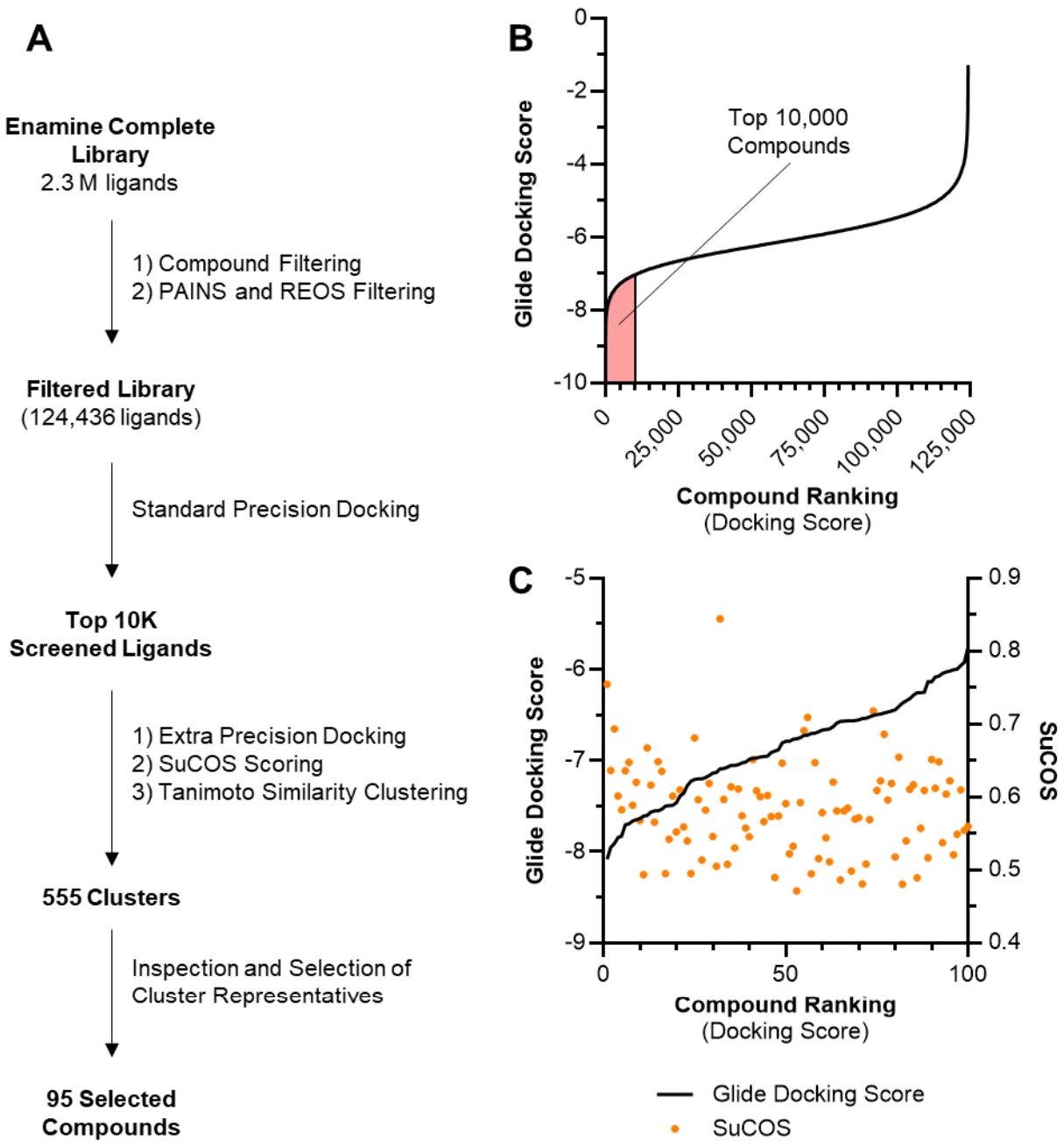


Figure 7.6: A) FBXO31 VLS pipeline. B) Standard precision docking results. Compounds sorted by Glide docking score.[188] Top 10,000 compounds are highlighted in red. C) Selected compounds for biophysical assays, sorted by Glide docking score (black line), calculated SuCOS (orange circles).[192]

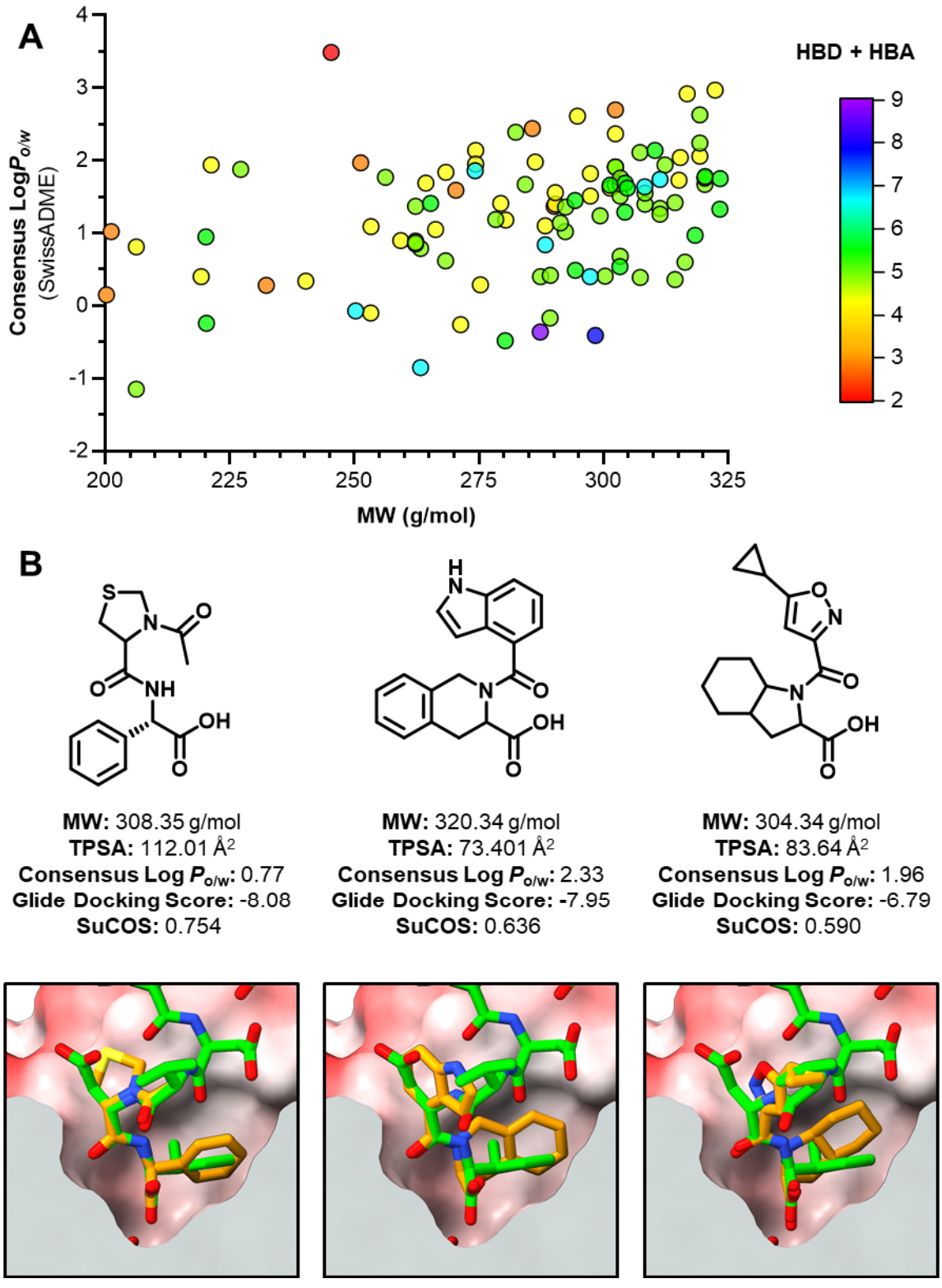


Figure 7.7: A) MW vs consensus Log$P_{o./w}$ (SwissADME) for the 95 selected compounds.[194] Data points coloured by total number of hydrogen bond donors and acceptors. B) Three examples of the compounds selected for biophysical assays. Predicted binding pose (orange) is overlaid with CD1 peptide (green). FBXO31 surface is coloured by electrostatic potential, as previously.

