Structure and function of E3 ligase complexes in immune signalling regulation

Linda Makhlouf

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
School of Molecular and Cellular biology

October 2023

Intellectual property and publication statement

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyrighted material and that no quotation from the thesis may be published without proper acknowledgement.

Work from **Chapter 3** will be included in a manuscript which is currently under preparation.

Linda Makhlouf, Mukul Mishra, Luca Busino, Stephanie Wright and Elton Zeqiraj. *Mechanism of MyD88 recognition by the E3 ligase adaptor SPOP. Manuscript under preparation.*

Work from **Chapter 6** is currently under a second round of review at Nature.

Linda Makhlouf, Joshua J. Peter, Helge M. Magnussen, Rohan Thakur, David Millrine, Thomas C. Minshull, Grace Harrison, Joby Varghese, Frederic Lamoliatte, Martina Foglizzo, Thomas Macartney, Antonio N. Calabrese, Elton Zeqiraj and Yogesh Kulathu. *Mechanism of 60S ribosomal subunit recognition and UFMylation by the UFM1 E3 ligase complex*. Nature. *In review.*

Acknowledgements

First of all, I'd like to express my sincere gratitude to my supervisors Elton Zeqiraj and Stephanie Wright for their guidance over the past few years and for taking the time to have weekly meetings with me. I'd also like to thank my co-supervisor Richard Bayliss for his support. I acknowledge the Wellcome Trust for funding and the opportunity to work on this PhD project. I'd also like to give a huge thank you to the members of the Zeqiraj lab: Martina Foglizzo, Miriam Walden, Lisa Campbell, Daniel Maskell, Mohd Syed Ahanger, Emma Cowan, Francesca Chandler, Laura Marr, Laura Musgrove, Jordan Liburd and Rehab Alsofayan. Thank you all for all of your help and your great company over the past few years. I'd like to give a special thanks to Martina Foglizzo for all of her assistance at the beginning of my PhD, and to Mohd Syed Ahanger for all of his help with X-ray crystallography.

I am very grateful to our collaborators from the Busino lab: Mukul Mishra and Luca Busino for their valuable advice and assistance for the SPOP-MyD88 project. I'd also like to thank the Yogesh Kulathu lab, particularly Yogesh Kulathu, Helge Magnussen, Joshua Peter and Rohan Thakur for their hard work and excellent collaboration for the UFL1 ligase project. A lot of the work carried out in this thesis was conducted at the fantastic FBS research facilities at the University of Leeds. In particular I'd like to thank everyone in the EM facility: Rebecca Thompson, Emma Hesketh, Louie Aspinall, Yehuda Halfon, Oksana Degtjarik, Joshua White, Tom O'Sullivan and Martin Fuller. Also, thank you the mass spectrometry facility, in particular Ranjani Ganji for her assistance, and the PIXC facility, particularly Iain Manfield for his help with ITC.

Lastly, I'd like to thank my family: Lila, Kamel and Hannah Makhlouf. Thank you for your continuous support and encouragement, not only for my PhD but for life in general.

Abstract

E3 ligases perform the final step of an enzyme cascade reaction where ubiquitin, or ubiquitin-like molecules, are covalently attached to substrates. Understanding the structural basis of how these E3 ligase complexes assemble and modify substrates is key to ongoing efforts to study their roles in biology.

The immune signalling protein MyD88 is ubiquitylated by the E3 ligase SPOP as a mode of negatively regulating immune signalling. Here I show that MyD88 binds to SPOP using an unusual extended binding motif and a 2.0 Å X-ray crystal structure of SPOP bound to a MyD88 peptide reveals the molecular basis of this interaction. I also present low-resolution cryo-electron microscopy (cryo-EM) data of the SPOP oligomer and the high-order oligomeric SPOP-MyD88 complex, potentially revealing a new conformation of oligomeric E3 ligase-substrate complexes.

The UFL1 E3 ligase complex catalyses the transfer of the ubiquitin-like molecule UFM1 and UFMylates ribosomes. I present the cryo-EM structure of the UFL1 E3 ligase complex bound to the 60S ribosome, which shows how this three-subunit ligase complex is poised to transfer UFM1 onto the 60S protein RPL26. The ligase obstructs the tRNA binding sites at one end and SEC61/peptide exit tunnel at the other end. Additionally, a long UFL1 loop extends into the peptidyl transferase centre (PTC), acting as a potential sensor of PTC occupancy status. The intricate interactions between the ligase complex and multiple ribosomal proteins provide a rationale for UFM1 modifications in regulating ribosome quality control. Collectively, these results provide a long-awaited structural mechanism for UFM1 transfer and lay the groundwork for investigating the mechanism of action and biological relevance of ribosome UFMylation.

Table of contents

Lists of Tables

Illustrative Material

Abbreviations

ΔH	Enthalpy change
ΔS	Entropy change
3DFlex	3D Flexible Refinement
3DVA	3-dimensional variability analysis
5-Flu	5-Carboxyfluorescein
A-tRNA	Aminoacyl-tRNA
Ac	Acetyl
AR	Androgen receptor
ASC1	Activating signal cointegrator
BTB	Broad complex, Tramtrack, and Bric-à-brac
C-term	C-terminus
Cal/mol	Calories per molar
CDK5RAP3	Cyclin-dependent kinase 5 regulatory subunit associated protein 3
Co-IP	Co-immunoprecipitation
CTD	C-terminal domain
DD	Death domain
ddH_20	Double distilled H_20
DLBCL	Diffuse large B-cell lymphoma
Dm	*Drosophila melanogaster* (Fruit fly)
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
DUB	Deubiquitylase
E-tRNA	Exit-tRNA
E. coli	*Escherichia coli*
E1	Ubiquitin-activating enzyme
E2	Ubiquitin-conjugating enzyme
E3	Ubiquitin ligase
EDTA	Ethylenediaminetetraacetic acid
eEF	Eukaryotic elongation factor
eIF	Eukaryotic initiation factor
EM	Electron microscopy
ER	Endoplasmic reticulum
ERα	Estrogen receptor-alpha

ERAD	ER-associated protein degradation
eRF	Eukaryotic release factor
FL	Full-length
FP	Fluorescence polarization
FT	Flow-through
GA	Glutaraldehyde
GST	Glutathione s-transferase
HECT	Homologous to the E6AP carboxyl terminus
His-tag	Histadine^{6}-tag
HPN	His-Pro-Asp
HRC 3C	Human rhinovirus 3C protease
Hs	*Homo sapiens*
HSC	Haematopoietic stem cell
IC_{50}	Half maximal inhibitory concentration
ID	Intermediate domain
IL1-R	Interleukin-1 receptor
IPTG	Isopropyl-β-D-thiogalactopyranoside
IRAK	Interleukin-1 receptor-associated kinase
ITC	Isothermal titration calorimetry
K	Lysine
kb	Kilo-base
K_d	Dissociation constant
KD	Kinase domain
kDa	Kilo-Daltons
Lab	Laboratory
LB	Luria Bertani
M	Molar
Ma	*Mesocricetus auratus* (Golden hamster)
MAPK	Mitogen-activated protein kinase
MATH	Meprin and TRAF homology
MBP	Maltose binding protein
mM	Millimolar
mP	Millipolarization
mRNA	Messenger RNA
MyD88	Myeloid differentiation primary response protein 88
MyD88^{DD-ID}	MyD88 residues 1-158

N-term	N-terminus
NaCl	Sodium chloride
nb	No binding
nM	Nanomolar
OD_{600}	Optical density at 600 nm
P-tRNA	Peptidyl-tRNA
PAMPs	Pattern-associated molecular patterns
Pdx1	Pancreatic duodenal homeobox 1
Pix	Pixels
PMSF	Phenylmethylsulfonyl fluoride
Poly-A	Poly-adenine
pS	Phospho-serine
pT	Phospho-threonine
PTC	Peptidyl-transferase centre
Puc	Puckered
pWH	Partial winged helix
RBD	RPL10a binding domain
RBR	RING-in between RING-RING
RING	Really interesting new gene
RNA	Ribonucleic acid
RPL	Ribosome protein large
rpm	Revolutions per minute
RQC	Ribosome-associated quality control
rRNA	Ribosome RNA
RT	Room temperature
SBC	SPOP binding consensus
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SPOP	Speckle-type BTB-POZ protein
$SPOP^{18-359}$	SPOP residues 18-359
$SPOP^{28-359}$	SPOP residues 28-359
SPOPL	Speckle-type POZ protein-like
$SPOP^{MATH}$	SPOP residues 28-166
SRC-3	Steroid receptor coactivator-3

TB	Terrific Broth
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TIR	Toll/Interleukin-1 receptor
TLR	Toll-like receptor
TMH	Transmembrane helix
TRAF-6	Tumour necrosis factor receptor-associated factor 6
tRNA	Transfer RNA
Ub	Ubiquitin
UBA5	Ubiquitin-like modifier activating enzyme 5
UBD	Ubiquitin binding domain
UFBP1	UFM1 binding protein
UFC1	Ubiquitin-fold modifier conjugating enzyme 1
UFIM	UFM1 interacting motif
UFL1	UFM1-specific ligase 1
UFM1	Ubiquitin-fold modifier 1
UFSP2	UFM1-specific protease 2
UUBD	UFM1 and UFBP1 binding domain
w/v	Weight per volume
WH	Winged helix
YCF	YxCxxxF motif
µM	Micromolar
π	Polar residue
φ	Non-polar residue

Chapter 1 - Introduction

1.1 - Ubiquitin

In 1974, an 8.6 kilo-Dalton (kDa) polypeptide was discovered that induced the differentiation of thymus-derived (T) cells and bone marrow-derived (B) cells [1]. This polypeptide, called ubiquitin, was found to be well conserved in eukaryotes, and was ubiquitously expressed in all cells. Ubiquitin was later shown to be covalently-linked to histone 2A, where the ubiquitin C-terminal glycine-76 was isopeptide-linked to a lysine side chain ε-amino group of the target protein, a process called ubiquitylation [2]. The ubiquitin protein is 76 amino acids long, with a β-grasp fold and a flexible C-terminal tail [3] (**Fig. 1.1**). Ubiquitin itself contains seven lysine (K) residues (K-6, K-11, K-27, K-29, K-33, K-48, K-63) which can also be ubiquitylated (**Fig. 1.1**).

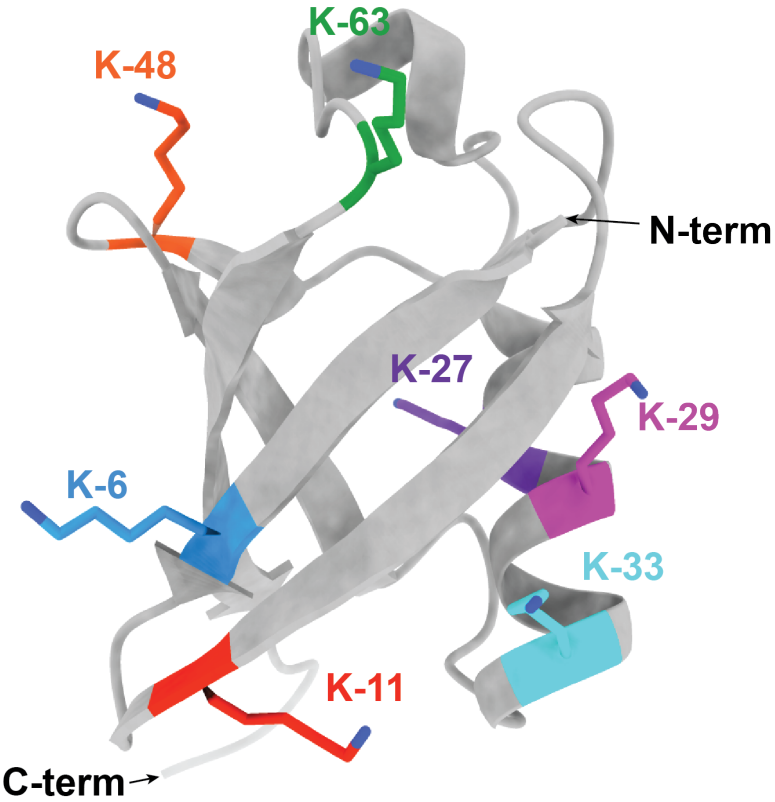


Figure 1.1 – Ubiquitin has a β-grasp fold

Structure of ubiquitin (PDB ID 1UBQ) [3] highlighting the seven lysine residues (K) that can be ubiquitylated to form poly-ubiquitin chains. N-terminus (N-term) and C-terminus (C-term) of ubiquitin are shown with black arrows.

1.2 - Ubiquitylation

Ubiquitylation of target proteins signals many cellular events, including gene transcription, proteasomal degradation and DNA damage repair [4]. There is a vast network of proteins and protein complexes which can add, read, and remove ubiquitin from target proteins. The ubiquitylation cascade is a multistep process involving several enzymes. These are: ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2) and ubiquitin ligase (E3), with approximately 9 E1 enzymes, 40 E2 enzymes and over 600 E3 enzymes across ubiquitin, or ubiquitin-like enzymes, in humans [5, 6].

In the ubiquitin enzyme cascade, ubiquitin is first activated by an E1 activating enzyme in an ATP-dependent reaction, resulting in its C-terminal G-76 adenylation [7]. This ubiquitin-adenylate intermediate is then attacked by the E1 enzyme active site cysteine, converting the ubiquitin C-terminal G-76 to a high-energy thiol ester intermediate [5, 8] (**Fig. 1.2**). Next, ubiquitin is transferred from the E1 to the active site cysteine of an E2 conjugating enzyme via a trans-thiolation reaction, forming a ubiquitin-E2 conjugate [9] (**Fig. 1.2**). This ubiquitin-E2 conjugate is then recruited to an E3 ligase enzyme, which facilitates the transfer of ubiquitin from the catalytic cysteine of the E2 to a target protein [6] (**Fig. 1.2**). Ubiquitin E2s contain a highly conserved catalytic histidine-proline-asparagine (HPN) motif adjacent to the catalytic cysteine which aids in ubiquitylation [10].

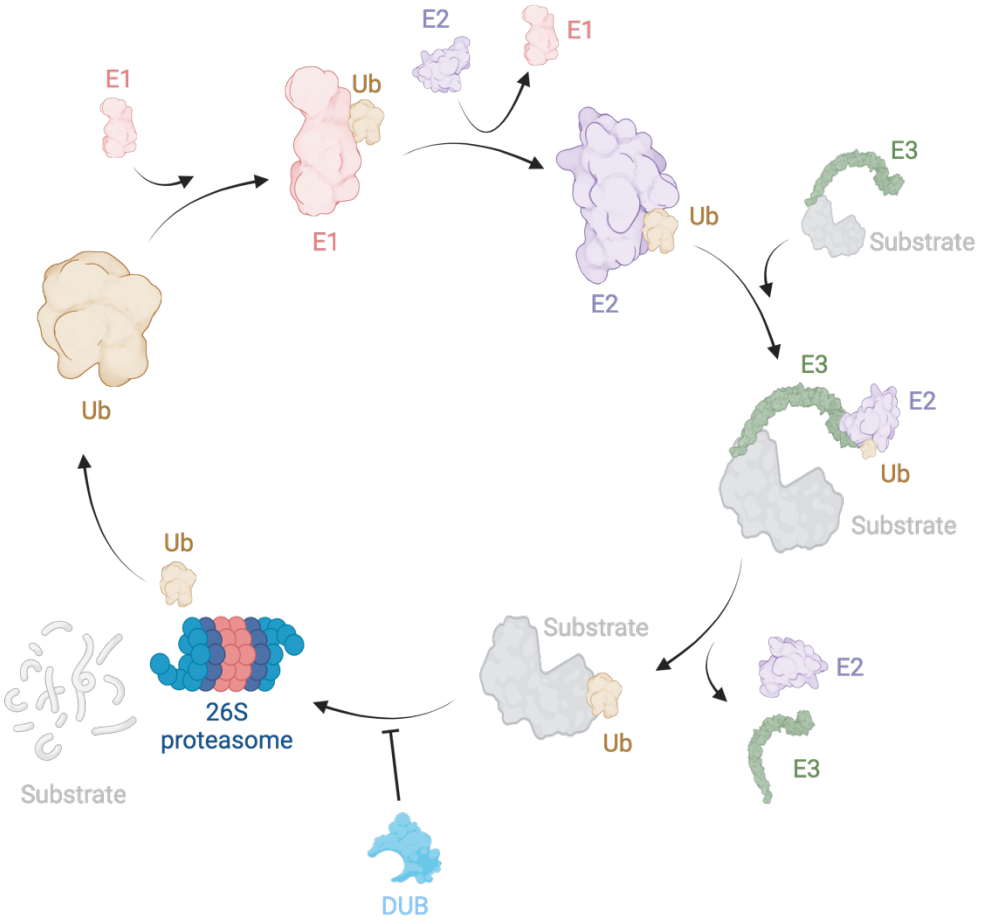


Figure 1.2 – Schematic showing the steps of the ubiquitylation enzyme cascade
E1 is ubiquitin-activating enzyme, E2 is ubiquitin-conjugating enzyme and E3 is ubiquitin ligase. DUB is deubiquitylase. Ubiquitin (Ub) binds to E1 and is activated. Ub is transferred from the E1 to the active site of E2. The Ub-E2 conjugate is recruited to substrate bound-E3. Ub is then isopeptide-linked to the substrate. The substrate is then targeted for downstream processes, such as protein degradation. DUB inhibits the ubiquitylation signal by removing Ub from substrate. Ub is then free to be attached to another substrate. Schematic created with BioRender.com.

E3 ligase complexes are able to catalyse multiple additions of ubiquitin onto substrates. Ubiquitin itself can be ubiquitylated on its lysine residues (K-6, K-11, K-27, K-29, K-33, K-48, K-63), or on its N-terminal methionine to create poly-ubiquitin chains with diverse topologies and length. Additionally, ubiquitin can also be acetylated on its lysine residues, or phosphorylated (T-12, S-57 and S-65), which can serve as either an inhibitor of ubiquitylation or as part of ubiquitin signalling, such as phospho-serine-65-ubiquitin which target mitochondrial proteins for

autophagosomal degradation (mitophagy) [11-15]. These poly-ubiquitin chains can be formed of a single type of linkage or of several different types of linkages (branched chains). This creates diverse poly-ubiquitin chains, called the ubiquitin code, providing a variety of intracellular signals. These signals target proteins for degradation by the 26S proteasome, or can serve as non-degradative signals, such as regulating protein-protein interactions, determining sub-cellular localisation, receptor signalling and cell differentiation (reviewed in [16]).

The most prevalent poly-ubiquitin linkage is K-48, which usually signals for target proteins to be degraded by the 26S proteasome [17]. Another common poly-ubiquitin chain linkage is K-63, which is typically a non-degradative signal although it can signal target substrates for degradation via the lysosomal pathway. K-63 chains are also involved in cellular signalling events, such as immune signalling and the DNA damage response [18]. M-1-linked poly-ubiquitin chains have been implicated as a mode of regulating the NF-κB signalling pathway [19-21]. The roles of K-6 and K-33 ubiquitylation are less clear, with these ubiquitin linkages implicated in regulating various proteins involved in the DNA damage response [22]. Lastly, K-11 poly-ubiquitin chain are involved in cell cycle regulation [23] and K-27 linked chains are involved in mitophagy regulation [24, 25].