No.	Enamine Catalogue No.	SMILES
1	Z1444321992	CC[C@H](C)[C@H](NC(=O)C1CSCN1C(=O)C)C(=O)O
2	Z1455370062	CC(=O)N1CSCC1C(=O)N[C@H](C(=O)O)C=2C=CC=CC2
3	Z239133844	OC(=O)C1CC2CCCCC2N1C(=O)C3CCCO3
4	Z1455181526	CC(=O)N1CSCC1C(=O)N2CCCC2C(=O)O
5	Z1455368826	CC(=O)N1CCCC1C(=O)N[C@H](C(=O)O)C=2C=CC=CC2
6	Z1013490068	OC(=O)C1CC2CCCCC2N1C(=O)C3CCCOC3
7	Z1443636729	CSC(C)C(=O)N1C(CC2CCCCC21)C(=O)O
8	Z1805575259	OC(=O)C1C2CCCC2CN1C(=O)C=3C=CN=CC3
9	Z1603334553	CC1=CC(C)=C(N1)C(=O)N2CC=3C=CC=CC3CC2C(=O)O
10	Z989049874	CC=1C=C(Br)OC1C(=O)N2CCCCC2C(=O)O
11	Z398843540	CN1C(Cl)=C(Cl)C=C1C(=O)N2CCCC2C(=O)O
12	Z2504201593	OC(=O)C(CC=1C=CN=CC1)NC(=O)C2CCC(F)(F)CC2
13	Z398909264	CC1=NC(C)=C(S1)C(=O)N2CC=3C=CC=CC3CC2C(=O)O
14	Z1258924920	CC1=NC(=CS1)C(=O)N2CCCCC2C(=O)O
15	Z1869646185	CC1=NSC=C1C(=O)N2CC=3C=CC=CC3CC2C(=O)O
16	Z1869527330	OC(=O)C1CCCCN1C(=O)C=2C=CN=C(OC3CCC3)C2
17	Z1079452422	OC(=O)C1CC2CCCCC2N1C(=O)C3=CSC=N3
18	Z448247582	CC12CSCC(=O)N2C(CS1)C(=O)O
19	Z415653262	OC(=O)C1CC=2C=CC=CC2CN1C(=O)C=3C=CC=C4NC=CC34
20	Z1455298195	CC(=O)N1CCCC1C(=O)N[C@@H](CC=2C=CC=CC2)C(=O)O
21	Z1013490840	OC(=O)C1CC=2C=CC=CC2CN1C(=O)C3CCCOC3
22	Z1603241006	CCC1=CC(C)=C(O1)C(=O)N2CCCCC2C(=O)O
23	Z1455178552	OC(=O)C1CCCN1C(=O)CCN2CCCC2=O
24	Z1033090390	COC(C(=O)NCC(C(=O)O)C=1C=CC=CC1)C=2C=CC=CC2
25	Z398910472	COC=1N=CC=CC1C(=O)N2CC=3C=CC=CC3CC2C(=O)O
26	Z398860328	COC=1N=CC=CC1C(=O)N2CCCCC2C(=O)O
27	Z1991713029	OC(=O)C(NC(=O)C=1C=CC=C2CCCC21)C3CCCOC3
28	Z2522368619	CC(=O)NC1(CCC(F)(F)CC1)C(=O)O
29	Z1033090292	OC(=O)C(CNC(=O)C=1C=CC=C2NC=CC12)C=3C=CC=CC3
30	Z2296018967	CN1C=CC=C1C(=O)NC(C(=O)O)C=2C=CC=3OCCC3C2
31	Z398861370	OC(=O)C1CCCCN1C(=O)C2=CN=C3C=CC=CN3C2=O
32	Z98652908	OC(=O)[C@H](CC=1C=CC=CC1)NC(=O)C2CCCO2
33	Z2256342258	COCC1=CC=C(O1)C(=O)N2CC3CCCC3C2C(=O)O
34	Z85881679	CC1=NOC(C)=C1CC(=O)NC(CC=2C=CC=CC2)C(=O)O
35	Z756429966	OC(=O)[C@@H]1CCCN1C(=O)C=2C=CC=CN2
36	Z1455185271	CC(NS(=O)(=O)C)C(=O)N1CCCC1C(=O)O
37	Z2256304509	CN(C(CC=1C=CC=CC1)C(=O)O)C(=O)C2CCCCCC2
38	Z2256341359	OC(=O)C1C2CCCC2CN1C(=O)C=3C=CC=C(CC#N)C3
39	Z1385252259	OC(=O)C1CCCCCC1NC(=O)C=2C=CC=CC2Cl
40	Z1444324930	CC[C@H](C)[C@H](NC(=O)C(C)OC=1C=CC=C(C1)C(=OC)C(=O)O
41	Z285197140	OC(=O)C1CC=2C=CC=CC2CN1C(=O)C=3NN=C4C=CC=CC34
42	Z398910720	CN1N=C(CCC1=O)C(=O)N2CC=3C=CC=CC3CC2C(=O)O
43	Z1455366610	CC1=CSC=C1C(=O)N[C@H](C(=O)O)C=2C=CC=CC2

44	Z316096034	CC=1C=CC=C(C(=O)N2CCCCC2C(=O)O)C1O
45	Z1805574931	COCC=1C=CC=C(C1)C(=O)N2CC3CCCC3C2C(=O)O
46	Z1455178130	OC(=O)C1CCCN1C(=O)C=2C=CC=C3OCOC32
47	Z1033088412	OC(=O)C(CNC(=O)C1=CC=CS1)C=2C=CC=CC2
48	Z398900354	OC(=O)C1CC2CCCCC2N1C(=O)C=3C=C(ON3)C4CC4
49	Z56944537	CC12CCC(=O)N2C(CS1)C(=O)O
50	Z1603732438	CC=1C=C(C(=O)NC(C(=O)O)C=2C=CC(F)=C(F)C2)N(C)N1
51	Z398978388	CC1CCC(CC1)(NC(=O)C2=NN(C)C(=O)CC2)C(=O)O
52	Z1444339415	CC1=CC=C(C(=O)N2CC=3C=CC=CC3CC2C(=O)O)C(=O)N1
53	Z1455301516	CC(C(=O)N[C@@H](CC=1C=CC=CC1)C(=O)O)C2=CC=CS2
54	Z169870676	CC=1C(Cl)=CC=CC1NC(=O)CC(C(=O)O)C=2C=CC=CC2
55	Z1031191022	CC=1C=CC(NC(=O)CC2(CC=3C=CC=CC3C2)C(=O)O)=C(C)C1
56	Z1444358830	CC1=CC(C)=C(O1)C(=O)N(CC(=O)O)CC=2C=CC=CC2
57	Z1385252595	CC=1C=C(ON1)C(=O)NC2CCCCCC2C(=O)O
58	Z756435668	OC(=O)[C@H](CC=1C=CC=CC1)NC(=O)CC(F)(F)F
59	Z1444366463	OC(=O)C(NC(=O)CN1N=CC=2C=CC=CC12)C=3C=CC=CC3
60	Z1033088462	CC(=O)NCC(C(=O)O)C=1C=CC=CC1
61	Z735912646	OC(=O)C1(CC(=O)NC=2C=CC=C3OCCOC23)CCCC1
62	Z1021439034	OC(=O)[C@H](CC=1C=CC=CC1)NC(=O)C=2C=CC=3N=CC=CC3C2
63	Z2504209369	CC1=NOC(C)=C1C(=O)NC(CC=2C=CN=CC2)C(=O)O
64	Z1603655621	CC(C(=O)NC(CC=1C=CC(F)=CC1)C(=O)O)N2C=CC=N2
65	Z119990890	CC(=O)N1C(CSC1C2=CC=CO2)C(=O)O
66	Z1455372678	OC(=O)[C@@H](NC(=O)C1=NNC=2C=CC=CC12)C=3C=CC=CC3
67	Z239132268	OC1CC(N(C1)C(=O)C=2C=CC=CC2O)C(=O)O
68	Z119989906	CS(=O)(=O)N1CCC=2C=CC=CC2C1CC(=O)O
69	Z2256304506	CN(C(CC=1C=CC=CC1)C(=O)O)C(=O)C2CCCOC2
70	Z1444368863	OC(=O)C(NC(=O)C=1C=CC=C(C1)N2C=CC=N2)C=3C=CC=CC3
71	Z398841896	OC(=O)C1CCCN1C(=O)CN2N=C3C=CC=CN3C2=O
72	Z1444823741	CCN1C=C(C=N1)C(=O)NC2CCCCCC2C(=O)O
73	Z1443802257	CN1C(=O)NC=C(C(=O)N2CCCCC2C(=O)O)C1=O
74	Z1455248035	CC(C)[C@H](NC(=O)C=1C=CC=C(N2CCNC2=O)C1C)C(=O)O
75	Z1603345885	NC=1N=CC=NC1C(=O)NC2(CC=3C=CC=CC3C2)C(=O)O
76	Z169904530	NC(=O)C=1C=CSC1NC(=O)CC2(CCCCCC2)C(=O)O
77	Z1539140153	CC=1C=CC(C(=O)NC2(CC=3C=CC=CC3C2)C(=O)O)=C(O)C1
78	Z444925418	CC1CCC(CC1)(NC(=O)CN2N=C(C)C=C2C)C(=O)O
79	Z133622412	CS(=O)(=O)NC1(CCCCC1)C(=O)O
80	Z1870166716	CN1C=C(C=CC1=O)C(=O)NC2(CCCCCC2)C(=O)O
81	Z756436472	OC(=O)[C@H](CC=1C=CC=CC1)NC(=O)C2=CN3C=CC=CC3=N2
82	Z1413700788	CN(CC(=O)N1CC(C(=O)O)C=2C=CC=CC12)C(=O)C
83	Z285571582	OC(=O)C(C=1C=CC=CC1)N2C=CC=N2
84	Z45564610	CCCS(=O)(=O)NC(CC(=O)O)C=1C=CC(Cl)=CC1
85	Z398859958	OC(=O)C1CCCCN1C(=O)C=2C=CC=3OCC(=O)NC3C2
86	Z316095884	OC(=O)C1CCCN1C(=O)C=2C=CC=3C=CC=CC3C2O
87	Z398803138	CC1=NC(CC(=O)NC(CC=2C=CC=CC2)C(=O)O)=CS1
88	Z1269158080	CC(=O)N1CS(=O)(=O)CC1C(=O)O
89	Z1917509746	OC(=O)C(NC(=O)NC=1C=CC=CC1)C=2C=CC=3C=CC=CC3C2
90	Z1444360091	OC(=O)CN(CC=1C=CC=CC1)C(=O)C2CCCO2

91	Z398860974	CC(OCC=1C=CC=CC1)C(=O)N2CCCCC2C(=O)O
92	Z1455367286	OC(=O)[C@@H](NC(=O)CC1=CN2C=CSC2=N1)C=3C=CC=CC3
93	Z56854683	OC(=O)C(CC=1C=CC(O)=CC1)NC2CCS(=O)(=O)C2
94	Z1455181452	OC(=O)C1CCCN1C(=O)C=2C=CC=C(C2)N3CCCNC3=O
95	Z1020687040	OC(=O)C(CC=1C=CC=CC1)NC(=O)CN2N=NC=3C=CC=CC23

Table 7.6: 95 compounds selected for biophysical assays. Compounds are sorted by SuCOS score (highest to lowest). Enamine catalogue number (2021),[189]

7.5 Discussion

There is currently a concerted effort within the TPD field to develop novel E3 ligase binders. FBXO31 appeared to be a promising E3 ligase target for chemical probe development because of the detailed characterisation of its substrate interactions; including biophysical assays and an FBXO31/SKP1/CD1 co-crystal structure.

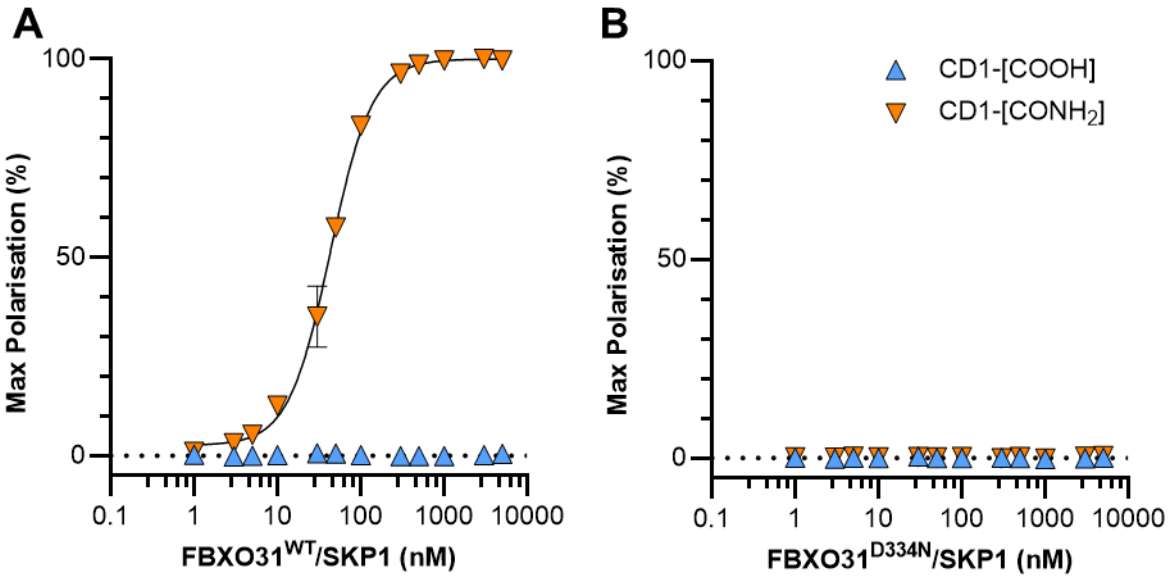
However, reproducing a high-resolution crystal structure proved to be frustrating. Despite producing FBXO31/SKP1 samples and preparing numerous crystallisation experiments, relatively few crystals were obtained, and none of these diffracted to a high enough resolution for crystal-ligand soaking. This showcases how mysterious and challenging protein crystallisation can be, and how published literature can often exclude the painful journeys behind successful structural studies.

Crystallisation attempts may have been hindered by presence of misfolded protein, which was indicated by the large presence of HSP during expression and purification, indicating misfolded or aggregated protein was being generated. For future studies, I recommend exploring expression of the FBXO31/SKP1 complex in *Sf9* insect cells (similar to successful work with KCTD9) or even mammalian cells.

Additionally, rather than using a codon-optimised entry clone (which substitutes each human codon with the most abundant codon within the expression cell), perhaps a codon-harmonised clone would yield better quality FBXO31/SKP1 samples.[195] Codon harmonisation maintains equivalent abundances of codons across cell types, thus if the human DNA gene contains a rarely used codon, a low abundance codon selected for use in the expression cell.[196] This slows down peptide synthesis, which may be necessary in allowing time for the correct folding of the protein.

I attempted to establish several biophysical assays in order to screen and validate compounds binding to FBXO31. Despite several preliminary experiments using numerous different techniques, I was unable to demonstrate the binding of either the CD1 or CD3 peptides to FBXO31. Without a reliable positive control, any subsequent compound screen would struggle to distinguish true binders from false positives

The trouble in demonstrating peptide binding to FBXO31 can explained by a recent study by M. F. Muhar *et al.* (2025), which identified that FBXO31 recognises and ubiquitinates proteins bearing a C-terminal amide modification.[197] These C-terminal amide-bearing proteins (CTAPs) were selectively engaged with high affinity over unmodified carboxylate protein. The researchers measured high affinity binding of several amide modified peptides, including CD1 (T286-I295), using FP assays. In contrast, an unmodified (carboxylate) fluorescein-labelled CD1 peptide showed no binding to FBXO31 whatsoever (Figure 7.8). Furthermore, the study demonstrated that FBXO31^{D334N}, a point mutation within the second β-barrel pocket, was unable to recognise amide-modified peptides; indicating that substrate recognition occurs within this pocket.