Differing chain linkages and chain lengths are recognised by specific reader proteins. These recognise the structural arrangements of the ubiquitin linkage types through single or multiple ubiquitin binding domains (UBD). The UBDs bind to a hydrophobic patch on ubiquitin (L-8, I-44, H-68, V-70), “reading” the poly-ubiquitin signal to enact specific downstream ubiquitin cellular signalling [26].

Conversely, these diverse ubiquitin modifications can also be removed from a substrate, terminating ubiquitin signalling. There are around 100 deubiquitylases (DUBs) in humans. DUBs are proteases that are able to

cleave the isopeptide link of ubiquitin from target proteins, disrupting ubiquitination signalling and freeing ubiquitin back to the ubiquitin pool [27, 28]. There are seven families of DUBs which act as papain-like cysteine proteases (USP, OTU, UHC, Josephin, MINDY, ZUP1) or zinc-dependent metallo-proteases (JAMM) [29]. Some of these DUBs have lost their catalytic activity and are classed as pseudo-DUBs [30]. These pseudo-DUBs still have roles in regulating the ubiquitin system where they can act as scaffolding proteins with other catalytically active DUBs [31].

1.3 - Ubiquitin E3 ligases

There are over 600 ubiquitin E3 ligases in *Homo sapiens* (*Hs*), which enable the ubiquitylation of a diverse range of substrates. E3 ligases belong to four main classes, with different ubiquitylation mechanisms.

1.3.1 - Homologous to the E6AP carboxyl terminus (HECT) E3 ligases

There are approximately 30 HECT family E3 ligases, which are characterised by the presence of a HECT domain at their C-terminus, with variable substrate binding domains at their N-terminus [32, 33]. The HECT domains consist of an N-terminal lobe that recruits the E2 enzyme and a flexible C-terminal lobe which contains a catalytic cysteine [32, 34]. HECT ligase-mediated ubiquitylation begins with the E2-ubiquitin conjugate binding to the N-terminal lobe (**Fig. 1.3**). Ubiquitin is then transferred to the HECT domain active site cysteine by trans-thiolation, resulting in a HECT ligase-ubiquitin thioester (**Fig. 1.3**) [33, 35]. The HECT ligase then catalyses the transfer of ubiquitin onto the target protein (**Fig. 1.3**) [34]. The poly-ubiquitin chain linkage type is determined by the C-terminal tail of HECT E3 ligases [36].

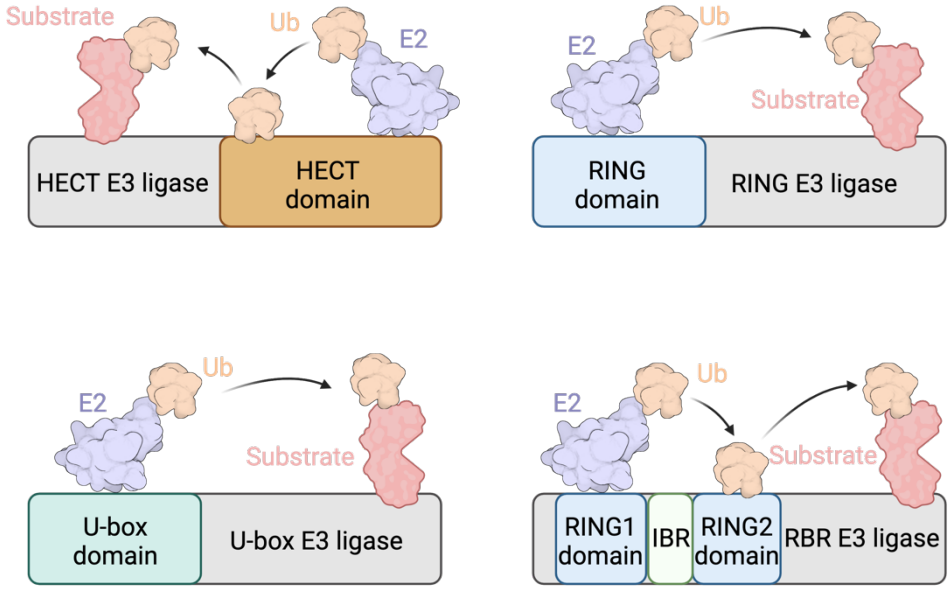


Figure 1.3 – Schematic of E3 ubiquitin ligase ubiquitylation mechanisms
Overview of the ubiquitin (Ub) transfer mechanism for different classes of Ub E3 ligases. The E2 conjugating enzyme binds to specific domains of E3 ligases. The E3 ligase then acts as a scaffold for the direct transfer of Ub from the E2 to the target substrate (RING and U-box E3 ligases). Alternatively, Ub is transferred to the E3 ligase active site cysteine, forming a Ub-E3 intermediate, before transfer of Ub to the target substrate (HECT and RBR E3 ligases). Created with BioRender.com.

1.3.2 - Really interesting new gene (RING) E3 ligases

The largest family of ligases are the RING ligases, which recruit E2-ubiquitin conjugates via their RING finger domain. RING domains contain a short α-helix and multiple zinc-binding residues, such as histidine and cysteine, which form cross-braced zinc finger motifs to coordinate zinc atoms [37]. These RING zinc fingers enable binding to E2 enzymes [38].

Unlike the HECT class of ligase enzymes, the RING ligases do not form a ligase-ubiquitin intermediate, instead facilitating ubiquitylation through a scaffold-type mechanism. The ubiquitin-E2 conjugate binds to the ligase RING domain (**Fig. 1.3**). The ligase recruits the substrate and ubiquitin is then transferred onto the substrate directly from the E2 enzyme (**Fig. 1.3**).

1.3.3 – Cullin-RING E3 ligases

The largest sub-group of RING ligases are the multi-subunit Cullin-RING ligases. This family of ligases contains a Cullin ligase (Cullin-1, Cullin-2, Cullin-3, Cullin-4a, Cullin-4b, Cullin-5, Cullin-7), which acts as a scaffold to recruit multiple proteins to form a functional E3 ligase complex (**Fig. 1.4**) [39]. The second component of the Cullin-RING ligase complexes are adaptor proteins which help to recruit substrate to the ligase complex. These adaptors bind to Cullin repeat domains at the N-terminus of the Cullin ligases (**Fig. 1.4**) [40]. The final component of the Cullin-RING ligases are RING domain containing proteins, such as Rbx1, which bind to the C-terminus of Cullins (**Fig. 1.4**) [40]. These RING proteins help to recruit charged E2 enzyme to the Cullin-RING ligase complex. The modification of the Cullin-RING ligase complex by the ubiquitin-like molecule, NEDD8, has been shown to activate its ubiquitylation activity by inducing a conformation change of the RING protein in relation to the Cullin C-terminus that favours ubiquitin transfer [41-43].

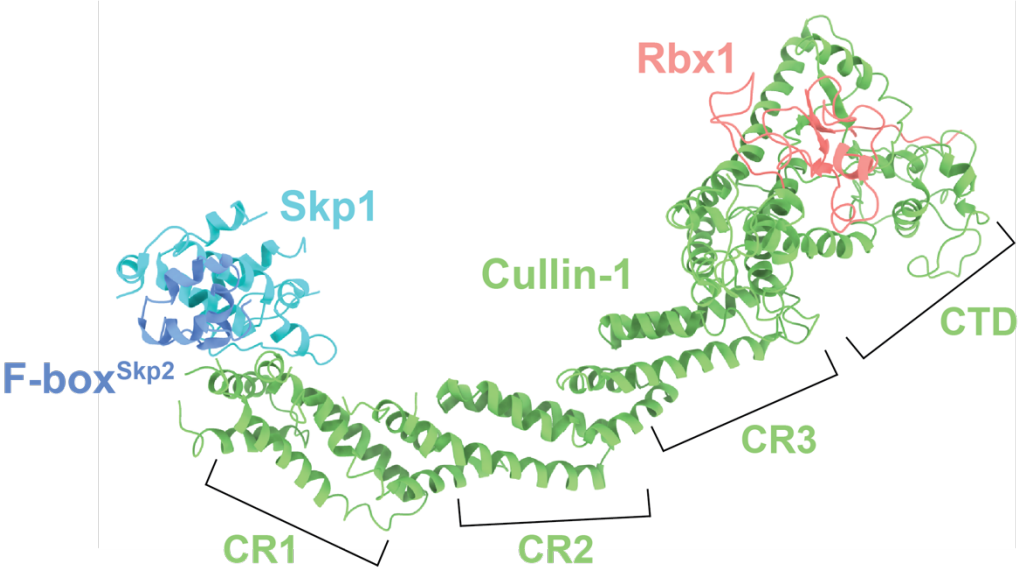


Figure 1.4 – Structure of a Cullin-1-RING ligase complex
Cullin-1 is the E3 ligase and contains three Cullin repeat (CR) domains and a C-terminal domain (CTD). The Skp1-F-box^{Skp2} complex binds to the N-terminus of Cullin-1 and acts as a substrate adaptor. Rbx1 binds to the CTD of Cullin-1 and recruits charged E2 enzyme to the complex. PDB ID 1LDK [40].

1.3.4 - U-box E3 ligases

The U-box E3 ligases are a smaller family of ligases structurally similar to the RING E3 ligases [44]. However, they possess a U-box domain instead of a RING domain, which is not capable of metal-chelation. Similar to the RING ligases, U-box ligases assist in the direct transfer of ubiquitin from the E2 enzyme [45] (**Fig. 1.3**).

1.3.5 - RING-between-RING (RBR) E3 ligases

The RBR ligases are characterised by the presence of a RING domain followed by an in-between-RING (IBR) domain, then a second RING domain. Unlike RING ligases, the second RING domain of RBR ligases has catalytic activity. The RBR ligases can be considered as a hybrid between RING and HECT ligases. They contain RING domains but use a ubiquitylation method similar to HECT ligases. First, the ubiquitin-E2 conjugate is recruited to a RING domain (RING1) of the RBR ligase (**Fig. 1.3**). Then, ubiquitin is transferred from the E2 to an active site cysteine on a second RING domain (RING2), aided by the IBR domain [46, 47] (**Fig. 1.3**). Ubiquitin is then transferred from the E3 ligase to the substrate [48, 49].

The diverse ubiquitylation mechanisms and unique structures and complex assemblies of the E3 ligases determine the poly-ubiquitin linkage specificity on the substrate. The poly-ubiquitin linkage specificity of RING and U-box E3 ligases is determined by selection of the E2, as they directly transfer ubiquitin to the substrate [50, 51]. Conversely, HECT and RBR ligases determine the poly-ubiquitin linkage specificity by using different sub-families of the ligase class, although the exact mechanisms remain poorly understood [4, 52].

Many E3 ligases use an adaptor protein for the recruitment of target proteins for ubiquitylation. These adaptors contain an E3 ligase-binding domain and a substrate-binding domain. In the case of the Cullin-3-RING

ligases, these adaptors recognise Cullin-3 using a Broad complex, Tramtrack, and Bric-à-brac domain (BTB) domain [53].

1.4 - Speckle-type POZ protein (SPOP)

The 42 kDa speckle-type POZ protein (SPOP) was first identified from the serum of a scleroderma patient [54]. SPOP is an E3 ubiquitin ligase adaptor, which binds to Cullin-3-Rbx1 E3 ligase complex and recruits substrates for ubiquitylation [55] (**Fig. 1.5**).

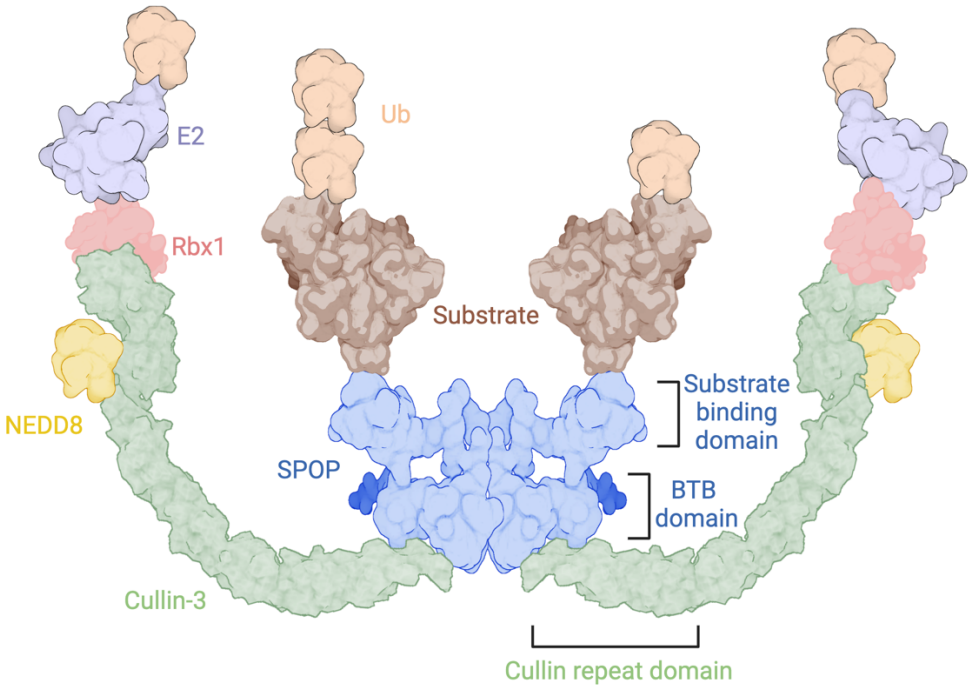


Figure 1.5 – SPOP-Cullin-3-Rbx1 ligase complex architecture
Ub is ubiquitin. Cullin-3 binds to the BTB-domain containing substrate adaptor, SPOP, via its N-terminal Cullin repeat domain. The substrate adaptor recruits substrate to the ligase complex via its substrate binding domain. At the opposite end of Cullin-3, RING domain-containing Rbx1 recruits Ub-E2 conjugate to the ligase complex. The ligase complex is NEDDylated which activates its ligase activity. The active Cullin-3-Rbx1 E3 complex facilitates the direct transfer of Ub from the E2 onto target substrate.

1.4.1 - SPOP in cancer

SPOP is frequently downregulated or mutated in several cancers such as prostate and endometrial cancer, and is overexpressed in clear cell renal cell carcinoma [56]. The most prevalent SPOP cancer-associated

mutations are found in the substrate binding cleft of the MATH domain (F-133, F-102, W-131, K-129, Y-87), preventing substrate binding and ubiquitylation, ultimately leading to the stabilisation of substrate and dysregulation of their various pathways (**Fig. 1.6 a, b**) [57, 58].

SPOP is mutated in between 4.4-28.6% of prostate cancer cases, as well as being downregulated in between 25.2-93.5% of cases, leading to dysregulation of its target substrates and resulting in tumorigenic effects [56]. One of the most well-studied SPOP-prostate cancer pathways is that of the steroid receptor coactivator-3 (SRC-3)/androgen receptor (AR) pathway [59, 60]. In response to low androgen levels, such as the hormone testosterone, SRC-3 associates with AR leading to a signalling cascade through the mitogen-activated protein kinase (MAPK) pathway, which regulates gene expression of cell growth and proliferation factors. When SPOP is mutated, it is unable to ubiquitylate SRC-3 and AR resulting in their stabilisation, which in turn leads to increased cell growth and survival signalling, further promoting tumorigenesis.

1.4.6 - SPOP substrates

SPOP regulates proteins involved in a variety of biological processes, particularly in genomic maintenance (transcription regulation, chromatin remodelling and DNA damage repair) as well as in various intracellular signalling pathways (immune and stress signalling) (**Table 1.1**). The large repertoire of substrates prompt SPOP to have both a nuclear and a cytoplasmic localisation, depending on the substrate. Additionally, some substrates have been shown to undergo liquid-liquid phase separation with SPOP, which has been suggested to help enhance ubiquitylation of substrates by increasing the local concentration of ligase and substrate [61].

Protein	Function or role	SBC sequence	UniProt ID	Ref.
MyD88_1	Immune signalling	VSSTS	Q99836	[62]
MyD88_2	“	VDSSV	“	[62]
Pdx1	Regulation of insulin gene transcription	VTSGE	P52945	[63]
Puc	JNK pathway phosphatase	VTSTT	O46122	[55]
Cl_1	Transcriptional activator and repressor of the hedgehog pathway	PSSTS	P19538	[64]
Cl_2	“	VSSST	“	[64]
GEMININ	DNA replication inhibitor	VSSST	O75496	[65]
TP53BP1	Double-strand DNA break repair protein	ASSSS	Q12888	[66]
BRD3	Transcription regulation	ADTTT	Q15059	[67]
MacroH2A 1	Transcription regulation	ADSTT	O75367	[55]
DAXX_2	Transcription regulation	ISSSR	Q9UER7	[61]
DAXX_3	“	VSSTS	“	[55]
DAXX_4	“	ADSST	“	[55]
DAXX_5	“	VTSSL	“	[61]
GLI2_1	Transcription regulation in hedgehog pathway	VSSGT	P10070	[68]
GLI2_2 (?)	“	VSSTV	“	[68]
GLI2_3 (?)	“	ASSTS	“	[68]
GLI2_4 (?)	“	VSSTV	“	[68]
GLI3_1	“	VTSTV	P10071	[69]
GLI3_2	“	ASSTT	“	[69]
GLI3_3	“	PSTSS	“	[69]
DEK	Chromatin organisation	ADSST	P35659	[70]
BRD4	Transcription regulation	ADTTT	O60885	[71]
NCOA3	Nuclear receptor binding and translation regulation	VSSTG	Q9Y6Q9	[59]
DDIT3	Transcription factor in ER stress response	VTSTS	P35638	[72]
ERG	Transcription regulation	VTSSS	P11308	[73]
SENP7	Protease in cGAS-STING pathway	LSSSS	Q9BQF6	[74]
SETD2	Histone methyltransferase	ISSNS	Q9BYW2	[75]
CDC20	Cell division regulator	GKSSS	Q12834	[76]
SIRT2 (?)	NAD-dependent protein deacetylase	ISTSA	Q8IXJ6	[77]
EGLN2_1	Prolyl hydroxylase	GSSSE	Q96KS0	[78]
EGLN2_2	“	ATSTT	“	[78]
C-MYC_1	Transcription factor	VCSTS	P01106	[79]
C-MYC_2	“	PTTSS	“	[79]
INF2	Actin polymerization	ADSTS	Q27J81	[80]
HDAC6	Histone deacetylase	PSSSK	Q9UBN7	[81]
AR	Androgen receptor	ASSTT	P10275	[82]
ERα	Nuclear hormone receptor	GSTSS	P03372	[83]
HIPK2_1	Serine/threonine-protein kinase involved in transcription regulation	ASSTS	Q9H2X6	[84]

HIPK2_2	“	ASSTT	“	[84]
BRD2	Chromatin binding	ADTTT	P25440	[71]
NANOG	Transcription regulator in embryonic stem cells	PDSST	Q9H9S0	[85]
ATF2_1	Transcription activator	PTSST	P15336	[86]
ATF2_2	“	ATSTT	“	[86]
FASN_1	Biosynthesis of long-chain saturated fatty acids	ACSSS	A0A0U1 RQF0	[87]
FASN_2	“	LDSTS	“	[87]
CAPRIN1	Regulator of mRNA transport, translation and stability	VSSTS	Q14444	[88]
CDCA5	Regulator of sister chromatid cohesion in mitosis	AESSS	Q96FF9	[89]
PDPK1	Serine/threonine master kinase	VSSSS	O15530	[90]
DHX9	Nucleic acid helicase	PWTSS	Q08211	[91]
BRMS1	Transcriptional repressor	GSSRS	Q9HCU9	[92]
FLI-1	Transcriptional activator	VTSSS	Q01543	[93]

Table 1.1 – List of known SPOP substrates
SBC is SPOP binding consensus sequence. Proteins with more than one SBC are numbered (i.e., _1) Proteins with a (?) are known SPOP binders but with an unknown SBC sequence. A predicted SBC based on sequence similarity is suggested. Ref. is reference.

In the SPOP-Cullin-3 ligase complex, Rbx1 has been shown to recruit E2 enzymes capable of forming K-48 ubiquitin linkages. In addition, SPOP was shown to ubiquitylate PrLZ using K-48 linkages and HIPK2 and INF2 using K-27, K-29, K-33, K48 and K-63 linkages, with the non-K48 linkages likely occurring due to the formation of mixed-linkage chains formed by other ligases [80, 84, 94].

1.4.2 - Domain architecture of SPOP

SPOP contains an N-terminal Meprin and TRAF homology (MATH) domain (residues 28-166), a central BTB dimerisation domain (residues 177-296), a C-terminal domain (CTD) dimerisation domain (residues 298-359) and a nuclear localisation sequence (residues 365-374) [55, 84, 95] (**Fig. 1.6 a, b**).

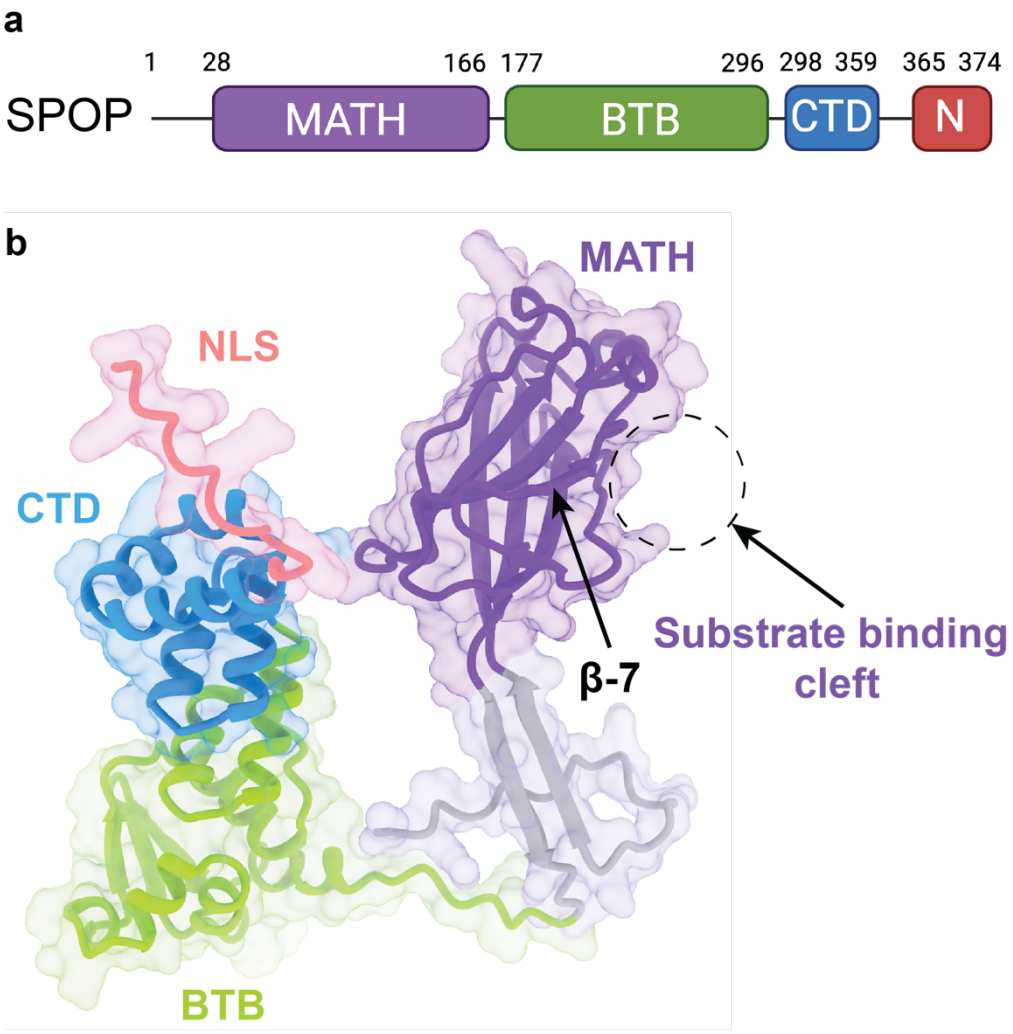


Figure 1.6 – SPOP protein domains and AlphaFold predicted structure of SPOP
a Domain organisation of SPOP protein. MATH is Meprin and TRAF homology domain. BTB is Broad complex, Tramtrack, and Bric-à-brac domain. CTD is C-terminal domain. N is nuclear localisation signal.
b AlphaFold [96] structure prediction model of SPOP, with protein domains highlighted as in **a**. NLS is nuclear localisation signal. The MATH domain substrate binding cleft is indicated with a black arrow.

1.4.3 - MATH domain

The MATH domain is a substrate recruiting domain that binds to serine and threonine-rich sequences on substrates [55]. SPOP substrates contain a five-residue SPOP binding consensus (SBC) motif composed of φ-π-S-S/T-S/T (φ is a nonpolar residue; π is a polar residue) [55]. These SBCs are typically found in intrinsically disordered regions of the substrate, with substrates typically containing multiple low affinity

degrons within their sequence for multi-valent binding to multiple SPOP MATH domains [55].

The *Drosophila melanogaster* phosphatase Puckered (Puc) is a regulator of the JNK signalling pathway and is a substrate of SPOP [97]. To date, Puc has the highest known affinity of binding to SPOP [55]. The binding affinity of a Puc SBC peptide (^{93}LACDE**VTSTT**SSSTA^{107}) to the MATH domain of SPOP was measured with a K_d of 3.70 ± 0.03 μM, whereas the majority of SBCs have affinities in the mid to high μM range [55]. A crystal structure of a Puc peptide bound to the SPOP MATH domain revealed Puc's SBC (^{98}VTSTT^{102}) bound to a central cleft of the MATH domain, proximal to β-7 (**Fig. 1.6 b, Fig. 1.7**). This hydrophobic cleft contains several critical residues for substrate binding (Y-87, F-102, Y-123, W-131, and F-133) (**Fig. 1.7**). Additionally, SPOP residues K-129 and E-130 hydrogen bond with the well conserved serine at position 3 of the SBC of the substrate (**Fig. 1.7**). These SPOP residues are frequently mutated in cancer, preventing substrate binding, leading to accumulation of substrate and dysregulation of critical intracellular processes.

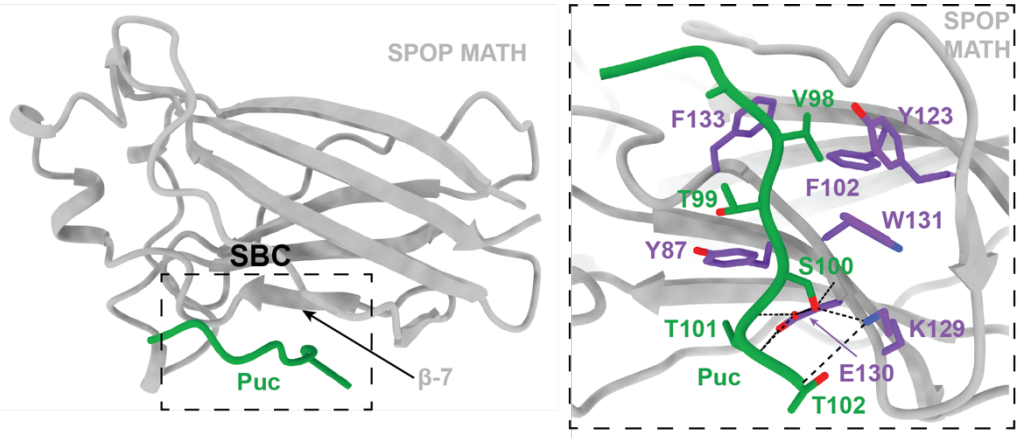


Figure 1.7 – Puckered (Puc) peptide bound to the SPOP MATH domain
Left shows an overview of Puc peptide (PDB ID 3HQL) [55] bound near beta-strand 7 of SPOP MATH. Right shows a zoomed in view of the interactions between the Puc SPOP binding consensus (SBC) sequence and SPOP MATH. Key SPOP residues involved in substrate binding are shown in purple. Hydrogen bonds are shown as black dashed lines.

Another SPOP substrate, Pancreatic duodenal homeobox 1 (Pdx1), is a transcription factor that regulates insulin gene expression and β-cell survival [98]. The mode of Pdx1 binding to SPOP is unusual in comparison to other SPOP-substrate structures, as additional residues other than the SBC are also involved in SPOP binding (**Fig. 1.8**) [99].

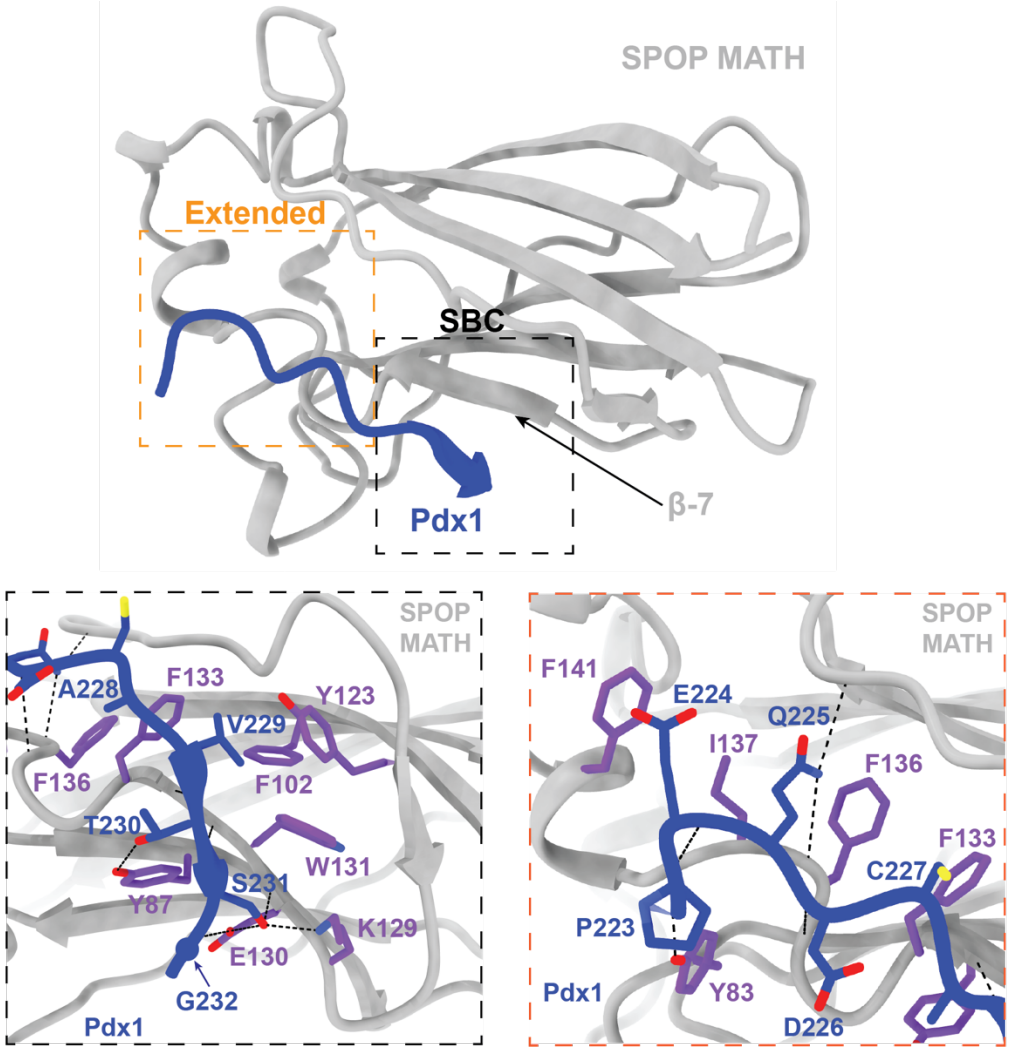


Figure 1.8 - Pancreatic duodenal homeobox 1 (Pdx1) peptide bound to the SPOP MATH domain

Top shows an overview of a Pdx1 peptide (PDB ID 6F8F) [99] bound near beta-strand 7 of the SPOP MATH domain. Bottom left shows a zoomed in view of the interactions between the Pdx1 SPOP binding consensus (SBC) sequence and SPOP MATH. Key SPOP residues involved in substrate binding are shown in purple. Hydrogen bonds are shown as black dashed lines. Bottom right shows the extended binding region of the Pdx1 peptide.

1.4.4 - BTB domain

The BTB domains from two SPOP monomers come together to form homodimers via a hydrophobic interface formed of BTB α-helices 1, 2 and 5 (**Fig. 1.9**) [55]. In addition, the BTB domain also mediates the interaction with the Cullin-3 N-terminal Cullin repeat domain via BTB α-helices 3 and 4 and β-strand 5 [95]. Mutations at the dimerisation interface (L186D, L190D, L193D and I217K) prevented dimer formation but did not reduce the binding affinity to Cullin-3 [55]. Despite maintaining a WT-level binding affinity to Cullin-3, the monomeric SPOP showed reduced ubiquitylation efficiency, suggesting that a higher-order assembly of the SPOP-Cullin-3 complex is needed for efficient ubiquitylation [55].

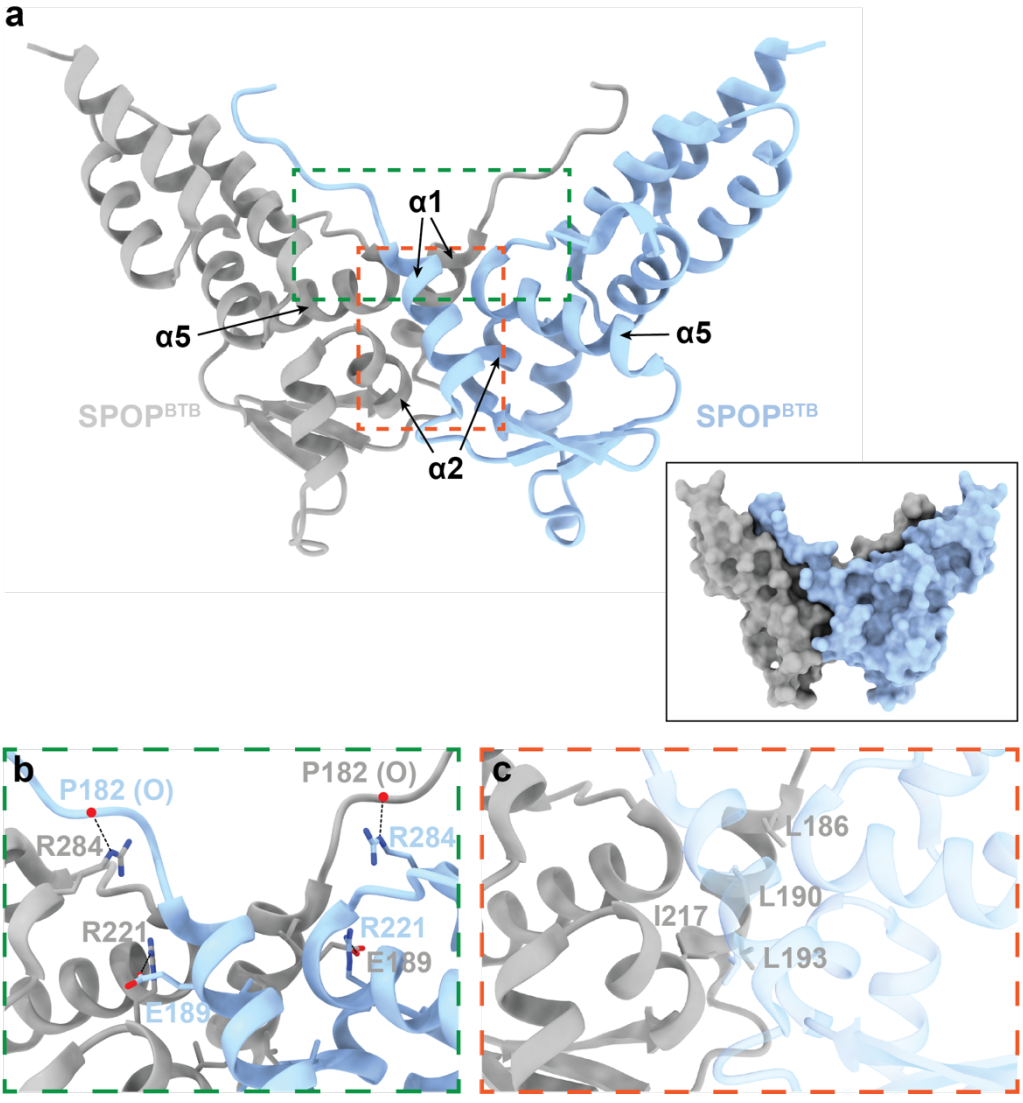


Figure 1.9 – Structure of the SPOP BTB domain dimer

a Structure of the SPOP BTB dimer (PDB ID 3HTM) [55] highlighting α-helices 1, 2 and 5 which are involved in dimerisation. Top left is a cartoon representation and bottom right is a surface representation of the same view. Dashed line boxes show locations of zoomed in views shown in **b** and **c**.
b Zoomed-in view of **a** showing hydrogen bonding residues of the BTB dimer. Hydrogen bond represented as black dashed line.
c Zoomed-in view of **a** showing hydrophobic residues of the BTB dimer. Residues for one SPOP BTB molecule are labelled but the same residues are also involved in dimer formation in the second SPOP BTB molecule.

1.4.5 - CTD domain

The CTD is an additional homo-dimerisation domain that acts in tandem to the BTB domain. A 1.5 Å crystal structure of a $SPOP^{270-374}$ dimer which included the CTD domain, NLS and part of the BTB domain showed a dimerisation interface across the CTD domains, predominantly between the α11 helices [100]. This interface buried hydrophobic residues V-349, A-350, L-356 and A-357, with hydrogen bonding observed between Y-353 and N-326 (**Fig. 1.10**) [100].