C

Peptide	K_d	B_{max}	R^2
▲ Cyclin D1$^{285\text{-}295}$-[COOH]	n.b.	-	-
▼ Cyclin D1$^{285\text{-}295}$-[$CONH_2$]	43.9 nM	115.6	0.98

Figure 7.8: Replotted FP assay data generated by M. F. Muhar *et al.* SD error bars.[197] A) Titration of FBXO31^{WT}/SKP1 complex against fluorescein-labelled CD1 peptide (2–10 nM) B) Titration of FBXO31^{D334N}/SKP1 complex instead C) Non-linear fit (GraphPad) of data presented in Figure 7.8A.

Although M. F. Muhar *et al.* did not obtain any structural data to support amide recognition by FBXO31, re-examining the mFo-DFc difference map of the FBXO31/SKP1/CD1 dataset (PDB 5VZU) reveals the presence of negative difference density at residues D294 and I295 of the CD1 peptide (Figure 7.9B).[91] This suggests that the peptide was overbuilt in the original model, and while cyclin D1 remains a confirmed substrate of FBXO31, the initial structural characterisation of the substrate interactions appear to be incorrect, or at least incomplete.

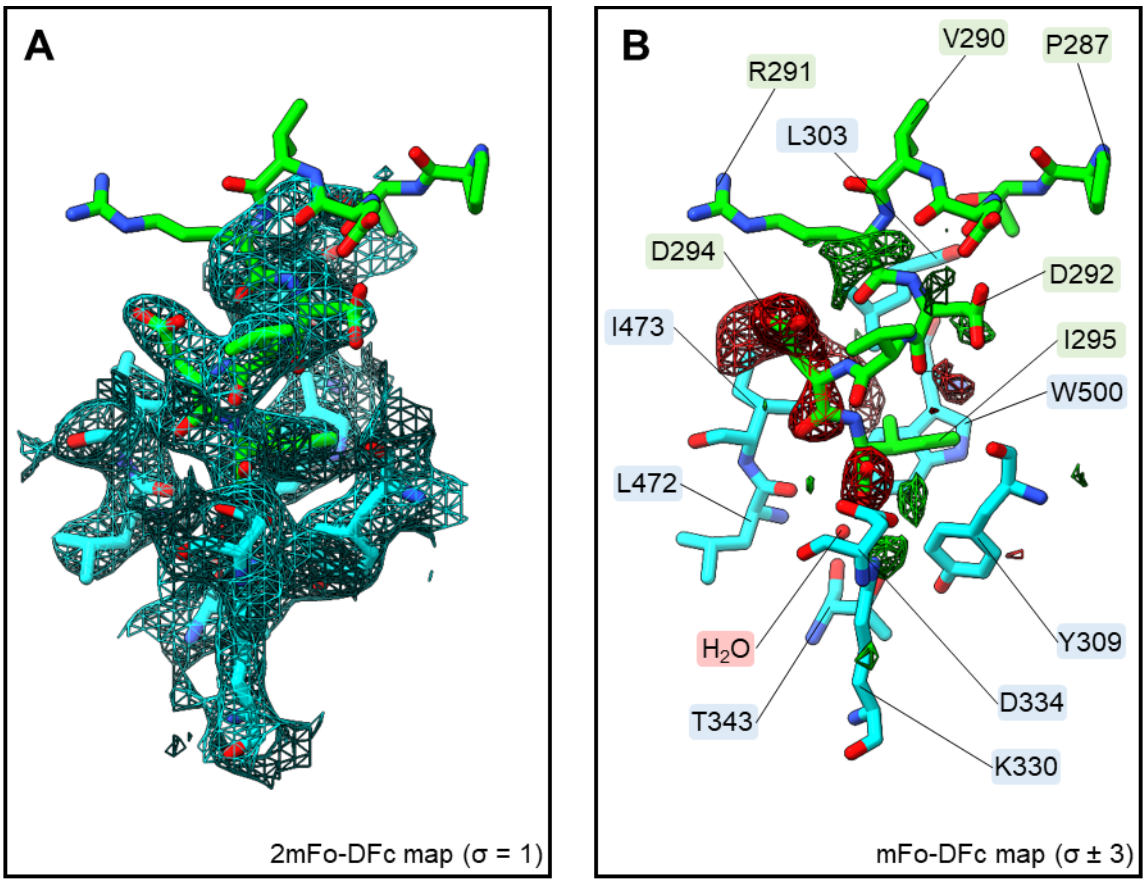


Figure 7.9: A) 2mFo-DFc map around cyclin D1 peptide and interacting FBXO31 residues (σ = 1). Chains B and E (PDB: 5VZU).[91] B) mFo-DFc map (σ ± 3).

These revelations have serious implications for the VLS performed in this project. The FBXO31/SKP1/CD1 dataset was used to define the docking site, and the reported CD1 peptide binding mode was used for compound selection and calculation of their SuCOS. Nevertheless, an interested scientist might still find it worthwhile to screen the selected compounds against FBXO31, particularly the D334N mutant; which is able to recognise unmodified (carboxylate) substrates. Furthermore, the compounds can still be used for screening against proteins that are known to recognise carboxylic acid moieties, such as KLHDC2 or GOPC.[137,158]

References

8.0 References

1. Damgaard, R.B., *The ubiquitin system: from cell signalling to disease biology and new therapeutic opportunities.* Cell Death & Differentiation, 2021. **28**(2): p. 423-426.
2. Dikic, I. and B.A. Schulman, *An expanded lexicon for the ubiquitin code.* Nature Reviews Molecular Cell Biology, 2023. **24**(4): p. 273-287.
3. Komander, D. and M. Rape, *The ubiquitin code.* Annu Rev Biochem, 2012. **81**: p. 203-29.
4. Vijay-Kumar, S., C.E. Bugg, and W.J. Cook, *Structure of ubiquitin refined at 1.8Åresolution.* Journal of Molecular Biology, 1987. **194**(3): p. 531-544.
5. Randles, L. and K.J. Walters, *Ubiquitin and its binding domains.* Front Biosci (Landmark Ed), 2012. **17**(6): p. 2140-57.
6. Schwertman, P., S. Bekker-Jensen, and N. Mailand, *Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers.* Nature Reviews Molecular Cell Biology, 2016. **17**(6): p. 379-394.
7. Clague, M.J. and S. Urbé, *Integration of cellular ubiquitin and membrane traffic systems: focus on deubiquitylases.* FEBS j, 2017. **284**(12): p. 1753-1766.
8. Haglund, K. and I. Dikic, *The role of ubiquitylation in receptor endocytosis and endosomal sorting.* Journal of Cell Science, 2012. **125**(2): p. 265-275.
9. Davis, C., B.L. Spaller, and A. Matouschek, *Mechanisms of substrate recognition by the 26S proteasome.* Current Opinion in Structural Biology, 2021. **67**: p. 161-169.
10. Huang, X., et al., *An atomic structure of the human 26S proteasome.* Nature Structural & Molecular Biology, 2016. **23**(9): p. 778-785.
11. Verma, R., et al., *Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome.* Science, 2002. **298**(5593): p. 611-5.
12. Tai, H.-C. and E.M. Schuman, *Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction.* Nature Reviews Neuroscience, 2008. **9**(11): p. 826-838.
13. Yao, Y., et al., *MDM2: current research status and prospects of tumor treatment.* Cancer Cell International, 2024. **24**(1): p. 170.
14. Yamano, H., *APC/C: current understanding and future perspectives.* F1000Res, 2019. **8**.
15. Kannt, A. and I. Đikić, *Expanding the arsenal of E3 ubiquitin ligases for proximity-induced protein degradation.* Cell Chemical Biology, 2021. **28**(7): p. 1014-1031.
16. Yang, Q., et al., *E3 ubiquitin ligases: styles, structures and functions.* Mol Biomed, 2021. **2**(1): p. 23.
17. Metzger, M.B., et al., *RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination.* Biochim Biophys Acta, 2014. **1843**(1): p. 47-60.
18. Nguyen, H.C., W. Wang, and Y. Xiong, *Cullin-RING E3 Ubiquitin Ligases: Bridges to Destruction.* Subcell Biochem, 2017. **83**: p. 323-347.
19. Qian, H., et al., *Structure and Function of HECT E3 Ubiquitin Ligases and their Role in Oxidative Stress.* J Transl Int Med, 2020. **8**(2): p. 71-79.
20. Wang, X.S., et al., *The unifying catalytic mechanism of the RING-between-RING E3 ubiquitin ligase family.* Nature Communications, 2023. **14**(1): p. 168.
21. Snyder, N.A. and G.M. Silva, *Deubiquitinating enzymes (DUBs): Regulation, homeostasis, and oxidative stress response.* J Biol Chem, 2021. **297**(3): p. 101077.
22. Hu, M., et al., *Crystal Structure of a UBP-Family Deubiquitinating Enzyme in Isolation and in Complex with Ubiquitin Aldehyde.* Cell, 2002. **111**(7): p. 1041-1054.
23. Vu, B., et al., *Discovery of RG7112: A Small-Molecule MDM2 Inhibitor in Clinical Development.* ACS Med Chem Lett, 2013. **4**(5): p. 466-9.
24. Qi, S.M., et al., *Targeting USP7-Mediated Deubiquitination of MDM2/MDMX-p53 Pathway for Cancer Therapy: Are We There Yet?* Front Cell Dev Biol, 2020. **8**: p. 233.
25. Grou, C.P., et al., *The de novo synthesis of ubiquitin: identification of deubiquitinases acting on ubiquitin precursors.* Sci Rep, 2015. **5**: p. 12836.
26. Andreeff, M., et al., *Results of the Phase I Trial of RG7112, a Small-Molecule MDM2 Antagonist in Leukemia.* Clin Cancer Res, 2016. **22**(4): p. 868-76.
27. Tsai, J.M., et al., *Targeted protein degradation: from mechanisms to clinic.* Nature Reviews Molecular Cell Biology, 2024. **25**(9): p. 740-757.
28. Bondeson, D.P., et al., *Catalytic in vivo protein knockdown by small-molecule PROTACs.* Nat Chem Biol, 2015. **11**(8): p. 611-7.