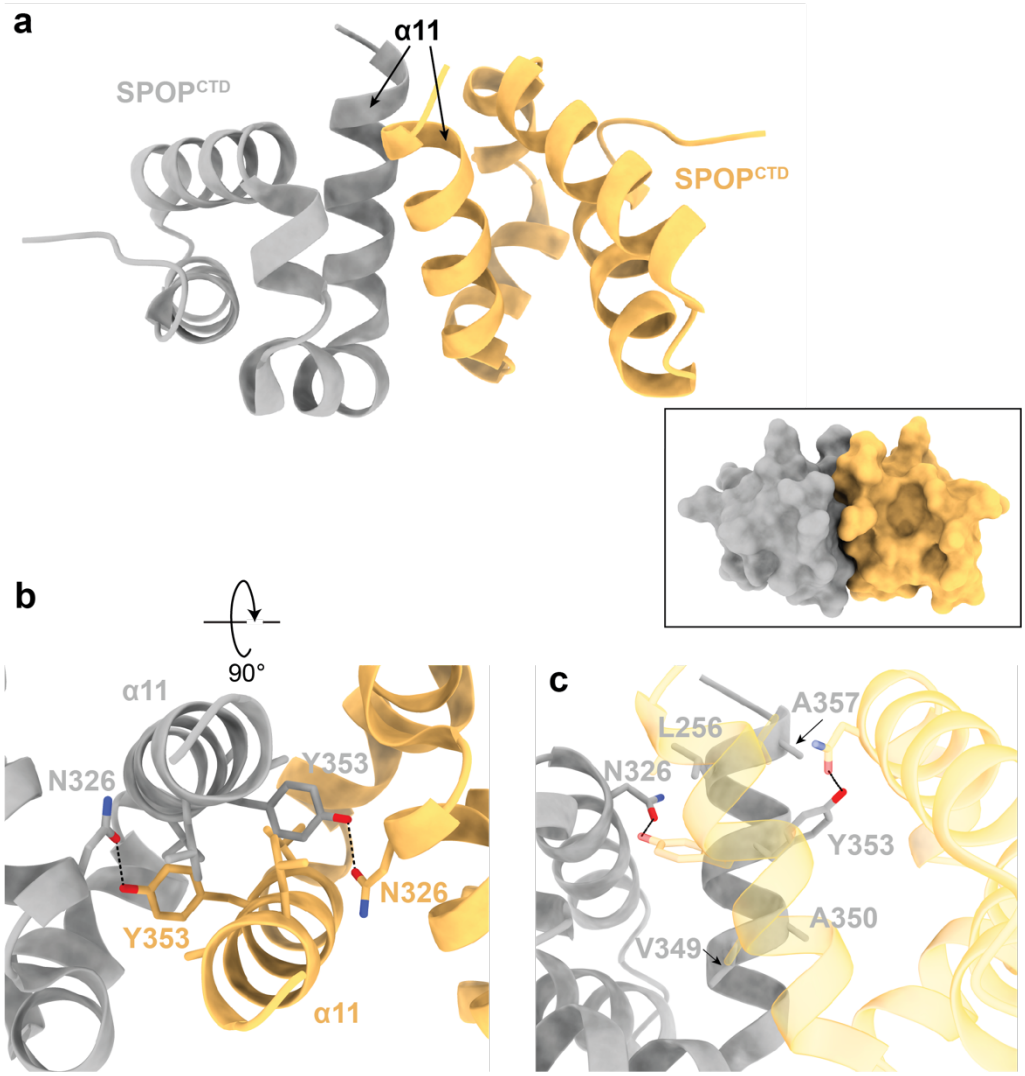


Figure 1.10 – Structure of the SPOP CTD dimer
a Structure of the SPOP CTD dimer (PDB ID 4HS2) [100] highlighting α-helix 11, which is involved in CTD dimerisation. Top left is a cartoon representation and bottom right is a surface representation of the same view.
b Zoomed-in view of **a** rotated 90 degrees horizontally showing hydrogen bonding residues across the CTD dimer. Hydrogen bond represented as black dashed line.
c Zoomed-in view of **a** showing dimerisation residues across the CTD dimer. Residues for one SPOP CTD molecule are labelled but the same residues are also involved in dimer formation in the second SPOP CTD molecule.

A truncation of SPOP which included both the BTB and CTD dimerisation domains ($SPOP^{169-374}$) showed the protein undergoing high-order oligomerisation, which was abolished when mutating Y353E in the CTD, which showed only dimer formation [100]. This dimerisation ability of the BTB and CTD domains allow for the high-order oligomerisation of SPOP, where BTB dimers and CTD dimers form a head-to-head, tail-to-tail

arrangement (**Fig. 1.11 a**). This high-order oligomerisation of SPOP greatly enhances the ubiquitylation efficiency of substrate by allowing the recruitment of multiple substrate SBCs to the SPOP oligomer [69, 95].

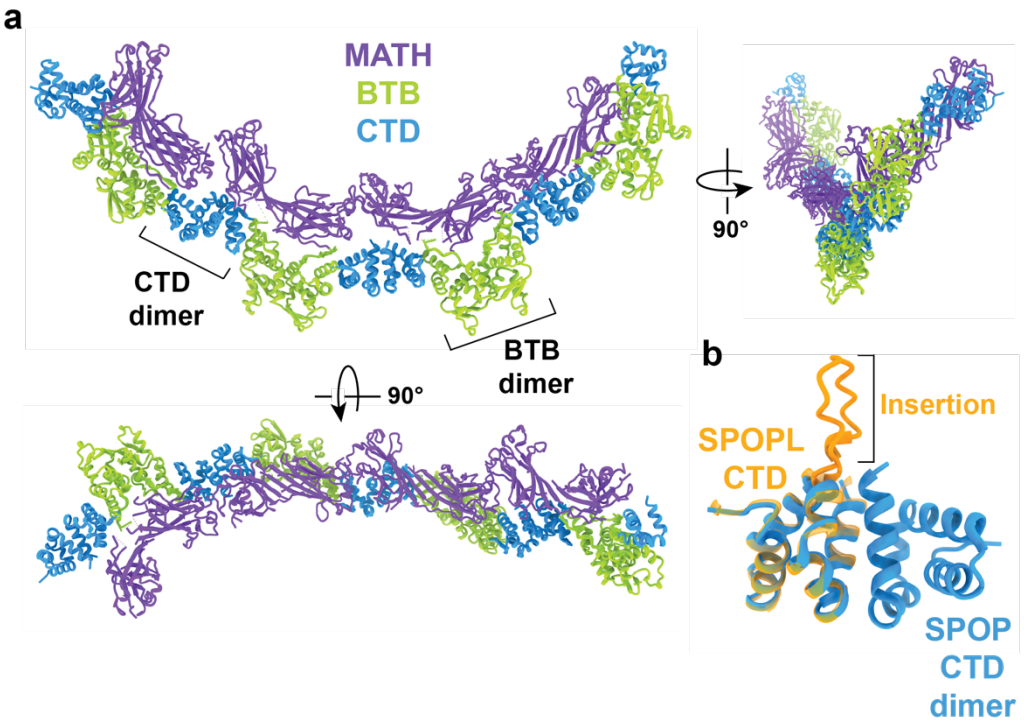


Figure 1.11 - Predicted SPOP oligomer structure
a Predicted structural model of the SPOP oligomer [100] made by superimposing overlapping crystal structures of SPOP domain truncations. MATH is Meprin and TRAF homology domain. BTB is Broad complex, Tramtrack, and Bric-à-brac domain. CTD is C-terminal domain.
b Alphafold model of the SPOPL CTD superimposed onto the structure of the SPOP CTD dimer (PDB ID 4HS2) [100]. An 18 amino acid insertion in the SPOPL CTD is indicated.

A paralog of SPOP called speckle-type POZ protein-like (SPOPL) shares 81% sequence similarity with SPOP. The MATH and BTB domains are well conserved but SPOPL contains an 18 amino acid insertion in its CTD, preventing high-order oligomerisation (**Fig. 1.11 b**) [95]. Both SPOP and SPOPL are substrate adaptors for Cullin-3-RING E3 ligase, but act upon different substrate proteins. In particular, SPOPL targets EPS15 at endosomes which assists in endocytic trafficking [101]. SPOPL is also thought to act as a SPOP inhibitor. SPOPL is unable to form high-order oligomers and is limited to dimer formation only. Due to the high

sequence similarity, SPOPL is able to be incorporated into SPOP oligomers, effectively capping the SPOP oligomer and reducing the oligomer length [95]. This in turn limits the number of substrates that the SPOP oligomer can recruit, reducing the ubiquitylation efficiency of target substrates.

1.5 - Immune signalling by the Myddosome complex

Tissue damage and infection results in a response in the bone marrow termed emergency haematopoiesis, resulting in rapid proliferation of haematopoietic stem cells (HSCs) and cell differentiation into the myeloid lineage [102]. Emergency haematopoiesis is initiated through pattern-associated molecular patterns (PAMPs) or cytokines IL-1 and IL-18 binding to Toll-like receptors (TLRs) or Interleukin-1 receptor (IL1-R) on the cell surface of HSCs [103, 104]. IL1-R is a member of the TLR family, which are pattern recognition receptors. They have a conserved cytoplasmic Toll/Interleukin-1 receptor (TIR) domain and leucine rich extracellular domain, which bind PAMPs [105]. Binding of cytokines or PAMPs to the extracellular domain triggers a receptor conformational change and dimerization for the recruitment of downstream immune signalling effectors. After resolution of injury or infection, the termination of emergency haematopoiesis is crucial to ensure homeostasis and to allow differentiation of HSCs into other blood cell lineages. A failure to regulate emergency haematopoiesis can result in chronic inflammation, autoimmune disease and cancer.

1.5.1 - The Myddosome complex

All TLRs, with the exception of TLR3, transduce their signal through a large, multi-protein complex called the Myddosome [106, 107]. The Myddosome complex comprises of the 33 kDa Myeloid differentiation primary response 88 (MyD88) protein together with IL1-R-associated kinases (IRAKs); IRAK4 and either IRAK1 or IRAK2 (52 kDa, 77 kDa and

69 kDa respectively). IRAK4 is required for Myddosome assembly; however, the Myddosome complex can form interchangeably with either IRAK1 or IRAK2 protomers.

MyD88 contains an N-terminal Death domain (DD), which binds to IRAKs (via their DDs) and a C-terminal toll-interleukin-1 receptor (TIR) domain, which associates with TLRs (**Fig. 1.12 a, b**) [106]. IRAK1, IRAK2 and IRAK4 contain an N-terminal DD and a C-terminal serine/threonine kinase domain (KD) (**Fig. 1.12 a, b**).

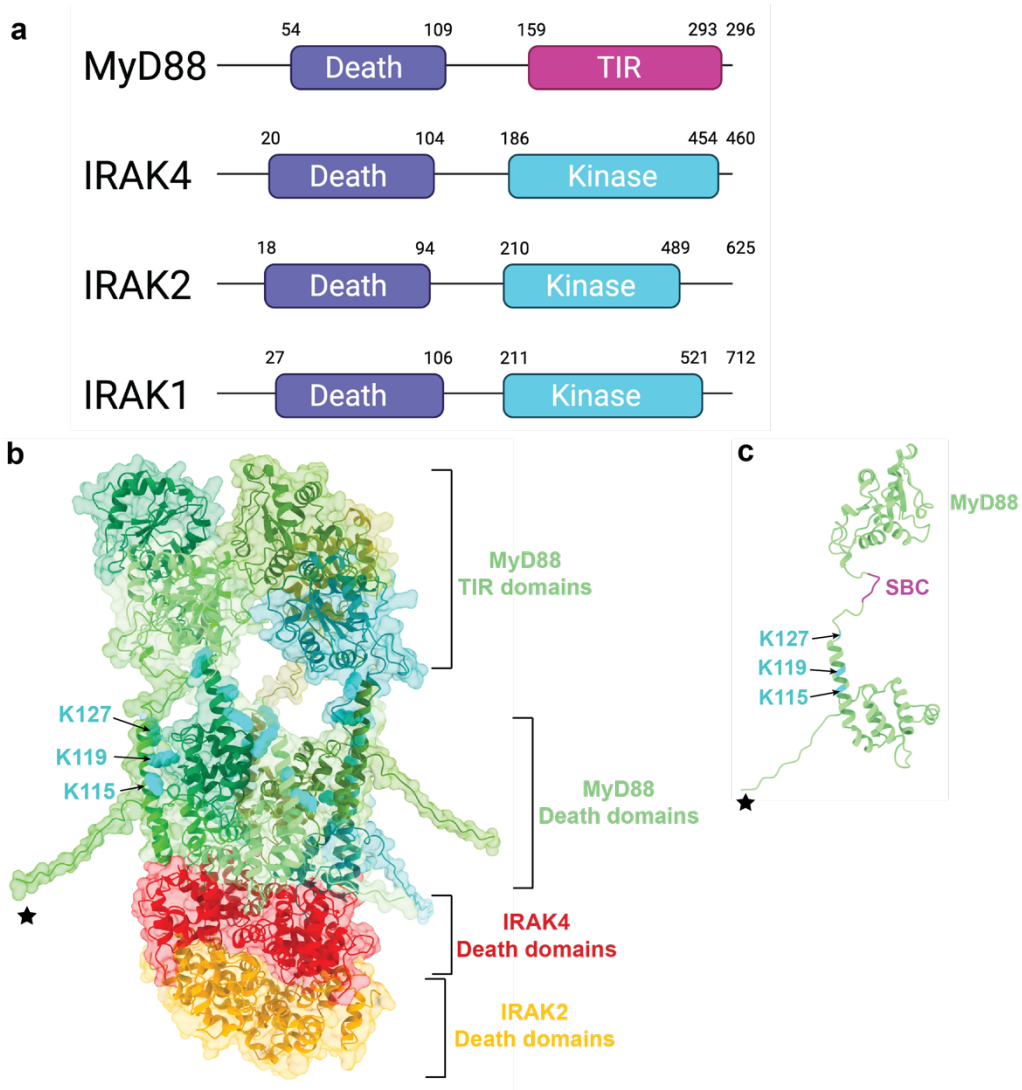


Figure 1.12 – Schematic of Myddosome protein domain organisation and composite model of the Myddosome
a Domains of the Myddosome-complex proteins. Amino acid numbers are indicated. TIR is Toll-like-Interleukin-2 receptor domain.

b Structure of the Death domain only Myddosome (PDB ID 3MOP) [108] which includes 6 MyD88 Death domains, 4 IRAK4 Death domains and 4 IRAK2 Death domains. AlphaFold [96] model of full-length MyD88 is superimposed onto the Myddosome by aligning the Death domains. Putative MyD88 lysines which may be ubiquitylated by SPOP are shown as spheres in cyan. Kinase domains of IRAK4 and IRAK2 are not shown.
c AlphaFold model of a single protomer of MyD88. Black star indicates the same MyD88 protomer as in **b**. The SPOP-binding consensus (SBC) sequence is located in an unstructured region between the Death and TIR domains, proximal to the potential ubiquitylations sites.

1.5.2 - The Myddosome signalling pathway

The Myddosome serves as a signalling relay between activated TLRs and downstream effector molecules (**Fig. 1.13**). To initiate Myddosome signalling, MyD88 binds to the cytoplasmic TIR domain of IL-1R after receptor activation and dimerisation, resulting in the oligomerisation of MyD88 [109]. MyD88 is able to oligomerise through both DD and TIR domain interactions [110]. Once assembled at the TLR, MyD88 recruits IRAK4 [106]. IRAK4 has intrinsic kinase activity and undergoes trans-autophosphorylation of T-342, T-345, and S-346 in its activation loop, which is greatly increased upon DD association with MyD88 [111]. The KD of IRAK4 has also been shown to dimerise, which in turn has been shown to increase the autophosphorylation activity of IRAK4 [112]. Once associated with IRAK4 at the Myddosome, the IRAK2 or IRAK1 activation loop is then phosphorylated by IRAK4 to activate them [113].

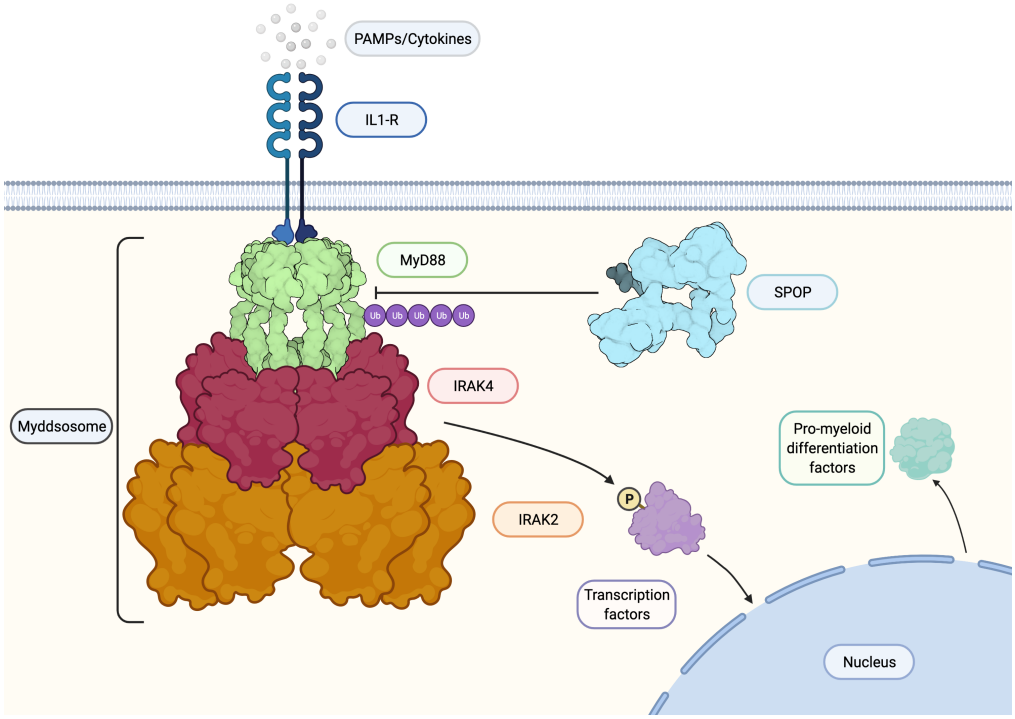


Figure 1.13 – The Myddosome immune signalling pathway
The Myddosome complex (MyD88, IRAK4 and IRAK2) assembles at the interleukin-1 receptor (IL1-R) in response to sensing cytokines of pattern-associated molecular patterns (PAMPs). After Myddosome activation, IRAK4 phosphorylates downstream effector molecules, eventually leading to the transcription of pro-inflammation factors. The E3 ligase adaptor SPOP negatively regulates Myddosome signalling by ubiquitylating MyD88 at the Myddosome. Ub is ubiquitin. P is phosphoryl group. Figure made in biorender.com.

Once the Myddosome is fully assembled at the activated TLR, IRAK4 then phosphorylates various transcription factors such as PU.1, C/EBPβ and C/EBPα, which eventually results in the transcription of pro-myeloid differentiation factors [114]. This Myddosome signalling cascade also occurs in terminally differentiated immune cells to activate the pro-inflammatory response, via the activation of the E3 ubiquitin ligase, tumour necrosis factor receptor-associated factor 6 (TRAF6) [115]. TRAF6 then activates transcription factors, such as NF-κB and AP-1, which leads to the transcription of pro-inflammation genes such as inflammatory cytokines.

1.5.3 - SPOP ubiquitylates MyD88

MyD88 has been shown to be regulated by the ubiquitin system. The HECT ligase Smurf K-48 poly-ubiquitylates MyD88 in response to the anti-inflammatory cytokine TGF-β, resulting in MyD88s degradation [116]. Additionally, MyD88 is K-63 poly-ubiquitylated by an unknown ligase, which further promotes immune signalling. The DUB CYLD acts as a negative regulator of immune signalling by cleaving K-63-linked chains from MyD88 [117].

SPOP was recently shown to terminate Myddosome signalling through the K-48-linked ubiquitylation of MyD88 [62, 118]. A MyD88 SBC was identified as ^{134}VDSSV^{138}, which is situated in a predicted unstructured region between the DD and TIR domain (**Fig. 1.12 c**) [62]. This SBC is likely sub-optimal for SPOP binding as position 5 diverges from the consensus, with an uncharged valine in place of a small, polar serine or threonine. Similar to other SBCs, this MyD88 SBC was shown to be sensitive to phosphorylation, with a complete inhibition of SPOP binding when both S-136 and S-137 are phosphorylated; however, the enzymes that regulate the phosphorylation state of the MyD88 SBC are currently unknown [55, 62]. A study of chicken SPOP and MyD88 found that MyD88 degradation was promoted by SPOP-mediated ubiquitylation on three MyD88 lysine residues; K-118, K-124 and K-143 [118]. Both K-118 (K-115 in humans) and K-124 (K-119 or K-127 in humans) are structurally conserved but K-143 is not, suggesting that other lysine residues may be modified on human MyD88 or that K-115, K-119 and K-127 are sufficient for function (**Fig. 1.12 c**).