29. Bondeson, D.P., et al., *Lessons in PROTAC Design from Selective Degradation with a Promiscuous Warhead.* Cell Chemical Biology, 2018. **25**(1): p. 78-87.e5.
30. Mares, A., et al., *Extended pharmacodynamic responses observed upon PROTAC-mediated degradation of RIPK2.* Communications Biology, 2020. **3**(1): p. 140.
31. Békés, M., D.R. Langley, and C.M. Crews, *PROTAC targeted protein degraders: the past is prologue.* Nature Reviews Drug Discovery, 2022. **21**(3): p. 181-200.
32. Liu, Y., et al., *Expanding PROTACtable genome universe of E3 ligases.* Nature Communications, 2023. **14**(1): p. 6509.
33. Pike, A., et al., *Optimising proteolysis-targeting chimeras (PROTACs) for oral drug delivery: a drug metabolism and pharmacokinetics perspective.* Drug Discovery Today, 2020. **25**(10): p. 1793-1800.
34. Hornberger, K.R. and E.M.V. Araujo, *Physicochemical Property Determinants of Oral Absorption for PROTAC Protein Degraders.* Journal of Medicinal Chemistry, 2023. **66**(12): p. 8281-8287.
35. Wurz, R.P., et al., *Affinity and cooperativity modulate ternary complex formation to drive targeted protein degradation.* Nat Commun, 2023. **14**(1): p. 4177.
36. Gough, S.M., et al., *Oral Estrogen Receptor PROTAC Vepdegestrant (ARV-471) Is Highly Efficacious as Monotherapy and in Combination with CDK4/6 or PI3K/mTOR Pathway Inhibitors in Preclinical ER+ Breast Cancer Models.* Clin Cancer Res, 2024. **30**(16): p. 3549-3563.
37. Campone, M., et al., *VERITAC-2: A global, randomized phase 3 study of ARV-471, a proteolysis targeting chimera (PROTAC) estrogen receptor (ER) degrader, vs fulvestrant in ER+/human epidermal growth factor receptor 2 (HER2)- advanced breast cancer.* Journal of Clinical Oncology, 2023. **41**(16_suppl): p. TPS1122-TPS1122.
38. Sasso, J.M., et al., *Molecular Glues: The Adhesive Connecting Targeted Protein Degradation to the Clinic.* Biochemistry, 2023. **62**(3): p. 601-623.
39. Jiang, W., et al., *Facilitating the development of molecular glues: Opportunities from serendipity and rational design.* European Journal of Medicinal Chemistry, 2024. **263**: p. 115950.
40. Oleinikovas, V., et al., *From Thalidomide to Rational Molecular Glue Design for Targeted Protein Degradation.* Annu Rev Pharmacol Toxicol, 2024. **64**: p. 291-312.
41. Hagner, P.R., et al., *CC-122, a pleiotropic pathway modifier, mimics an interferon response and has antitumor activity in DLBCL.* Blood, 2015. **126**(6): p. 779-89.
42. Shah Zaib Saleem, R., M.P. Schwalm, and S. Knapp, *Expanding the ligand spaces for E3 ligases for the design of protein degraders.* Bioorganic & Medicinal Chemistry, 2024. **105**: p. 117718.
43. Liu, Z., Y. Xiang, and G. Sun, *The KCTD family of proteins: structure, function, disease relevance.* Cell & Bioscience, 2013. **3**(1): p. 45.
44. Lee, T.E., et al., *Structural determinant for assembly of mammalian K+ channels.* Biophys J, 1994. **66**(3 Pt 1): p. 667-73.
45. Jiang, W., et al., *Structural basis for the ubiquitination of G protein βγ subunits by KCTD5/Cullin3 E3 ligase.* Science Advances, 2023. **9**(28).
46. Nguyen, D.M., et al., *Structure and dynamics of a pentameric KCTD5/CUL3/Gβγ E3 ubiquitin ligase complex.* Proceedings of the National Academy of Sciences, 2024. **121**(17).
47. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
48. Esposito, L., et al., *AlphaFold-Predicted Structures of KCTD Proteins Unravel Previously Undetected Relationships among the Members of the Family.* Biomolecules, 2021. **11**(12): p. 1862.
49. Holm, L., *DALI and the persistence of protein shape.* Protein Science, 2020. **29**(1): p. 128-140.
50. Esposito, L., N. Balasco, and L. Vitagliano, *Alphafold Predictions Provide Insights into the Structural Features of the Functional Oligomers of All Members of the KCTD Family.* International Journal of Molecular Sciences, 2022. **23**(21): p. 13346.
51. Dementieva, I.S., et al., *Pentameric assembly of potassium channel tetramerization domain-containing protein 5.* J Mol Biol, 2009. **387**(1): p. 175-91.
52. Pinkas, D.M., et al., *Structural complexity in the KCTD family of Cullin3-dependent E3 ubiquitin ligases.* Biochem J, 2017. **474**(22): p. 3747-3761.

53. Zuo, H., et al., *Structural basis for auxiliary subunit KCTD16 regulation of the GABA(B) receptor.* Proc Natl Acad Sci U S A, 2019. **116**(17): p. 8370-8379.
54. Miller, K.A., et al., *BTB domain mutations perturbing KCTD15 oligomerisation cause a distinctive frontonasal dysplasia syndrome.* J Med Genet, 2024. **61**(5): p. 490-501.
55. Ji, A.X., et al., *Structural Insights into KCTD Protein Assembly and Cullin3 Recognition.* Journal of Molecular Biology, 2016. **428**(1): p. 92-107.
56. Zheng, S., et al., *Structural basis for KCTD-mediated rapid desensitization of GABA(B) signalling.* Nature, 2019. **567**(7746): p. 127-131.
57. Brockmann, M., et al., *Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling.* Nature, 2017. **546**(7657): p. 307-311.
58. He, H., et al., *Cullin3/KCTD5 induces monoubiquitination of ΔNp63α and impairs its activity.* FEBS Lett, 2018. **592**(13): p. 2334-2340.
59. De Smaele, E., et al., *Identification and Characterization of KCASH2 and KCASH3, 2 Novel Cullin3 Adaptors Suppressing Histone Deacetylase and Hedgehog Activity in Medulloblastoma.* Neoplasia, 2011. **13**(4): p. 374-IN23.
60. Marneros, Alexander G., et al., *Mutations in KCTD1 Cause Scalp-Ear-Nipple Syndrome.* The American Journal of Human Genetics, 2013. **92**(4): p. 621-626.
61. Angrisani, A., et al., *The emerging role of the KCTD proteins in cancer.* Cell Communication and Signaling, 2021. **19**(1): p. 56.
62. Teng, X., et al., *KCTD: A new gene family involved in neurodevelopmental and neuropsychiatric disorders.* CNS Neurosci Ther, 2019. **25**(7): p. 887-902.
63. Zhang, X., et al., *Kctd9 Deficiency Impairs Natural Killer Cell Development and Effector Function.* Frontiers in Immunology, 2019. **10**.
64. Ji, A.X. and G.G. Privé, *Crystal Structure of KLHL3 in Complex with Cullin3.* PLOS ONE, 2013. **8**(4): p. e60445.
65. Vetting, M.W., et al., *Pentapeptide Repeat Proteins.* Biochemistry, 2006. **45**(1): p. 1-10.
66. Zhang, R. and M.A. Kennedy, *Current Understanding of the Structure and Function of Pentapeptide Repeat Proteins.* Biomolecules, 2021. **11**(5): p. 638.
67. Mistry, J., et al., *Pfam: The protein families database in 2021.* Nucleic Acids Research, 2020. **49**(D1): p. D412-D419.
68. Feng, L., et al., *The pentapeptide-repeat protein, MfpA, interacts with mycobacterial DNA gyrase as a DNA T-segment mimic.* Proceedings of the National Academy of Sciences, 2021. **118**(11): p. e2016705118.
69. Hegde, S.S., et al., *A fluoroquinolone resistance protein from Mycobacterium tuberculosis that mimics DNA.* Science, 2005. **308**(5727): p. 1480-3.
70. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.
71. Chen, T., et al., *KCTD9 contributes to liver injury through NK cell activation during hepatitis B virus-induced acute-on-chronic liver failure.* Clinical Immunology, 2013. **146**(3): p. 207-216.
72. Zhang, X., et al., *Interference with KCTD9 inhibits NK cell activation and ameliorates fulminant liver failure in mice.* BMC Immunology, 2018. **19**(1).
73. Yao, H., et al., *KCTD9 inhibits the Wnt/β-catenin pathway by decreasing the level of β-catenin in colorectal cancer.* Cell Death & Disease, 2022. **13**(9).
74. Gupta, V.A. and A.H. Beggs, *Kelch proteins: emerging roles in skeletal muscle development and diseases.* Skeletal Muscle, 2014. **4**(1): p. 11.
75. Stogios, P.J. and G.G. Privé, *The BACK domain in BTB-kelch proteins.* Trends Biochem Sci, 2004. **29**(12): p. 634-7.
76. Canning, P., et al., *Structural Basis for Cul3 Protein Assembly with the BTB-Kelch Family of E3 Ubiquitin Ligases.* Journal of Biological Chemistry, 2013. **288**(11): p. 7803-7814.
77. 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.
78. Chen, Z., et al., *Identification of a PGXPP degron motif in dishevelled and structural basis for its binding to the E3 ligase KLHL12.* Open Biology, 2020. **10**(6): p. 200041.
79. Li, X., et al., *Crystal structure of the Kelch domain of human Keap1.* J Biol Chem, 2004. **279**(52): p. 54750-8.
80. Canning, P., et al., *Structural Basis for Cul3 Protein Assembly with the BTB-Kelch Family of E3 Ubiquitin Ligases*.* Journal of Biological Chemistry, 2013. **288**(11): p. 7803-7814.

81. Angers, S., et al., *The KLHL12–Cullin-3 ubiquitin ligase negatively regulates the Wnt–β-catenin pathway by targeting Dishevelled for degradation.* Nature Cell Biology, 2006. **8**(4): p. 348-357.
82. Gao, C. and Y.G. Chen, *Dishevelled: The hub of Wnt signaling.* Cell Signal, 2010. **22**(5): p. 717-27.
83. 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.
84. Jin, L., et al., *Ubiquitin-dependent regulation of COPII coat size and function.* Nature, 2012. **482**(7386): p. 495-500.
85. Akopian, D., C.A. McGourty, and M. Rapé, *Co-adaptor driven assembly of a CUL3 E3 ligase complex.* Molecular Cell, 2022. **82**(3): p. 585-597.e11.
86. Shami Shah, A., et al., *PLEKHA4/kramer Attenuates Dishevelled Ubiquitination to Modulate Wnt and Planar Cell Polarity Signaling.* Cell Reports, 2019. **27**(7): p. 2157-2170.e8.
87. 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.
88. Waldman, A.D., J.M. Fritz, and M.J. Lenardo, *A guide to cancer immunotherapy: from T cell basic science to clinical practice.* Nature Reviews Immunology, 2020. **20**(11): p. 651-668.
89. Shuken, S.R., *An Introduction to Mass Spectrometry-Based Proteomics.* Journal of Proteome Research, 2023. **22**(7): p. 2151-2171.
90. Buchan, D.W.A. and D.T. Jones, *The PSIPRED Protein Analysis Workbench: 20 years on.* Nucleic Acids Res, 2019. **47**(W1): p. W402-w407.
91. Li, Y., et al., *Structural basis of the phosphorylation-independent recognition of cyclin D1 by the SCF^{FBXO31} ubiquitin ligase.* Proceedings of the National Academy of Sciences, 2018. **115**(2): p. 319-324.
92. Strausberg, R.L., et al., *The mammalian gene collection.* Science, 1999. **286**(5439): p. 455-7.
93. Strain-Damerell, C., et al., *Screening and Production of Recombinant Human Proteins: Ligation-Independent Cloning.* Methods Mol Biol, 2021. **2199**: p. 23-43.
94. Bond, S.R. and C.C. Naus, *RF-Cloning.org: an online tool for the design of restriction-free cloning projects.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W209-13.
95. Jian-Ping Yang, F.K., Lansha Peng, *Novel Vectors for Baculovirus Expression.* Genetic Engineering & Biotechnology News, 2009. **29**(21).
96. Liu, J., et al., *An Improved Method of Preparing High Efficiency Transformation Escherichia coli with Both Plasmids and Larger DNA Fragments.* Indian J Microbiol, 2018. **58**(4): p. 448-456.
97. Laemmli, U.K., *Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-685.
98. Wingfield, P.T., *Overview of the Purification of Recombinant Proteins.* Current Protocols in Protein Science, 2015. **80**(1).
99. Labrou, N.E., *Protein Purification: An Overview*. 2014, Humana Press. p. 3-10.
100. Hong, P., S. Koza, and E.S. Bouvier, *Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates.* J Liq Chromatogr Relat Technol, 2012. **35**(20): p. 2923-2950.
101. Jungbauer, A. and R. Hahn, *Chapter 22 Ion-Exchange Chromatography*. 2009, Elsevier. p. 349-371.
102. Free, R.B., L.A. Hazelwood, and D.R. Sibley, *Identifying Novel Protein-Protein Interactions Using Co-Immunoprecipitation and Mass Spectroscopy.* Current Protocols in Neuroscience, 2009. **46**(1).
103. Yang, P.-C. and T. Mahmood, *Western blot: Technique, theory, and trouble shooting.* North American Journal of Medical Sciences, 2012. **4**(9): p. 429.
104. Demichev, V., et al., *DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput.* Nature Methods, 2020. **17**(1): p. 41-44.
105. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nature Methods, 2016. **13**(9): p. 731-740.
106. Harris, L., et al., *Evaluating Proteomics Imputation Methods with Improved Criteria.* J Proteome Res, 2023. **22**(11): p. 3427-3438.
107. Abramson, J., et al., *Accurate structure prediction of biomolecular interactions with AlphaFold 3.* Nature, 2024. **630**(8016): p. 493-500.