Since the interaction between SPOP and MyD88 was initially identified in 2019, several papers have confirmed this interaction, each with slightly different conclusions as to how SPOP interacts with MyD88 to terminate Myddosome signalling [62, 118-120]. Guillamot *et al.* suggest that SPOP-mediated ubiquitylation of MyD88 likely occurs at the level of the Myddosome to terminate Myddosome signalling [62]. SPOP-mediated

ubiquitylation of MyD88 was increased upon co-expression of IRAK4, suggesting that ubiquitylation of MyD88 occurs at the Myddosome complex level rather than of free MyD88 [62].

Li *et al.* found that overexpression of chicken SPOP promoted the ubiquitylation and proteasomal degradation of chicken MyD88 [118]. In macrophage cells, decreased SPOP expression and ubiquitylation site mutants of MyD88 led to increased cytokine production when TLR/Myddosome signalling was activated with lipopolysaccharide, suggesting that SPOP-mediated ubiquitylation of MyD88 results in negative regulation of Myddosome signalling [118].

Hu *et al.* report that SPOP prevents the self-association of MyD88 in response to TLR activation; however, they did not report any detectable ubiquitylation of MyD88 after SPOP overexpression [120]. However, TLR4 activation was shown to cause the translocation of SPOP from the nucleus to the cytoplasm, perhaps to target the Myddosome pathway [120]. Under knockdown of SPOP and stimulation of TLR signalling, MyD88 aggregation was increased in comparison to when SPOP was expressed, suggesting that SPOP regulates Myddosome assembly during TLR signalling [120].

Jin *et al.* suggested that SPOP induces the ubiquitylation of MyD88 however, this ubiquitylation was non-degradative and prevented Myddosome signalling by blocking Myddosome assembly [119].

These differing reports all agree on SPOP's interaction with MyD88, although the exact mechanism and outcome of this interaction requires further work. Gaining the structure of the SPOP-MyD88 complex may reveal SPOP's mechanism of ubiquitylation and how this might relate to the Myddosome complex assembly.

1.6 - UFMylation

UFMylation is a post-translation modification present in metazoans. It is similar to ubiquitylation, where ubiquitin-fold modifier 1 (UFM1) is covalently attached to lysine residues on target proteins, either as a monomer or as a lysine-linked polymer [121]. The UFM1 molecule shares a similar size and β-grasp fold to ubiquitin, despite having little amino acid sequence similarity [122]. Like ubiquitylation, UFMylation occurs through a three-step enzymatic reaction involving the E1 activating enzyme, ubiquitin-like modifier activating enzyme 5 (UBA5), the E2 conjugating enzyme, ubiquitin-fold modifier conjugating enzyme 1 (UFC1) and an E3 ligase enzyme, UFM1-specific ligase 1 (UFL1) (**Fig. 1.14**) [123]. UFM1 has six lysines (K-3, K-7, K-19, K-34, K-41 and K-69) with only K-69 having been shown to form poly-UFM1 chains [124]. Two de-UFMylases have been identified UFM1-specific proteases 1 and 2 (UFSP1 and UFSP2), which can remove UFM1 modifications from target proteins.

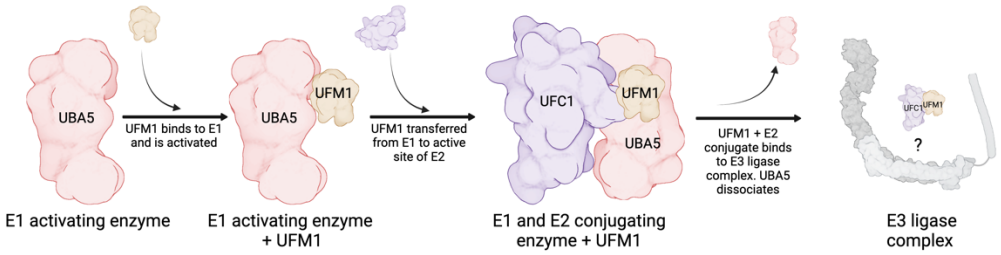


Figure 1.14 – Schematic of the UFMylation enzyme cascade
Pro-UFM1 is processed by the de-UFMylase UFSP2 to remove the last two amino acids (Ser-84 and Cys-85), generating mature UFM1 with a C-terminal Gly-83. Mature UFM1 is activated by the E1 enzyme UBA5 by forming a thioester bond between UFM1 Gly-83 and the catalytic Cys-250 of UBA5. UFM1 is then transferred to the active site Cys-116 of the E2 enzyme, UFC1. The UFM1-E2 conjugate is then recruited to the E3 ligase UFL1 for UFMylation of target proteins.

UFMylation of target proteins is typically involved in non-proteolytic signalling, such as protein complex assembly and functional activation. UFMylation occurs to maintain ER homeostasis during haematopoiesis and immune signalling [125, 126]. Additionally, an abnormal UFM1 cascade

results in ER stress and is associated diseases such as cancer, ischemic heart disease, diabetes, atherosclerosis and schizophrenia, suggesting a fundamental role for UFMylation in cell growth and development [127-129].

1.7 - The UFL1 E3 ligase complex

UFMylation of target proteins is mediated by a multi-protein E3 ligase complex called the UFL1 ligase complex. This complex consists of UFL1, UFM1 binding protein 1 (UFBP1; also known as DDRGK1) and cyclin-dependent kinase 5 regulatory subunit-associated protein 3 (CDK5RAP3).

The 90 kDa protein UFL1 is the only known UFM1 E3 ligase [130]. This is an unusual E3 ligase as it does not share any domain homology with the RING, U-box and HECT classes of ubiquitin E3 ligases. UFL1 forms a complex with UFBP1, with an AlphaFold prediction showing that the N-terminal region of UFL1 binds to the C-terminus of UFBP1 [131]. The 36 kDa protein UFBP1 acts as a UFL1 adaptor which localises the entire ligase complex to the ER membrane via an N-terminal transmembrane helix (TMH) [132]. UFBP1 also contains a nuclear localisation sequence, allowing the ligase complex to also target nuclear substrates. UFBP1 is essential for UFL1 ligase activity, which is thought to UFMylate substrates using a scaffold-type E3 ligase mechanism, as no catalytic cysteine has been found [131]. Lastly, the 53 kDa protein CDK5RAP3 forms a ternary complex with UFL1 and UFBP1. CDK5RAP3 was initially thought to act as a inhibitor of ligase activity but was shown to be required for ligase specificity towards the ribosome [131].

There is currently little structural information regarding the UFM1 enzymes and complexes, with the exception of the UBA5 E1 enzyme and the UFC1 E2 conjugating enzyme, which broadly resemble the ubiquitin enzymes [133]. The full-length or partial structures of any of the

E3 ligase components are yet to be elucidated, with the exception of a short UFBP1 peptide bound to UFM1 [134].

Known UFMylation substrates include activating signal cointegrator (ASC1), a transcriptional co-activator of estrogen receptor-alpha (ERα), as well as ERα itself, which increases gene expression of ERα and stability of ERα respectively [124, 129]. Additionally, Histone H4 is UFMylated during the DNA damage repair pathway [135]. p53 UFMylation stabilises it by preventing is ubiquitylation and degradation, promoting its tumour-suppressive activity [136]. Recently the ribosome protein RPL26 was shown to UFMylated in a pathway implication in ribosome quality control [137-139].

1.8 - Ribosome protein translation

The ribosome is a large multi-subunit ribozyme present in most cells that synthesizes new polypeptides in a process termed translation. It is composed of approximately 60% ribosome RNA (rRNA) and 40% ribosome proteins, with two main subunits: a large subunit (60S subunit in humans) and a small subunit (40S subunit in humans), which come together to form a ribosome (80S in humans) (**Fig. 1.15**). Translation occurs in three main steps: initiation, where the ribosome subunits assemble on the messenger RNA (mRNA), elongation, where transfer RNAs (tRNAs) bring amino acids to the ribosome for polypeptide chain formation, and termination, where the newly formed polypeptide chain is released from the ribosome.

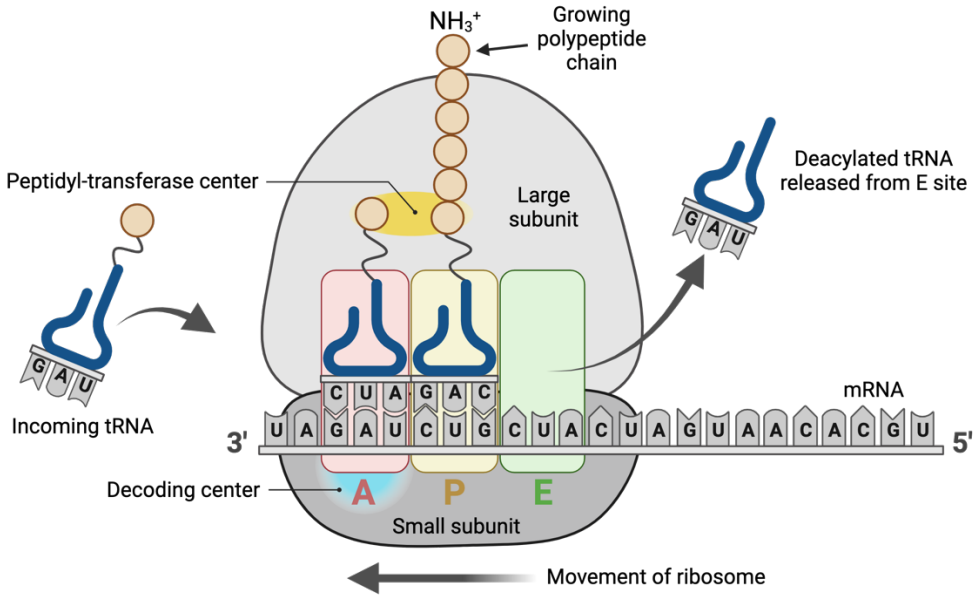


Figure 1.15 – Schematic of a ribosome during protein translation and elongation tRNA is transfer RNA. A is aminoacyl-tRNA binding site. P is peptidyl-tRNA binding site. E is exit-tRNA binding site. Adapted from "Translation: Elongation Phase (EPA Site)", by BioRender.com (2023). Retrieved from https://app.biorender.com/biorender-templates.

1.8.1 - Initiation

In eukaryotes, protein translation is initiated through the action of family of proteins called eukaryotic initiation factors (eIF). First, eIF2 brings a methionine-tRNA to the P-site of the 40S subunit [140]. This tRNA-40S complex is then recruited to the 5' end of mRNA by eIF3 and eIF4 factors, where the 40S reads through the mRNA from 5' to 3' in its decoding centre [141-143]. After recognition of the AUG start codon in the 40S decoding centre though base pairing of the anticodon loop of the methionine-tRNA, eIF2 hydrolyses GTP to release from the complex (**Fig. 1.16**) [144, 145]. Next, eIF1A and eIF5B recruits the 60S subunit [146, 147]. The 60S binds, with the methionine and tRNA acceptor stem sitting in the P-site (peptidyl tRNA; P-tRNA) of the large subunit (**Fig. 1.15**). eIF5B then hydrolyses GTP to release all factors from the fully assembled 80S ribosome [148]. The ribosome can now begin translating the mRNA sequence.

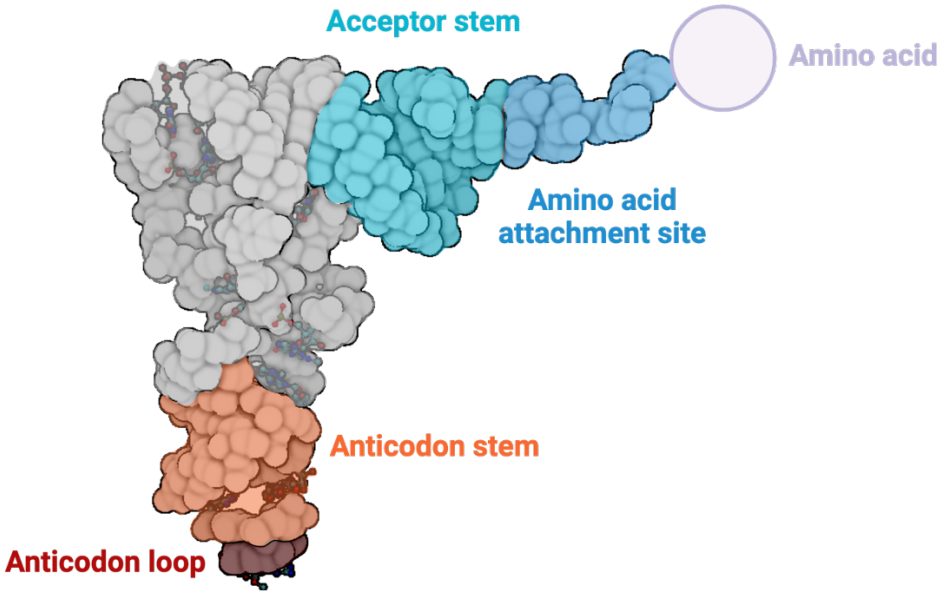


Figure 1.16 – Features of an aminoacyl-transfer RNA (tRNA)
Surface representation of a tRNA molecular (PDB ID: 4TNA) [149]. Key sites involved in protein translation are highlighted. Orange are 40S subunit binding. Blue are 60S subunit binding. Created with BioRender.com.

1.8.2 - Elongation

Protein chain elongation is assisted by eukaryotic elongation factors (eEF). An aminoacyl-tRNA (A-tRNA) molecule is brought to the A-site of the ribosome with eEF1A [150]. The A-tRNA base pairs with the second mRNA codon, leading to GTP hydrolysis and dissociation of eEF1A. The A-tRNA is now fully situated in the A-site with the amino acid sat in the peptidyl-transferase centre (PTC) of the 60S subunit [151, 152]. The PTC is a site composed entirely of rRNA that catalyses critical protein translation processes, such as peptide bond formation and peptide release (**Fig. 1.15**). Peptide bond formation occurs by transfer of the amino acid situated in the P-site onto the A-site amino acid. Catalysis is mediated by rRNA base A3908 (in *H. sapiens*; A2062 in *E. coli*) which assists in the nucleophilic attack of the ester carboxyl group of P-tRNA by the α-amino group of A-tRNA [153-156]. This step results the rotation of the 40S subunit which is coupled to the translocation tRNA to the next ribosome site, where the deacylated tRNA in the P site is now in the E-site and the

peptide chain-tRNA in the A-site is now in the P-site [157]. The translocation of tRNA and mRNA is assisted by eEF2 [158]. The E-tRNA is then released from the E-site and the cycle is ready to begin again. Peptide chain elongation continues, with the growing peptide chain moving from the PTC, through the peptide exit tunnel exit the ribosome at the other side (**Fig. 1.15**).

1.8.3 - Termination

Termination of translation occurs in response to an mRNA stop codon (UGA, UAA, UAG) in the 40S decoding centre. Termination requires the release of polypeptide from the ribosome which is catalysed by eukaryotic release factors (eRF). First eRF1-eRF3-GTP binds to the A-site, with eRF1 recognising the stop codon.

eRF1 is composed of three domains. First is the N-terminal domain which senses mRNA stop codons through conserved TASNIKS and YxCxxxF (YCF) motifs [159-161] (**Fig. 1.17**). A central catalytic domain inserts into the A-site of the PTC to catalyse the hydrolysis of the peptide-tRNA bond using a conserved catalytic GGQ motif, thereby releasing nascent polypeptide from the ribosome [162] (**Fig. 1.17**). Lastly, the C-terminal domain of eRF1 interacts with another eukaryotic release factor, eRF3, which couples the GTP hydrolysis of eRF3 to the hydrolysis activity of eRF1's central domain [163, 164] (**Fig. 1.17**). eRF1 structurally resembles a tRNA molecule with the catalytic domain resembling an A-tRNA acceptor stem and the N-terminal domain resembling an A-tRNA anti-codon stem (**Fig. 1.16, 1.17**).

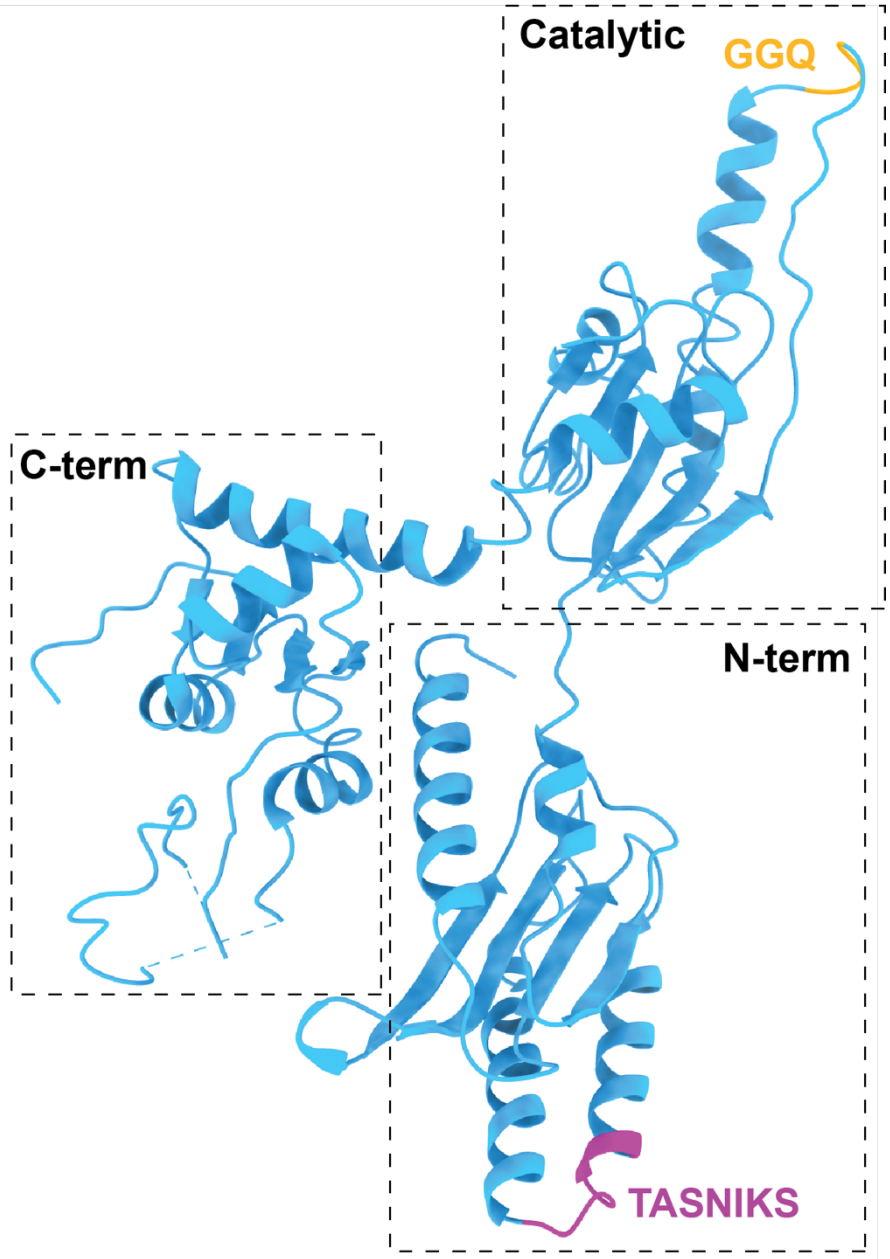


Figure 1.17 – Eukaryotic release factor 1 (eRF1) domain organisation
Structure of eRF1 (PDB ID 1dt9) [165]. Dashed line boxes denote three domains of eRF1: N-terminal (N-term) domain, central catalytic domain and C-terminal (C-term) domain. mRNA sensing TASNIKs motif highlighted in magenta. Catalytic GGQ motif highlighted in yellow.