108. Elfmann, C. and J. Stülke, *PAE viewer: a webserver for the interactive visualization of the predicted aligned error for multimer structure predictions and crosslinks.* Nucleic Acids Research, 2023. **51**(W1): p. W404-W410.
109. Mirdita, M., et al., *ColabFold: making protein folding accessible to all.* Nature Methods, 2022. **19**(6): p. 679-682.
110. Cianfrocco, M.A., et al., *COSMIC2: A Science Gateway for Cryo-Electron Microscopy Structure Determination*, in *Practice and Experience in Advanced Research Computing 2017: Sustainability, Success and Impact*. 2017, Association for Computing Machinery: New Orleans, LA, USA. p. Article 22.
111. Maveyraud, L. and L. Mourey, *Protein X-ray Crystallography and Drug Discovery.* Molecules, 2020. **25**(5): p. 1030.
112. Smyth, M.S., *x Ray crystallography.* Molecular Pathology, 2000. **53**(1): p. 8-14.
113. Winter, G., *xia2: an expert system for macromolecular crystallography data reduction.* Journal of Applied Crystallography, 2010. **43**(1): p. 186-190.
114. Winter, G., et al., *DIALS: implementation and evaluation of a new integration package.* Acta Crystallographica Section D, 2018. **74**(2): p. 85-97.
115. McCoy, A.J., et al., *Phaser crystallographic software.* Journal of Applied Crystallography, 2007. **40**(4): p. 658-674.
116. Emsley, P., et al., *Features and development of Coot.* Acta Crystallographica Section D Biological Crystallography, 2010. **66**(4): p. 486-501.
117. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine.* Acta Crystallographica Section D, 2012. **68**(4): p. 352-367.
118. Williams, C.J., et al., *MolProbity: More and better reference data for improved all-atom structure validation.* Protein Sci, 2018. **27**(1): p. 293-315.
119. Carroni, M. and H.R. Saibil, *Cryo electron microscopy to determine the structure of macromolecular complexes.* Methods, 2016. **95**: p. 78-85.
120. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nature Methods, 2017. **14**(3): p. 290-296.
121. Bepler, T., et al., *Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs.* Nat Methods, 2019. **16**(11): p. 1153-1160.
122. Scheres, S.H.W., *RELION: Implementation of a Bayesian approach to cryo-EM structure determination.* Journal of Structural Biology, 2012. **180**(3): p. 519-530.
123. Afonine, P.V., et al., *Real-space refinement in PHENIX for cryo-EM and crystallography.* Acta Crystallographica Section D, 2018. **74**(6): p. 531-544.
124. Croll, T.I., *ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps.* Acta Crystallographica Section D Structural Biology, 2018. **74**(6): p. 519-530.
125. LAGUERRE, M., et al., *MLPP: A Program for the Calculation of Molecular Lipophilicity Potential in Proteins.* Pharmacy and Pharmacology Communications, 1997. **3**(5-6): p. 217-222.
126. Ghose, A.K., V.N. Viswanadhan, and J.J. Wendoloski, *Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods.* The Journal of Physical Chemistry A, 1998. **102**(21): p. 3762-3772.
127. Fiskin, E., et al., *Structural basis for the recognition and degradation of host TRIM proteins by Salmonella effector SopA.* Nature Communications, 2017. **8**(1): p. 14004.
128. Lin, D.Y., et al., *Biochemical and structural studies of a HECT-like ubiquitin ligase from Escherichia coli O157:H7.* J Biol Chem, 2011. **286**(1): p. 441-9.
129. Ward, J.J., et al., *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, 2004. **20**(13): p. 2138-2139.
130. Fairhead, M. and M. Howarth, *Site-specific biotinylation of purified proteins using BirA.* Methods Mol Biol, 2015. **1266**: p. 171-84.
131. Beugelink, J.W., et al., *Structural Basis for Recognition of the FLAG-tag by Anti-FLAG M2.* Journal of Molecular Biology, 2024. **436**(16): p. 168649.
132. Whitby, F.G., et al., *Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes.* J Biol Chem, 1998. **273**(52): p. 34983-91.
133. Burger, D., et al., *Crystal Structures of the Human Doublecortin C- and N-terminal Domains in Complex with Specific Antibodies.* J Biol Chem, 2016. **291**(31): p. 16292-306.
134. Krissinel, E. and K. Henrick, *Detection of Protein Assemblies in Crystals*. 2005, Springer Berlin Heidelberg. p. 163-174.

135. Couillard-Despres, S., et al., *Doublecortin expression levels in adult brain reflect neurogenesis.* Eur J Neurosci, 2005. **21**(1): p. 1-14.
136. Sébastien, M., et al., *Doublecortin restricts neuronal branching by regulating tubulin polyglutamylation.* Nature Communications, 2025. **16**(1): p. 1749.
137. Scott, D.C., et al., *E3 ligase autoinhibition by C-degron mimicry maintains C-degron substrate fidelity.* Molecular Cell, 2023. **83**(5): p. 770-786.e9.
138. Noemi, et al., *The IntAct database: efficient access to fine-grained molecular interaction data.* Nucleic Acids Research, 2022. **50**(D1): p. D648-D653.
139. Luck, K., et al., *A reference map of the human binary protein interactome.* Nature, 2020. **580**(7803): p. 402-408.
140. Brückner, A., et al., *Yeast Two-Hybrid, a Powerful Tool for Systems Biology.* International Journal of Molecular Sciences, 2009. **10**(6): p. 2763-2788.
141. Deane, C.M., et al., *Protein Interactions.* Molecular & Cellular Proteomics, 2002. **1**(5): p. 349-356.
142. Roux, K.J., et al., *BioID: A Screen for Protein-Protein Interactions.* Current Protocols in Protein Science, 2018. **91**(1).
143. Pontén, F., et al., *A global view of protein expression in human cells, tissues, and organs.* Mol Syst Biol, 2009. **5**: p. 337.
144. Tong, S., et al., *MLN4924 (Pevonedistat), a protein neddylation inhibitor, suppresses proliferation and migration of human clear cell renal cell carcinoma.* Scientific Reports, 2017. **7**(1): p. 5599.
145. Chen, Z., et al., *Disease-associated KBTBD4 mutations in medulloblastoma elicit neomorphic ubiquitylation activity to promote CoREST degradation.* Cell Death & Differentiation, 2022. **29**(10): p. 1955-1969.
146. Sherman, B.T., et al., *DAVID Ortholog: an integrative tool to enhance functional analysis through orthologs.* Bioinformatics, 2024. **40**(10).
147. Thul, P.J. and C. Lindskog, *The human protein atlas: A spatial map of the human proteome.* Protein Sci, 2018. **27**(1): p. 233-244.
148. Oughtred, R., et al., *The <scp>BioGRID</scp> database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions.* Protein Science, 2021. **30**(1): p. 187-200.
149. Uhlén, M., et al., *Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
150. Dubiel, W., et al., *The COP9 Signalosome: A Multi-DUB Complex.* Biomolecules, 2020. **10**(7): p. 1082.
151. Maris, C., C. Dominguez, and F.H. Allain, *The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression.* Febs j, 2005. **272**(9): p. 2118-31.
152. Ho, W.-H.J., et al., *Activation of the viral sensor oligoadenylate synthetase 2 (Oas2) prevents pregnancy-driven mammary cancer metastases.* Breast Cancer Research, 2022. **24**(1).
153. Mahmud, S.A., L.S. Manlove, and M.A. Farrar, *Interleukin-2 and STAT5 in regulatory T cell development and function.* JAK-STAT, 2013. **2**(1): p. e23154.
154. Smith, M.R., L.R.F. Satter, and A. Vargas-Hernández, *STAT5b: A master regulator of key biological pathways.* Frontiers in Immunology, 2023. **13**.
155. Castro, W., et al., *The transcription factor Rfx7 limits metabolism of NK cells and promotes their maintenance and immunity.* Nature Immunology, 2018. **19**(8): p. 809-820.
156. Shi, H., et al., *BCL11A Is Oncogenic and Predicts Poor Outcomes in Natural Killer/T-Cell Lymphoma.* Frontiers in Pharmacology, 2020. **11**.
157. Cheng, J. and W. Guggino, *Ubiquitination and Degradation of CFTR by the E3 Ubiquitin Ligase MARCH2 through Its Association with Adaptor Proteins CAL and STX6.* PLoS ONE, 2013. **8**(6): p. e68001.
158. Amacher, J.F., et al., *Stereochemical Determinants of C-terminal Specificity in PDZ Peptide-binding Domains.* Journal of Biological Chemistry, 2013. **288**(7): p. 5114-5126.
159. Feng, S.H., C.Q. Xia, and H.B. Shen, *CoCoPRED: coiled-coil protein structural feature prediction from amino acid sequence using deep neural networks.* Bioinformatics, 2022. **38**(3): p. 720-729.
160. Wang, H.-C., et al., *DNA Mimic Proteins: Functions, Structures, and Bioinformatic Analysis.* Biochemistry, 2014. **53**(18): p. 2865-2874.

161. Huttlin, E.L., et al., *Dual proteome-scale networks reveal cell-specific remodeling of the human interactome.* Cell, 2021. **184**(11): p. 3022-3040.e28.
162. Ibarra, A. and M.W. Hetzer, *Nuclear pore proteins and the control of genome functions.* Genes Dev, 2015. **29**(4): p. 337-49.
163. Mosalaganti, S., et al., *AI-based structure prediction empowers integrative structural analysis of human nuclear pores.* Science, 2022. **376**(6598): p. eabm9506.
164. Meng, M., et al., *Zinc finger protein ZNF638 regulates triglyceride metabolism via ANGPTL8 in an estrogen dependent manner.* Metabolism, 2024. **152**: p. 155784.
165. Zhao, B., et al., *Structural Elucidation of Peptide Binding to KLHL-12, a Substrate Specific Adapter Protein in a Cul3-Ring E3 Ligase Complex.* Biochemistry, 2020. **59**(8): p. 964-969.
166. Krissinel, E. and K. Henrick, *Inference of Macromolecular Assemblies from Crystalline State.* Journal of Molecular Biology, 2007. **372**(3): p. 774-797.
167. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* Journal of Molecular Biology, 1982. **157**(1): p. 105-132.
168. Kumar, M., et al., *ELM—the Eukaryotic Linear Motif resource—2024 update.* Nucleic Acids Research, 2023. **52**(D1): p. D442-D455.
169. Kipreos, E.T. and M. Pagano, *The F-box protein family.* Genome Biology, 2000. **1**(5): p. reviews3002.1.
170. Zhang, X., et al., *Interference with KCTD9 inhibits NK cell activation and ameliorates fulminant liver failure in mice.* BMC Immunol, 2018. **19**(1): p. 20.
171. Shimasaki, N., A. Jain, and D. Campana, *NK cells for cancer immunotherapy.* Nature Reviews Drug Discovery, 2020. **19**(3): p. 200-218.
172. Seetharam, D., et al., *Targeting ZNF638 activates antiviral immune responses and potentiates immune checkpoint inhibition in glioblastoma.* bioRxiv, 2024: p. 2024.10.13.618076.
173. Foster, B.M., et al., *Critical Role of the UBL Domain in Stimulating the E3 Ubiquitin Ligase Activity of UHRF1 toward Chromatin.* Mol Cell, 2018. **72**(4): p. 739-752.e9.
174. Garber, K., *The PROTAC gold rush.* Nature Biotechnology, 2022. **40**(1): p. 12-16.
175. Okabe, H., et al., *A critical role for FBXW8 and MAPK in cyclin D1 degradation and cancer cell proliferation.* PLoS One, 2006. **1**(1): p. e128.
176. Yu, Z.K., J.L. Gervais, and H. Zhang, *Human CUL-1 associates with the SKP1/SKP2 complex and regulates p21(CIP1/WAF1) and cyclin D proteins.* Proc Natl Acad Sci U S A, 1998. **95**(19): p. 11324-9.
177. Lin, D.I., et al., *Phosphorylation-dependent ubiquitination of cyclin D1 by the SCF(FBX4-alphaB crystallin) complex.* Mol Cell, 2006. **24**(3): p. 355-66.
178. Santra, M.K., N. Wajapeyee, and M.R. Green, *F-box protein FBXO31 mediates cyclin D1 degradation to induce G1 arrest after DNA damage.* Nature, 2009. **459**(7247): p. 722-725.
179. Scheich, C., et al., *Vectors for co-expression of an unrestricted number of proteins.* Nucleic Acids Research, 2007. **35**(6): p. e43-e43.
180. Schönfeld, H.-J., et al., *The DnaK Chaperone System of Escherichia coli: Quaternary Structures and Interactions of the DnaK and GrpE Components (∗).* Journal of Biological Chemistry, 1995. **270**(5): p. 2183-2189.
181. Claassens, N.J., et al., *Bicistronic Design-Based Continuous and High-Level Membrane Protein Production in Escherichia coli.* ACS Synthetic Biology, 2019. **8**(7): p. 1685-1690.
182. Graham, A.I., et al., *Severe zinc depletion of Escherichia coli: roles for high affinity zinc binding by ZinT, zinc transport and zinc-independent proteins.* J Biol Chem, 2009. **284**(27): p. 18377-89.
183. Ng, Y.Z., P.A. Baldera-Aguayo, and V.W. Cornish, *Fluorescence Polarization Assay for Small Molecule Screening of FK506 Biosynthesized in 96-Well Microtiter Plates.* Biochemistry, 2017. **56**(40): p. 5260-5268.
184. Jankovics, H., et al., *Grating-coupled interferometry reveals binding kinetics and affinities of Ni ions to genetically engineered protein layers.* Scientific Reports, 2020. **10**(1): p. 22253.
185. Capelli, D., V. Scognamiglio, and R. Montanari, *Surface plasmon resonance technology: Recent advances, applications and experimental cases.* TrAC Trends in Analytical Chemistry, 2023. **163**: p. 117079.
186. Degorce, F., et al., *HTRF: A technology tailored for drug discovery - a review of theoretical aspects and recent applications.* Curr Chem Genomics, 2009. **3**: p. 22-32.

187. Lin, K. and G. Wu, *Isothermal Titration Calorimetry Assays to Measure Binding Affinities In Vitro.* Methods Mol Biol, 2019. **1893**: p. 257-272.
188. Friesner, R.A., et al., *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy.* Journal of Medicinal Chemistry, 2004. **47**(7): p. 1739-1749.
189. Grygorenko, O.O., *Enamine Ltd.: The Science and Business of Organic Chemistry and Beyond.* European Journal of Organic Chemistry, 2021. **2021**(47): p. 6474-6477.
190. Kralj, S., M. Jukič, and U. Bren, *Comparative Analyses of Medicinal Chemistry and Cheminformatics Filters with Accessible Implementation in Konstanz Information Miner (KNIME).* Int J Mol Sci, 2022. **23**(10).
191. Baell, J.B. and J.W.M. Nissink, *Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017—Utility and Limitations.* ACS Chemical Biology, 2018. **13**(1): p. 36-44.
192. Leung, S., et al., *SuCOS is Better than RMSD for Evaluating Fragment Elaboration and Docking Poses*. 2019.
193. Grigoryan, A.V., et al., *Spatial chemical distance based on atomic property fields.* Journal of Computer-Aided Molecular Design, 2010. **24**(3): p. 173-182.
194. Daina, A., O. Michielin, and V. Zoete, *SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules.* Scientific Reports, 2017. **7**(1): p. 42717.
195. Fu, H., et al., *Codon optimization with deep learning to enhance protein expression.* Scientific Reports, 2020. **10**(1): p. 17617.
196. Wright, G., et al., *CHARMING: Harmonizing synonymous codon usage to replicate a desired codon usage pattern.* Protein Sci, 2022. **31**(1): p. 221-231.
197. Muhar, M.F., et al., *C-terminal amides mark proteins for degradation via SCF–FBXO31.* Nature, 2025. **638**(8050): p. 519-527.