After binding of the eRF complex, eRF3 hydrolyses GTP which triggers the eRF1 catalytic domain to hydrolyse the peptide-tRNA bond. During eRF1-mediated tRNA-peptide hydrolysis, the glutamine in the GGQ motif is thought to coordinate a water molecule to mediate nucleophilic attack of the peptide-tRNA ester bond [165]. This frees the polypeptide to exit the peptide exit tunnel for protein folding. The ribosome can now be

disassembled to remove mRNA, deacylated tRNA and release factors, allowing protein translation to re-initiate.

1.9 - Protein translation stalling and ribosome quality control

Ribosome stalling during protein translation can occur due to many reasons such as defective mRNA, absence of an in-frame stop codon and inappropriate read through of the mRNA poly-adenine (poly-A) tail. This can result in aberrant polypeptides being produced leading to protein aggregation and toxic effects in the cell. Several mechanisms exist in the cell to surveil and counteract their production, such as the ubiquitin-proteasome system.

Ribosome stalling by read through of the mRNA poly-A tail (AAA codons) is thought to occur in approximately 5-10% of all protein translation events in mammalian cells [166]. This is due to difficulty in decoding the poly-A mRNA codons, as well effects of the polypeptide [167]. Decoding of the poly-A tail results in the translation of a poly-lysine chain. These poly-basic residues are thought to interact with the highly acidic, RNA-rich peptide exit tunnel, slowing down translation which can lead to ribosome collisions on the mRNA [168].

1.9.1 - Ribosome quality control pathway

In the cytosol, a ribosome stalling event triggers the ribosome-associated quality control (RQC) pathway, eventually leading to ribosome recycling and the degradation of the arrested polypeptide. First, the ubiquitin E3 ligase ZNF598 ubiquitylates the 40S subunit proteins RPS10 and RPS20 [169]. This signals for the ribosome to be split, separating the 40S from the 60S-polypeptide-tRNA complex [170]. This is achieved by recruitment of ribosome recycling factors Hbs1 and Pelota to the stalled ribosome [171]. The GTPase activity of Hbs1 is coupled to a conformational change in Pelota, situated in the A-site [172]. This conformational change allows the

recruitment of ABCE1, which in turn couples its ATPase activity to split the ribosome subunits. The NEMF/Listerin ubiquitin E3 ligase complex is then recruited to the remaining 60S-peptide-tRNA complex, which ubiquitylates the nascent polypeptide. The ubiquitylated peptide is then released from the tRNA by ANKZF1, where it is then degraded by p97 [173, 174].

The ribosome quality control pathway for ER-bound ribosomes is less clear. ER-ribosomes are localised to the ER membrane via the heterotrimeric SEC61 complex. SEC61 is a translocon that binds to the ribosome at the peptide exit tunnel. During ER-protein translation, the nascent polypeptide exits directly from the peptide exit tunnel, through the SEC61 translocon into the ER-lumen without being exposed to the cytosol. However, the cytosolic RQC pathway relies on Listerin ubiquitylating the stall polypeptide as it exits the peptide exit tunnel which in the case of ER-bound ribosomes is inaccessible to Listerin. Therefore, further mechanisms are needed for ribosome quality control at the ER-membrane.

1.10 - The UFL1 ligase complex in ribosome quality control

Multiple recent studies implicate the UFMylation machinery and UFMylation process in ER homeostasis. Yet the identification of UFMylation substrates localised at the ER membrane have remained elusive. The ribosome 60S subunit protein ribosome protein large 26 (RPL26) was identified as a target for UFMylation, a process that only occurs on ER-bound ribosomes [137, 139, 175]. Additionally, this UFM1 modification was shown to be removed by the ER membrane-localised de-UFMylase UFSP2, providing a mode of negative regulation of ribosome UFMylation [137].

The RPL26 protein sits adjacent to the exit of the peptide exit tunnel, close to where the SEC61 translocon binds. Two lysines at the surface exposed C-terminus of RPL26, K-134 and to a lesser degree K-132, were shown to be UFMylated [137]. This ribosome UFMylation was associated with rescuing translation stalling after a failure of co-translational protein translocation into the ER. A failure in the UFMylation pathways lead to ER-associated protein degradation (ERAD) burden by build-up of aberrant polypeptide [137]. ERAD is a protein quality control mechanism where misfolded polypeptides are exported from the ER lumen to the cytosol for degradation by the proteasome.

Treating ribosomes with various translation-blocking drugs promoted UFL1 ligase complex recruitment to the ribosome and RPL26 UFMylation [175]. The most marked affect was observed after anisomycin treatment, with more modest UFMylation observed after treatment with harringtonine and cycloheximide [175]. These three agents all bind to the PTC of the ribosome to induce ribosome stalling with different mechanisms. Anisomycin blocks ribosome translation by preventing peptidyl transfer during elongation [176]. Harringtonine inhibits the initial elongation step by blocking the A-site [177]. Cyclohexamide inhibits the eEF2-mediated translocation of tRNA on the ribosome [178]. This difference in UFMylation levels perhaps points to the UFL1 ligase specifically recognising ribosomes that have a specific stalled conformation within the PTC.

Treatment of cells with the proteasome inhibitor MG-132 stabilised an overexpressed cytosol-destined stall peptide, whereas an overexpressed ER-destined stall peptide was still degraded, suggesting that ribosome stalled poly-peptides translated by ER-bound ribosomes are not degraded by the cytosolic proteasome pathway [175]. Further, the ER stall peptide was not stabilised by inhibition of the ERAD protein HRD1, suggesting that ER stalled peptides are also not degraded by the ERAD pathway [175]. It was instead suggested that RPL26 UFMylation eventually results in the export of stalled nascent polypeptide chains from

the ER lumen for degradation through the lysosomal degradation pathway [175]. The various reports showing UFL1 ligase complex recruitment to the ribosome and UFMylation of ribosome protein RPL26 suggest a role in stalled ribosome rescue, stall polypeptide degradation and ribosome recycling, although the exact functions remain unclear.

1.11 - Aims of the project

While much effort has focussed on understanding the Myddosome signalling cascade that underpins immune signalling, the termination of this signalling by the E3 ligase adaptor, SPOP, is not so well understood. A key question that remains is what is the structural and functional basis of SPOP interacting with MyD88 in order to terminate Myddosome immune signalling? A minimal binding site between the MATH domain of SPOP and three residues of MyD88 was recently proposed; however, the binding affinity and any additional interactions between SPOP and the Myddosome complex have not yet been explored. Currently, there are no full-length (FL) structures of the SPOP-Myddosome complex or of the individual proteins that make up this complex. Therefore, understanding the structural basis of how the SPOP-MyD88 complex assembles will help elucidate its biological functions. The work in **Chapters 3-5** will try to address this.

The ribosome was recently identified as a substrate of the UFL1 E3 ligase complex. Currently there are no FL structures of the UFL1 ligase complex and the mechanism of ribosome recognition is unknown. Additionally, it is unknown what exactly this ribosome UFMylation signals for, although it has been implicated in rescuing stalled ribosomes. The work in **Chapter 6** will aim to address these questions using a structural biology approach.

Gaining the structure of these multimeric E3 ligase complexes is key to understanding how they enact substrate modification, which is key to

understanding their roles in biology. To this end, the following aims of this thesis are proposed.

Aim 1: Characterise the interactions between SPOP and MyD88 using binding assays and X-ray crystallography

Aim 2: Reveal the oligomeric assembly of SPOP using cryo-EM

Aim 3: Elucidate the high-order structure of the SPOP-MyD88 oligomer by cryo-EM

Aim 4: Gain the high-resolution structure of the UFL1 ligase complex bound to the ribosome by cryo-EM

Chapter 2 - Materials and Methods

2.1 - Protein expression plasmids used in this study

Fusion protein	Truncation or mutation	Plasmid	Expression system	Cloned by	Expressed and purified by
His-MBP-3C-SPOP^{MATH}	SPOP$^{28-166}$	Modified pGEX-6P-1	Bacterial	S. Wright lab, University of Leeds	L. Makhlouf
His-MBP-3C-SPOP$^{28-296}$	SPOP$^{28-296}$	pOPINM	Bacterial	L. Makhlouf	L. Makhlouf
His-MBP-3C-SPOP$^{28-359}$	SPOP$^{28-359}$	pOPINM	Bacterial	S. Wright lab, University of Leeds	L. Makhlouf
His-MBP-3C-SPOP$^{18-359}$	SPOP$^{18-359}$	pOPINM	Bacterial	S. Wright lab, University of Leeds	L. Makhlouf
His-Avi-SUMO-MyD88	FL	pUCDM	Insect	GenScript	-
His-TEV-MyD88	FL	pProEX HTb	Bacterial	L. Makhlouf	L. Makhlouf
His-TEV-MyD88^{DD-ID}	MyD88$^{1-158}$	pProEX HTb	Bacterial	L. Makhlouf	L. Makhlouf

UFL1 ligase complex (His-TEV-UFL1, GST-3C-UFC1-His-3C-UFM1, MBP-3C-CDK5-Rap3, StrepII-3C-UFBP1)	UFBP1$^{29-314}$, UFC1^{C116K}, UFM1$^{1-83}$	-	Bacterial	Y. Kulathu lab, University of Dundee	Y. Kulathu lab, University of Dundee

Table 2.1 – Protein expression plasmids used in experiments
All gene sequences are from *Homo sapiens*. FL is full length. His is hexa-histidine tag. TEV is tobacco etch virus cleavage site. MBP is maltose binding protein tag. 3C is human rhinovirus 3C protease cleavage site. GST is glutathione s-transferase tag. StrepII is a W-S-H-P-Q-F-E-K peptide tag.

2.2 - Agarose gel electrophoresis

DNA samples were analysed on 1% weight per volume (w/v) agarose gels prepared with Tris-Borate-Ethylenediaminetetraacetic acid (EDTA) (TBE) buffer and 1:1000 SYBR safe DNA gel stain (Invitrogen). TBE buffer (VWR Life Sciences) contained 89 mM Tris, 89 mM Boric acid and 2 mM EDTA. DNA samples were mixed with 6x Purple Gel Loading Dye (NEB), diluting to a 1x loading dye concentration prior to gel running. 5 μL of Quick-Load Purple 1kb Plus DNA ladder (NEB) was added to one lane as a DNA base pair length standard. Gels were run at 100 V for 1 hour before imaging the gel under UV light.

2.3 - Plasmid DNA transformation into DH5α chemical competent cells

Plasmids were transformed into DH5α *Escherichia coli* (*E. coli)* cells for plasmid amplification. For plasmid transformation, 50 μL of chemically competent DH5α cells were thawed on ice. 10 ng of plasmid in a volume not exceeding 5 μL was added to the cells and the tube was flicked 5

times to mix. The mixture was incubated on ice for 30 minutes followed by heat shock in a water bath at 42 °C for 30 seconds. The tube was then placed on ice for 5 minutes. 900 μL of Miller Luria Bertani (LB) broth (10 g Tryptone, 5 g Yeast extract, 10 g NaCl per L) was added to the mixture, followed by incubation at 37°C for 1 hour shaking at 220 revolutions per minute (rpm). 150 μL of transformation reaction was plated onto LB-agar plates containing antibiotic for plasmid selection. Plates were incubated overnight at 37°C to allow colony formation.

2.4 - Plasmid DNA purification from bacterial cells

To isolate plasmid DNA from transformed *E. coli* cells, 5 mL of LB with antibiotic was inoculated with cells from an LB-agar plate colony. Cells were grown overnight at 37 °C shaking at 220 rpm. Cells were pelleted at 4,000 x g for 5 minutes and the supernatant was discarded. Plasmid DNA was isolated from the pelleted cells using the QIAprep Spin Miniprep kit (Qiagen) following the manufacturers protocol. Plasmid DNA was eluted from the spin column in 30 μL of milli-Q water.

2.5 - Bacterial expression of SPOP^{MATH}

SPOP^{MATH} (residues 28-116) with an N-terminal His-tag, MBP-tag and a HRV 3C protease cleavage site was cloned into a modified pGEX-6P-1 vector by the Dr Stephanie Wright lab, University of Leeds (**Table 2.1**). The SPOP^{MATH} plasmid was transformed into BL21 (DE3) chemical competent *E. coli* cells (NEB), following the manufacturers protocol. The transformation reaction was plated onto LB-agar plates supplemented with 100 μg/mL ampicillin antibiotic. A confirmed clone was used to inoculate 20 mL of LB with 100 μg/mL ampicillin and grown overnight at 37 ºC shaking at 220 rpm. For protein expression, 1 L of Terrific Broth (TB) medium (Millipore; 24 g Yeast extract, 12 g Peptone, 9.4 g di-Potassium hydrogen phosphate, 2.2 g Potassium dihydrogen phosphate per L) supplemented with 4 mL of glycerol was inoculated with the 20 mL

His-MBP-3C-SPOP^{MATH} BL21 (DE3) overnight culture and grown at 37 ºC until an optical density at 600 nm (OD_{600}) of 0.8. Protein expression was induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the cells were incubated overnight at 18ºC shaking at 220 rpm. Cells were harvested by centrifugation at 4,000 x g.

2.6 – Purification of SPOP^{MATH}

A BL21 (DE3) cell pellet expressing His-MBP-3C-SPOP^{MATH} was resuspended in 60 mL lysis buffer (25 mM HEPES pH 7.4, 150 mM NaCl, 20 mM Imidazole, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF)). Cells were sonicated for 10 minutes, pulsing for 1 second on and 4 seconds off. The lysate was clarified by centrifugation at 30,000 x g. Clarified lysate was incubated with 5 mL of equilibrated His-Pur Ni-NTA resin (Thermo Scientific) for 1 hour, rolling, at 4 ºC. The Ni-NTA resin was washed with ~50 mL of lysis buffer and the protein was eluted from the resin in lysis buffer containing 200 mM Imidazole. The elution was pooled and supplemented with ~580 μg of His-HRV 3C protease to cleave the His-MBP-tag. The protein was dialysed with SnakeSkin 3.5 kDa cut-off dialysis membrane (Thermo Fisher) overnight in dialysis buffer (25 mM HEPES pH 7.4, 150 mM NaCl, 1 mM DTT). Dialysed protein was incubated with Ni-NTA for 3 hours at 4 ºC to remove the cleaved His-MBP tag, any un-cleaved protein and the His-tagged protease. The flow-through (FT) containing tag-cleaved SPOP^{MATH} was collected and further purified by size exclusion chromatography (SEC) using an ÄKTA Pure system (Cytivia) with either a HiLoad 16/600 Superdex 75 pg column or a Superose6 Increase 10/300 gl column (Cytivia) in gel filtration buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP)). The purified protein was analysed by sodium dodecyl-sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).

2.7 - SDS-PAGE

Protein samples were diluted in 4x SDS-PAGE reducing buffer (240 mM Tris pH 6.8, 40% glycerol, 8% SDS, 0.04% Bromophenol blue, 5% β-mercaptoethanol), diluting to a 1x concentration of SDS-PAGE reducing buffer. Samples were boiled for 4 minutes at 95 °C. 12% Bis-Acrylamide gels were run in Tris-Glycine running buffer (25 mM Tris pH 8.3, 192 mM Glycine, 0.1% w/v SDS) at 200 V for 50 minutes. 5 μL of Precision Plus Protein Dual Color Standards (BioRad) was used as a molecular weight marker. Gels were stained with SimplyBlue SafeStain (Invitrogen) and imaged using a ChemiDoc XRS+ imaging system (BioRad).

2.8 - α-His western blot

SDS-PAGE gels were transferred to nitrocellulose membrane at 25 V for 47 minutes using a Trans-Blot Turbo transfer system (BioRad) in Towbin's buffer (25 mM Tris, 192 mM glycine, 20% w/v methanol). The nitrocellulose membrane was washed with double distilled H_20 (dd H_20) and blocked in 5% w/v BSA in PBTS (1x phosphate buffered saline, 0.05% Tween-20) at RT for 1 hour. Membrane was washed 3x with PBST and incubated with 1:2500 of mouse α-His^6-tag Monoclonal Antibody (Invitrogen) in PBST overnight at 4ºC. Membrane was washed 3x with PBST and incubated with 1:2500 rabbit α-mouse IgG SuperClonal Secondary antibody, HRP conjugate (Invitrogen) for 1 hour at RT. Membrane was washed 3x and blot was visualised using SuperSignal West Pico PLUS Chemiluminescent Substrate (ThermoFisher) according to manufacturer's instruction. Blot was imaged in a ChemiDoc XRS+ imaging system (BioRad) using the high-sensitivity default settings.

2.9 – Cloning and bacterial expression of $SPOP^{28-359}$ and $SPOP^{18-359}$

$SPOP^{28-359}$ and $SPOP^{18-359}$ with an N-terminal His-tag, MBP-tag and a HRV 3C protease cleavage site (His-MBP-3C) was cloned into a modified pOPINM plasmid by the Dr Stephanie Wright lab, University of Leeds. Plasmids were transformed into BL21 (DE3) cells and protein expression was performed as stated for the $SPOP^{MATH}$ protein expression.

2.10 – Purification of $SPOP^{28-359}$ and $SPOP^{18-359}$

$SPOP^{28-359}$ and $SPOP^{18-359}$ were purified in a similar manner to $SPOP^{MATH}$, except the protein was dialysed using a 10 kDa molecular weight cut-off SnakeSkin dialysis membrane (Thermo Fisher) and the SEC purification was performed using a HiLoad 16/600 Superdex 200 pg column (Cytivia).

2.11 - Glutaraldehyde crosslinking of $SPOP^{28-359}$

To optimise the glutaraldehyde (GA) crosslinking conditions of $SPOP^{28-359}$, six reactions were set up in gel filtration buffer, where $SPOP^{28-359}$ was at a final concentration of either 20 µM or 40 µM and GA (Sigma) was at a final concentration of either 0.00%, 0.05% or 0.10%. Reactions were incubated with on ice for 30 minutes. SDS-PAGE reducing buffer was added to samples before analysing the samples by SDS-PAGE. 5 µL HiMark Pre-Stained Protein Standard (Thermo Fisher) was used as the molecular weight standard. Samples were run on a NuPAGE 3-8% Tris-acetate 12-well gel (Thermo Fisher) run in NuPAGE Tris-Acetate SDS buffer (Thermo Fisher). The gel was run at 150 V for 68 minutes.

2.12 - Cloning, expression and purification of SPOP$^{28-296}$

To clone a bacterial expression plasmid for the expression of His-MBP-3C-SPOP$^{28-296}$, a stop codon was introduced after amino acid 296 of the His-MBP-3C-SPOP$^{28-359}$ in pOPINM plasmid by site-directed mutagenesis (see next section). The Sanger sequencing confirmed plasmid was then transformed in BL21 (DE3) cells from protein expression, as stated for the SPOP^{MATH} protein expression. SPOP$^{28-296}$ was purified as stated for the SPOP^{MATH} purification.

2.13 - Site-directed mutagenesis cloning

Site-directed mutagenesis (SDM) of a gene in a plasmid was achieved by PCR (**Fig. 2.1**). Forward and reverse primers containing mutant DNA bases were used to amplify the WT plasmid. 10 ng of the WT plasmid DNA template was amplified by PCR using hot start KOD polymerase (Novagen). Three reactions were performed with the addition of either 0%, 4% or 8% dimethyl sulfoxide (DMSO). The PCR reaction had an initial denaturation step at 95°C for 2 minutes. This was followed by 20 cycles consisting of a denaturing step at 95 °C for 20 seconds, a primer annealing step at the lowest primer melting temperature for 10 seconds and an extension step at 70 °C for 25 seconds/kb of DNA template. The PRC reaction was finished with a final extension step at 70 °C for 5 minutes. Reactions were treated with 1 μL of *DpnI* (NEB) in 1x Cutsmart buffer (NEB) for 30 minutes at 37 °C to remove any WT parental plasmid. *DpnI* treated reactions were purified using the QIAquick PCR purification kit (Qiagen) as per manufacturers protocol. DNA was eluted from the QIAquick column in 30 μL of milli-Q water and the DNA concentration was measured using a NanoDrop 8000 spectrophotometer (Thermo Scientific). 10 ng of DNA was transformed into DH5α chemical competent cells for plasmid amplification.

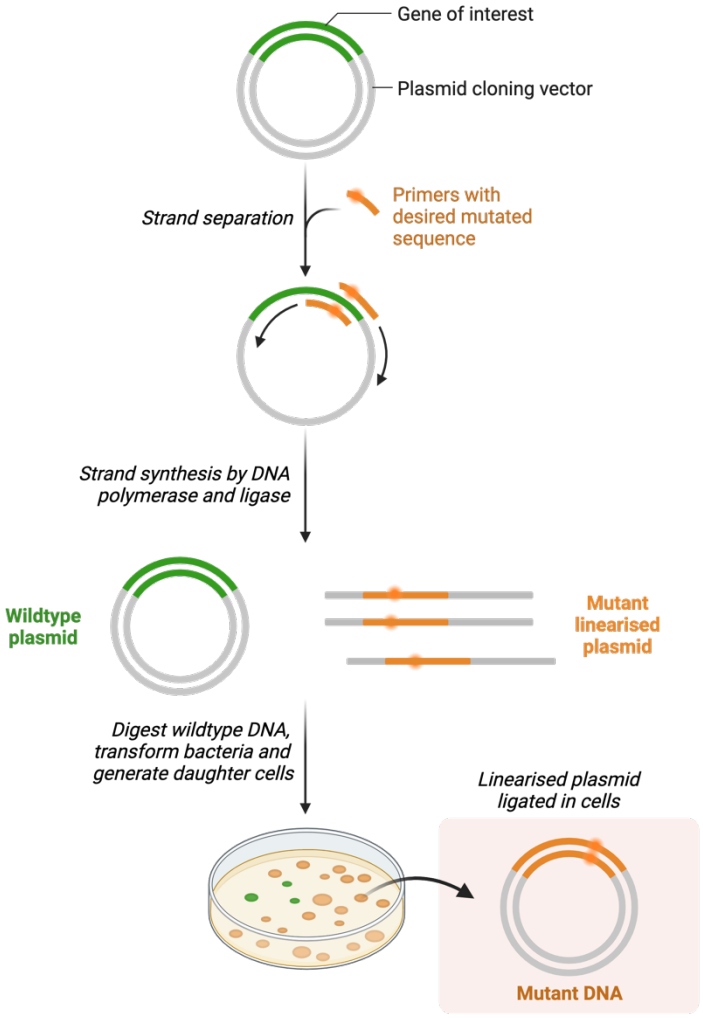


Figure 2.1 – Overview of site-directed mutagenesis (SDM) cloning
SDM cloning was used to generate a mutation in a gene of interest contained in a DNA plasmid. The wild type plasmid was amplified by PCR using primers containing the desired mutation sequence, resulting in a linearised plasmid containing the mutant gene. Wild type methylated plasmid DNA was digested using *DpnI* restriction enzyme. The SDM reaction was then transformed into *Escherichia coli* cells, where the linearised plasmid was then re-ligated. Schematic adapted from "Site-directed Mutagenesis", by BioRender.com (2023). Retrieved from https://app.biorender.com/biorender-templates.

2.14 – Cloning of full-length MyD88

For the bacterial expression of FL MyD88, MyD88 was cloned into a pPROEX HTb plasmid. This results in the expression of a fusion protein of an N-terminally tagged His-TEV-MyD88. To clone the FL MyD88 plasmid, the MyD88 gene was amplified by PCR from another plasmid (His^8-Avi-SUMO-MyD88 in pUCDM plasmid) using primers to add

restriction enzyme sites BamHI and XbaI at the 5' and 3' ends respectively. PCR products were analysed by agarose gel electrophoresis and the remaining reaction was purified using a QIAquick PCR purification kit (Qiagen) as per manufacturers protocol. DNA was eluted from the QIAquick column in 30 μL of milli-Q water. To digest the restriction site ends of the PCR amplified MyD88 gene, 1 μg of purified DNA was digested using BamHI-HF and XbaI restriction enzymes (NEB) in Cutsmart buffer (NEB). The reaction was incubated at 37°C for 30 minutes. The reaction was purified using the QIAquick PCR purification kit (Qiagen) as previous. To digest the empty pPROex HTb plasmid, 1 μg of plasmid was digested using BamHI-HF and XbaI restriction enzymes (NEB) in Cutsmart buffer (NEB). The reaction was incubated at 37°C for 30 minutes before being run on a 1% agarose gel. The band corresponding to the linear digested plasmid was excised and the digested plasmid was isolated from the gel using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturers protocol. The digested plasmid was eluted from the column in 25 μL of milli-Q water. The restriction digested MyD88 insert and digested plasmid were ligated in a 3:1 molar ratio using T4 DNA ligase (Roche). 50 ng of MyD88 insert, 28 ng of plasmid and 3μL of T4 DNA ligase was mixed in 1xligation buffer (Roche) to a total volume of 20 μL. The reaction was incubated overnight at 4 °C. 10 μL of the ligation mixture was transformed into 100 μL of DH5α cells as stated previously, except during the outgrowth step 400 μL of LB was added to the cells. 150 μL of cells were plated on LB-agar plates with 100 μg/mL ampicillin. The plate was incubated overnight at 37 °C. To confirm colony clones, 5 mL of LB medium with 100 μg/mL ampicillin was inoculated with individually picked colonies and grown overnight at 37 °C shaking at 220 rpm. The cells were pelleted and the plasmids extracted using the QIAprep Spin Miniprep kit (Qiagen) as described previously. Correct plasmid clones were confirmed by Sanger sequencing (Eurofins). Next, 10 ng of plasmid was transformed into BL21 (DE3) chemical competent *E. coli* cells (NEB), following the manufacturers protocol. 150 μL of transformation reaction was plated

onto LB-agar plates with 100 μg/mL ampicillin. The plate was incubated overnight at 37 °C. To confirm colony clones, 5 mL of LB medium with 100 μg/mL ampicillin was inoculated with individually picked colonies and grown overnight at 37 °C shaking at 220 rpm. The plasmid was extracted from cells as previous and correct clones were identified by PCR with gene specific primers.

2.15 – Cloning of MyD88^{DD-ID}

For the bacterial expression of a truncation of MyD88 (residues 1-158) which includes the Death domain and SBC containing intermediate domain (MyD88^{DD-ID}), a stop codon was introduced to the His-TEV-MyD88 in pPROex HTb plasmid after Pro-158 by SDM. This results in the expression of a fusion protein of an N-terminally tagged His-TEV-MyD88$^{1-158}$. SDM PCR was performed as stated previously (**Fig. 2.2**). 1 μL of *DpnI* treated SDM reaction was transformed into 50 μL of DH5α cells as previous. Correct clones with the introduced stop codon were identified by Sanger sequencing. 10 ng of plasmid was transformed into BL21 (DE3) for protein expression and transformed colonies were confirmed by PCR using gene specific primers.

2.16 - Bacterial expression of MyD88 and MyD88^{DD-ID}

For protein expression of His-TEV-MyD88 and His-TEV-MyD88^{DD-ID}, transformed BL21 (DE3) cells were used to inoculate 20 mL of LB with ampicillin and grown overnight at 37 ºC shaking at 220 rpm. TB media supplemented with 100 μg/mL ampicillin was inoculated with 20 mL of respective transformed BL21 (DE3) overnight LB culture and grown at 37 ºC shaking at 220 rpm until an OD_{600} of 0.8. Expression was induced with 0.2 mM IPTG and the cultures were incubated overnight at 18 ºC shaking at 220 rpm. Cells were harvested by centrifugation at 4,000 x g and were stored at -80 °C before purification.

2.17 - Purification of MyD88 and MyD88^{DD-ID}

BL21 (DE3) cell pellets from 1-3 L of cell culture expressing either His-TEV-MyD88 or His-TEV-MyD88^{DD-ID} were resuspended in 115 mL of lysis buffer containing 25 mM HEPES pH 7.4, 150 mM NaCl, 20 mM Imidazole, 1 mM Benzamidine and 0.075% β-mercaptoethanol. After 5 minutes, 0.2 mM PMSF was added to the resuspended cells. Cells were sonicated on ice for 10 minutes at 40% amplitude, pulsing for 1 second on and 4 seconds off. Lysed cells were pelleted for 30 minutes at 30,000 x g at 4 °C. The supernatant was incubated for 2.5 hours rolling at 4°C with 5-7.5 mL of Ni-NTA resin which was pre-equilibrated in lysis buffer. The Ni-NTA-protein mixture was transferred to a gravity flow column and was washed with 50 mL of lysis buffer. The protein was eluted from the Ni-NTA beads in elution buffer containing 25 mM HEPES pH 7.4, 150 mM NaCl, 200 mM Imidazole and 0.075% β-mercaptoethanol. Eluted protein was dialysed overnight using a 10 kDa cut-off SnakeSkin Dialysis Tubing membrane (Thermo Scientific) with ~370 μg of His-TEV protease, in dialysis buffer containing 25 mM HEPES pH 7.4, 150 mM NaCl, 5% glycerol and 1 mM TCEP. The cleaved protein was incubated with 5 mL of equilibrated Ni-NTA slurry for 1.5 hours, rolling at 4°C. The flow-through containing cleaved protein was concentrated using a 10 kDa molecular weight cut-off VivaSpin 6 PES centrifugal concentrator (Sartorius) by centrifuging at 3,000 x g for 10-minute spins, resuspending the protein after every spin. The protein was then separated by SEC using a Superose6 Increase 10/300 24 mL column (Cytivia). The column was equilibrated in gel filtration buffer containing 25 mM HEPES pH 7.4, 20-150 mM NaCl and 0.5 mM TCEP. 1-2 mL of protein was injected onto the column at a flow rate of 0.5 mL/minute. Peak fractions were analysed by SDS-PAGE and purified protein was flash frozen in liquid nitrogen and stored at -80°C for later use.

2.18 - Co-purification of MyD88^{DD-ID} and SPOP^{MATH}

To attempt to produce a complex of SPOP^{MATH} and MyD88^{DD-ID}, a co-immunoprecipitation (co-IP) of His-MBP-SPOP^{MATH} and MyD88^{DD-ID} was performed. His-MBP-SPOP^{MATH} was buffer exchanged into co-IP buffer containing 25 mM HEPES, 150 mM NaCl, 20 mM Imidazole and 0.075% β-mercaptoethanol using a Zeba Spin 7 kDa cut-off desalting column (Thermo Fisher), following manufacturers protocol. Two reactions were set up: a His-MBP-SPOP^{MATH} and MyD88^{DD-ID} binding experiment and a negative control with just MyD88^{DD-ID}. 100 μg of His-MBP-SPOP^{MATH} and 300 μg of MyD88^{DD-ID} was incubated with 100 μL of Ni-NTA resin for 2.5 hours, rolling at 4ºC. Ni-NTA beads were pelleted by centrifugation at 800 x g and beads were washed 3x in 500 μL of co-IP buffer. SDS-PAGE buffer was added to pelleted beads and the samples were boiled for 5 minutes before pelleting the beads again. 15 μL of the elution as well as the input, flow-through and wash were run on a 10-well 12% SDS-PAGE gel.

2.19 - Purification of the MyD88^{DD-ID} and SPOP$^{18-359}$ complex

To try to form a high-order MyD88^{DD-ID} and SPOP$^{18-359}$ oligomer complex for cryo-EM, the individual proteins were first purified, as mentioned previously. 1 mL of purified SPOP$^{18-359}$ at 0.7 mg/mL was mixed with 1.2 mL of purified MyD88^{DD-ID} at 0.67 mg/mL and incubated on ice for 30 minutes. The mixture was then injected onto a Superose6 Increase 10/300 column for purification of the complex by SEC.

2.20 – Preparation of UFL1 ligase-bound ribosomes (by Dr Joshua Peter)

The UFL1 ligase complex was prepared as described in [131]. In brief, UFM1 and CDK5RAP3 were individually expressed in BL21 (DE3) cells and purified using Ni^{2+}-NTA affinity chromatography. Cells were resuspended in lysis buffer containing 25 mM Tris pH 8, 300 mM NaCl,

10% glycerol, 2 mM DTT, 1 mM benzamidine, 1 mM AEBSF and 1x protease inhibitor cocktail (Roche). Cells were lysed by sonication and pelleted at 30,000x g for 30 minutes at 4 °C. The clarified lysate was incubated with Ni^{2+}-NTA resin for 2 hours in binding buffer containing 25 mM Tris pH 8, 300 mM NaCl, 10% glycerol and 10 mM imidazole. The resin was washed in wash buffer containing 25 mM Tris pH 8, 300 mM NaCl, 10% glycerol, 2 mM DTT and 20 mM imidazole and eluted in buffer containing 25 mM Tris pH 8, 300 mM NaCl, 10% glycerol, 2 mM DTT and 300 mM imidazole. The tags were cleaved using HRV 3C protease and a size exclusion step was performed using either a HiLoad 16/60 Superdex 75 pg or a HiLoad 16/60 Superdex 200 pg column (GE Healthcare Life Sciences) with buffer containing 25 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol and 2 mM DTT. $UFC1^{C116K}$ was expressed in BL21 (DE3) cells and purified using glutathione 4B-sepharose resin. Cells were lysed in buffer containing 25 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol and 2 mM DTT by sonication. The lysate was clarified and incubated with glutathione 4B-sepharose resin for 2 hours. The resin was washed in a high salt buffer (25 mM Tris pH 7.5, 500 mM NaCl, 10% glycerol and 2 mM DTT) followed by a low salt buffer (25 mM Tris pH 8, 150 mM NaCl, 10% glycerol and 2 mM DTT). The protein was eluted from the resin by tag cleavage using HRV 3C protease and a size exclusion step was performed as previous.

To prepare the UFC1-UFM1 conjugate, 30 μM UBA5, 30 μM $UFC1^{C116K}$ and 60 μM UFM1 were incubated in 25 mM HEPES pH 7.5, 200 mM NaCl, 10 mM MgCl2 and 10 mM ATP. The pH of the reaction mixture was adjusted to 9.8 with 0.5 M CAPS pH 11.5, and incubated for 18 hours at 23 °C. UFC1-UFM1 was subsequently separated from UBA5 and unreacted UFC1 and UFM1 using a HiLoad 26/600 Superdex 75 pg column in buffer containing 25 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT.

His-UFL1 and Strep-UFBP1 were co-expressed in BL21 codon plus RIPL cells (Agilent). Cells were harvested and lysed in buffer containing 25 mM Tris pH 8.0, 300 mM NaCl, 2 mM DTT, 1 mM benzamidine, 1 mM AEBSF and protease inhibitor cocktail. Cells were lysed by high pressure homogenization using an Emulsiflex C3 homogenizer (Avestin) and clarified as previous. The clarified lysate was first purified using HisTrap FF column (GE Healthcare Life Sciences). The column was washed with 25 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole and 2 mM DTT and eluted in buffer containing 25 mM Tris pH 8.0, 300 mM NaCl, 300 mM imidazole and 2 mM DTT. The elution was further purified using a StrepTrap column (GE Healthcare Life Sciences) equilibrated in 25 mM Tris pH 8.0, 300 mM NaCl and 2 mM DTT. The column was washed in equilibration buffer and eluted in 25 mM Tris pH 8.0, 300 mM NaCl, 2 mM DTT and 2.5 mM desthiobiotin. Lastly, the UFL1-UFBP1 complex was purified using a HiLoad 16/60 Superdex 200 pg column.

60S ribosomes were purified from HEK293 cells grown to ~80% confluency in 15 × 15 cm dishes supplemented with media containing high glucose DMEM supplemented with 10% v/v Fetal Bovine Serum (FBS), 50 mg/mL Penicillin Streptomycin, and 2 mM L-Glutamine. To harvest cells, media was removed by aspiration, washed with ice-cold PBS followed by removal of PBS by aspiration. Cells were scrapped in residual PBS and transferred to a 15 mL falcon. Cells were pelleted by centrifugation at 1000 x g for 3 minutes. Next, the cell pellets were resuspended in lysis buffer (15 mM Tris pH 7.6, 1500 mM NaCl, 10 mM $MgCl_2$, 1% (v/v) Triton-X 100, 2 mM DTT, RNAsin (60 units), 1x cOmplete mini protease inhibitor cocktail and mixed gently followed by incubation on ice for 10 min. Cell lysates were then centrifuged at 17,000 x g for 10 minutes and the supernatant was collected. The collected supernatant was layered directly onto a high-salt sucrose cushion containing 20 mM Tris pH 7.5, 500 mM KCl, 30% (v/v) sucrose, 10 mM $MgCl_2$, 0.1 mM EDTA pH 8.0 and 2 mM DTT. Total ribosomes were sedimented by centrifugation at 63,000 x g for 18 hours using a Type 70 Ti rotor (Beckman Coulter). The sedimented ribosomes were then resuspended

in buffer containing 20 mM Tris pH 7.5, 500 mM KCl, 7.5% (v/v) sucrose, 2 mM $MgCl_2$, 75 mM NH_4Cl, 2 mM puromycin and 2 mM DTT. The resolubilized pellet containing ribosomes was incubated at 4 °C for 1 hour and then at 37 °C for 1.5 hours. To isolate 40S and 60S ribosomal subunits, the solution was layered directly on a linear 10–30% sucrose gradient containing 20 mM Tris pH 7.5, 500 mM KCl, 6 mM $MgCl_2$ and 2 mM DTT. The 60S and 40S were separated by centrifugation at 49,123 x g for 9 hours 42 minutes at 4 °C using a SW41 Ti rotor (Beckman Coulter). Gradients were fractionated into 0.5 mL fractions using BioComp fractionating system. Fractions containing 60S ribosomal subunits were collected and exchanged into buffer containing 20 mM HEPES pH 7.2, 100 mM KCl, 5 mM $MgCl_2$ and 2 mM DTT.

To form the UFL1 ligase-60S ribosome complex, 10 µM of UFL1-UFBP1 and CDK5RAP3, 5 µM of UFC1-UFM1 and 1 µM of purified 60S ribosomes were incubated together in buffer containing 20 mM HEPES pH 7.2, 50 mM KCl, 5 mM $MgCl_2$ and 0.25 mM TCEP for 2 hours at 4 °C. Post incubation, samples were mixed with 0.05% glutaraldehyde for 30 seconds at 23 °C followed by quenching with 100 mM Tris pH 8.0 (final concentration). The crosslinked sample was then layered on a 10-30% sucrose gradient containing 20 mM HEPES pH 7.5, 50 mM KCl, 5 mM $MgCl_2$ and 0.25 mM TCEP and centrifuged using a TLS55 rotor at 24,000 rpm for 6 hours at 4 °C. The sucrose gradient was then manually fractionated into 100 µL fractions. The fractions containing UFL1 ligase complex-60S ribosome, as determined by immunoblotting, were then pooled, buffer exchanged in 25 mM HEPES pH 7.5, 50 mM KCl, 5 mM $MgCl_2$, 2 mM DTT and concentrated to 7.7 mg/mL.

2.21 – Preparation of UFMylated ribosomes (by Dr Joshua Peter)

An *in vitro* UFMylation reaction was performed by incubating 0.1 µM UBA5, 5 µM UFC1, 10 µM UFM1, 3 µM UFL1:UFBP1 complex, 5 µM

CDK5RAP3 and 1 µM 60S ribosomes in the presence of 5 mM $MgCl_2$ and 5 mM ATP. Then, 10 µM of isopeptide-linked UFC1-UFM1 was added to the reaction and further incubated at 4 °C for 2 h. The sample was then crosslinked with 0.05% glutaraldehyde and used directly for cryo-EM visualisation.

2.22 - Preparation of membrane-associated 60S ribosomes (by Mr Rohan Thakur)

Parental cells (WT-HEK293) or CDK5RAP3 KO cells (~80% confluency) grown in 10 x 15 cm dishes were washed briefly with ice cold PBS and collected in a 15 mL falcon tube. Cells were pelleted by centrifugation at 500 x g for 5 minutes. Cell pellets were resuspended in buffer containing 20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 µg/mL cycloheximide, 0.02% w/v digitonin, 1x cOmplete protease inhibitor cocktail – EDTA-free (Roche) and RNasin for 10 minutes on ice, followed by centrifugation at 17,000 x g for 10 minutes. The clarified supernatant is the cytosolic fraction and was discarded. The remaining membrane pellet was resuspended in lysis buffer containing 20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 µg/mL cycloheximide, 1% w/v Decyl Maltose Neopentyl Glycol (DMNG), 1x cOmplete protease inhibitor cocktail – EDTA-free (Roche) and RNasin for 15 minutes on ice, followed by centrifugation at 17,000 x g for 10 minutes. The clarified supernatant was collected and layered on a 10 - 30% sucrose gradient containing 20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 µg/mL cycloheximide, 0.01% DMNG followed by centrifugation at 36,000 rpm for 3 hours using SW41 Ti rotor. Fractions containing 60S ribosomes were collected and exchanged into buffer containing 20 mM HEPES pH 7.2, 100 mM KCl, 5 mM $MgCl_2$, 2 mM DTT and stored at -80 °C until use.

2.23 - Fluorescence polarization (FP) assay

The binding affinity between SPOP and substrate peptides was measured using an FP assay (**Fig. 2.2**).

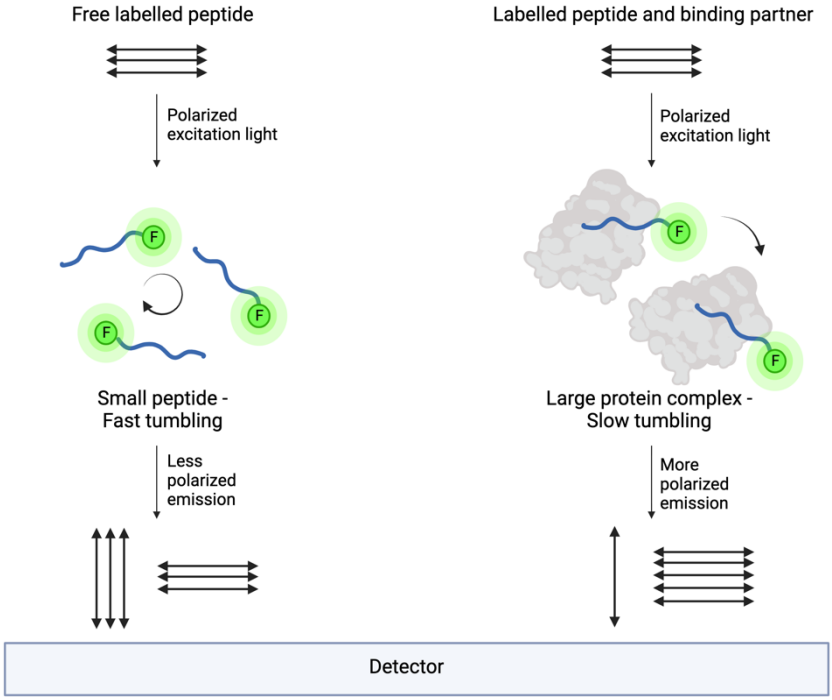


Figure 2.2 – Overview of a fluorescence polarization assay
A fluorescently-labelled peptide is mixed with its unlabelled binding partner. The fluorophore (F) on the peptide is then excited by polarised light. Arrows indicate direction of light. The level of polarization of the light that is emitted from the excited fluorophore is dependent on the tumbling speed of the molecule, with larger molecules, such as a peptide-protein complex, tumbling more slowly compared to the smaller free peptide. This degree of light polarization is measured by the detector in millipolarization (mP) units and allows the direct quantification of binding. The following equation is used to calculate fluorescence polarization (mP): $mP = ((F\| - F\perp)/(F\| + F\perp))/1000$, where $F\|$ is the fluorescence intensity parallel to the polarised excitation light and $F\perp$ is the fluorescence intensity perpendicular to the polarised excitation light. Created with BioRender.com.

All peptides used for FP assays were synthesised by Peptide Synthetics (Peptide Protein Research Limited), purified to >95% purity (**Table 2.2**). For the fluorescently-labelled peptides, peptides were labelled with 5-Carboxyfluorescein (5Flu), either via the N-terminal amino group (N-terminal label) or via the side chain ε-amino group of an added C-terminal lysine (C-terminal label). Peptides and proteins were diluted in FP buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 1 mM TCEP, 0.005%

Tween-20). All concentrations were measured in triplicate, in 384-well black flat-bottom low flange plates (Corning) and with a 20 µL reaction volume. Millipolarization (mP) measurements were taken in a Hidex Sense microplate reader (Hidex) at 25ºC, with an excitation of 490 nm (20 nm filter), emission of 520 nm (14 nm filter) and G-factor of 1. The optimum concentration of 5Flu-peptide was empirically tested, with a final working concentration range between 6-50 nM.

Protein	**Sequence**	**N-term label**	**C-term label**	**Mutation**
MyD88	^{133}AVDSSVP^{139}	5Flu	-	-
MyD88	^{125}AEKPLQVAAVDSSVPRT^{141}	5Flu	-	-
MyD88	^{13}PVSSTSS^{19}	5Flu	-	-
MyD88	^{125}AEKPLQVAAVDSSVP^{139}	-	K-5Flu	-
MyD88	^{133}AVDSSVP^{139}	-	K-5Flu	-
MyD88	^{13}PVSSTSS^{19}	-	K-5Flu	-
MyD88	^{133}AVDSSVP^{139}	-	-	-
MyD88	^{133}AVDSSVPRT^{141}	-	-	-
MyD88	^{125}AEKPLQVAAVDSSVPRT^{141}	-	-	-
MyD88	^{128}PLQVAAVDSSVPRT^{141}	-	-	-
MyD88	^{125}AEKPLQVAAVAAAVPRT^{141}	-	-	^{135}DSS^{137} to AAA
MyD88	^{133}AVAAAVP^{139}	-	-	^{135}DSS^{137} to AAA
MyD88	^{125}AEKPLQVAAVDS*S*VPRT^{141}	-	-	Phospho-S-136 and S-137
MyD88	^{125}AEKPLQVAAVDS*SVPRT^{141}	-	-	Phospho-S-136
MyD88	^{125}AEKPLQVAAVDSS*VPRT^{141}	-	-	Phospho-S-137
MyD88	^{125}AEKPLQVAAVDSSVPRT*^{141}	-	-	Phospho-T-141

MyD88	^{125}AEK$^{(Ac)}$PLQVAAVDSSVPRT^{141}	-	-	Acetyl-K-127
MyD88	^{125}AAKPLQVAAVDSSVPRT^{141}	-	-	E126A
MyD88	^{125}AEAPLQVAAVDSSVPRT^{141}	-	-	K127A
MyD88	^{125}AEKGLQVAAVDSSVPRT^{141}	-	-	P128G
MyD88	^{125}AEKPAQVAAVDSSVPRT^{141}	-	-	L129A
MyD88	^{125}AEKPLAVAAVDSSVPRT^{141}	-	-	Q130A
MyD88	^{125}AEKPLQVAAADSSVPRT^{141}	-	-	V134A
MyD88	^{125}AEKPLQVAAVASSVPRT^{141}	-	-	D135A
MyD88	^{125}AEKPLQVAAVDASVPRT^{141}	-	-	S136A
MyD88	^{125}AEKPLQVAAVDSAVPRT^{141}	-	-	S137A
Pdx1	^{220}VAEPEQDCAVTSGEE^{234}	-	K-5Flu	-
Puc	^{90}SRENLACDEVTSTTS^{104}	-	K-5Flu	-
MyD88-Puc	AEKPLQVAAVTSTTS	-	-	-
Gli2	^{1140}NNMPVQWNEVSSGTVD^{1156}	-	-	-
DEK	^{277}SVKSANVKKADSSTTKK^{292}	-	-	-
SRC3	^{91}NDDDVQKADVSSTGQGV^{107}	-	-	-
SENP7	^{192}DTDNLQSEQLSSSSDGS^{208}	-	-	-
SETD2	^{1363}DKGSVQAPEISSNSIKD^{1379}	-	-	-
CAPRIN 1	^{426}QPEATQVPLVSSTSEGY^{442}	-	-	-

Table 2.2 – Peptides used in fluorescence polarization assays
Residue numbers are indicated. MyD88-Puc is a chimera peptide with ^{125}AEKPLQVAA^{133} from MyD88 and ^{99}VTSTTS^{104} from Puc. * represents a phosphate group. (Ac) represents an acetyl group. 5Flu is 5-carboxyfluorescein. N-term is N-terminus. C-term is C-terminus.

For direct FP binding measurements using fluorescently-labelled peptides, a 2-fold dilution series of SPOP protein was plated and mixed with a constant concentration of 5Flu-labelled peptide. The measured triplicate mP values were averaged and normalised against a peptide only control. Data was plotted and analysed using Prism 9 software (GraphPad Prism version 9, GraphPad Software, Boston,

Massachusetts USA, www.graphpad.com). Error bars are standard error of the mean. A one-site non-linear binding curve was fit to the data to derive the dissociation constant (K_d) which was calculated as the concentration of SPOP at half the maximum mP.

For peptide competition FP binding measurements, a 2-fold dilution series of unlabelled peptide mixed with a constant concentration of 5Flu-labelled peptide and $SPOP^{MATH}$ was measured. The concentration of $SPOP^{MATH}$ was determined as the concentration which produced 70% of the maximum mP signal in the direct FP binding assay. The triplicate mP values were averaged and normalised against a no unlabelled peptide control. Data was plotted and analysed using Prism 9 software. Error bars are standard error of the mean. An [inhibitor] vs. response – variable slope (four parameters) curve was fit to the data to give the half maximal inhibitory concentration (IC_{50}), which was calculated as the concentration of unlabelled peptide half way between the top and the bottom of the curve.

2.24 - Isothermal titration calorimetry (ITC)

The binding interaction between $SPOP^{MATH}$ and various peptides was measured using ITC. Experiments were performed using a Microcal™ iTC_{200} system (Malvern). In all experiments, 300 μM of peptide was titrated into 30 μM of $SPOP^{MATH}$. All samples were in FP buffer and samples were degassed prior to measurements. Runs were performed with a reference power of 5 μcalories/second, stirring speed of 750 rpm, at 25 °C with a total of 20 injections. The first injection was 0.5 μL for 2 seconds, followed by 2 μL injections for 4 seconds. Spacing between the injections was 120 seconds. A heat of dilution control titration was performed where 300 μM peptide was titrated into FP buffer. Results were analysed using Origin software (Malvern). The data was normalised against the heat of dilution control and a binding isotherm curve was fitted

using a One Set of Sites model. The K_d was calculated from the reciprocal of the binding affinity (K).

2.25 - Mass photometry

The molecular mass of SPOP$^{28-359}$ was measured using a One^{MP} mass photometer (Refeyn). For mass photometry measurements, SPOP$^{28-359}$ was diluted in gel filtration buffer to concentrations of 3 μM and 300 nM. A clean coverslip with wells was placed on the mass photometer with a drop of immersion oil. Data were acquired using AcquireMP software (Refeyn) and was processed using DiscoverMP (Refeyn). The focus level of the well was found using the buffer free mode. Once focussed, the sample was added to the well and a movie was acquired using the normal mode with a regular image size. The movie was processed and calibrated against reference proteins of known mass (66-660 kDa). The resulting histogram shows the mass distribution of the sample. The median mass molecular mass and standard deviation is calculated from a Gaussian fit of the histogram peak.

2.26 - Mass spectrometry of MyD88^{DD-ID} (performed by Dr Ranjani Ganji)

To identify the additional SDS-PAGE bands present after MyD88^{DD-ID} purification, protein identification mass spectrometry was performed. Purified MyD88^{DD-ID} sample was run on a 12% SDS-PAGE gel and stained as stated previously. All subsequent processing of the sample and data analysis were performed by Dr Ranjani Ganji. Gel bands were excised and washed in 30% ethanol, incubating at 70 °C shaking until all stain was removed. Gel pieces were washed by incubating in 25 mM ammonium bicarbonate/50% acetonitrile for 10 minutes shaking. Cysteines were reduced by incubating with 100 μL of 10 mM DTT at 57 °C for 1 hour shaking. Supernatant was discarded and cysteines were alkylated with 100 μL of 55 mM iodoacetic acid, incubated at room

temperature in the dark for 45 minutes shaking. The supernatant was discarded. Gel pieces were dehydrated by addition of 100% acetonitrile and incubation for 5 mins at room temperature. Acetonitrile was removed and gel pieces were left to dry. Once dry, the gel slices were cooled on ice then covered with ice-cold trypsin solution (0.02 µg/µL in 25 mM ammonium bicarbonate) and left on ice for 10 minutes to rehydrate. Excess trypsin solution was removed and discarded and the gel slices were covered with a minimal amount of 25 mM ammonium bicarbonate. After briefly vortexing and centrifuging, the gel slices were incubated at 37 °C with shaking for 18 hours. The resulting digest was vortexed and centrifuged. The supernatant containing the peptides was recovered and added to a tube containing 5 µL acetonitrile/ water/ formic acid (60/35/5; v/v) to quench protease activity. 50 µL of acetonitrile/ water/ formic acid (60/35/5; v/v) was added to the gel slices and vortexed for an additional 10 minutes. The supernatant was pooled with the previous wash and one additional wash of the gel slices was performed. The pool of the three washes was dried by vacuum centrifugation. The peptides were reconstituted in 20 µL 0.1% aqueous trifluoroacetic acid. Digested peptides were processed using the S-TRAP Micro column (PROTIFI, NY, USA) following the manufacturer's instructions. Sample was solubilised in 50 mM triethylammonium bicarbonate (TEAB) containing 5% SDS. Reduction and alkylation were then performed. DTT was added to a final concentration of 20 mM before heating to 56 °C for 15 minutes with shaking. The sample was left to cool, then iodoacetamide was added to a final concentration of 40 mM, before heating to 20 °C for 15 minutes with shaking in the dark. Proteins were further denatured by acidification; phosphoric acid was added to a final concentration of 1.2%. Samples were then diluted with S-Trap binding buffer (100 mM TEAB pH 7.1 in methanol). 1 µg of trypsin reconstituted in 50 mM TEAB was added before quickly being loaded onto the S-trap column. The S-trap was washed by adding 150 µL binding buffer before being spun at 4000 x g for 30 seconds. 30 µL of 0.02 µg/µL trypsin was then added to the top of the S-trap. S-traps were loosely capped and placed in a 1.5 mL tube and

heated to 46 °C for 15 minutes. Digested peptides were eluted by first spinning the S-trap at 4,000 x g for 1 minute. Further elutions used 40 µL 50 mM TEAB, 40 µL 0.2% formic acid, and 30 µL 50% acetonitrile with 0.2% formic acid, prior to centrifugation. Elutions were combined then dried down prior to resuspension in 50 µL 0.2% formic acid containing 3% acetonitrile. For liquid chromatography tandem mass spectrometry of the digested peptides, sample was injected onto an in house-packed 20cm capillary column (inner diameter 75µm, 3.5µm Kromasil C18 media). An Ultimate 3000 nano liquid chromatography system was used to apply a gradient of 2–30% ACN in 0.1% formic acid over 30 minutes at a flow rate of 300 nL/min. Total acquisition time was 60 minutes, including column wash and re-equilibration. Peptides were eluted from the column and into an Orbitrap Exploris 240 Mass Spectrometer (ThermoFisher Scientific) via a nanospray flex ion source, using a capillary voltage of 2.7 kV. Precursor ion scans were acquired in the Orbitrap with resolution of 60,000. EASY-IC internal calibration was used for precursor ion scans. Up to 20 ions per precursor scan (charge state 2+ and higher) were selected for HCD fragmentation using a normalised collision energy of 30%. Fragments were measured in the Orbitrap at a resolution of 15,000. Dynamic exclusion of 30 seconds was used. Peptide data were processed with PEAKS Studio XPro (Bioinformatic Solutions Inc) and searched against the MyD88 sequence provided. Carbamidomethylation was selected as a fixed modification, variable modifications were set for oxidation of methionine and deamidation of glutamine and asparagine. Mass spectrometry mass tolerance was 20 ppm, and fragment ion mass tolerance was 0.05 Da. The peptide false discovery rate was set to 1%.

2.27 - *In vitro* 60S ribosome-Sec61 dissociation assays (by Mr Rohan Thakur)

In vitro 60S-Sec61 dissociation reaction was performed by incubating 0.05 µM membrane solubilized 60S ribosomes (60S-Sec61 solubilized

and enriched from CDK5RAP3 KO Cells) with 0.5 µM UBA5, 1 µM UFC1, 1 µM UFM1, 0.1 µM UFL1/UFBP1, 0.1 µM CDK5RAP3 in the presence of 5 mM $MgCl_2$ and 5 mM ATP at 37°C for 25 minutes. At the end of the reaction, the reaction mix was layered over a 10 - 50% sucrose gradient containing 20mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT and centrifuged at 36,000 rpm for 3 hours using a SW41 Ti rotor. Sucrose gradients were fractionated using BioComp fractionation system. The sucrose gradient fractions were separated on 4-12% SDS-PAGE gel and analysed for co-migration of Sec61β with 60S ribosomes by immunoblotting.

2.28 - SPOP^{MATH} and MyD88 peptide X-ray crystallography

Sample was prepared for X-ray crystallography by mixing a MyD88 peptide (^{125}AEKPLQVAAVDSSVPRT^{141}) with SPOP^{MATH}, in buffer containing 25 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP. SPOP^{MATH} (552 µM final concentration) and MyD88 peptide (1.75 mM final concentration) were mixed and incubated on ice for 4 hours. For crystal tray preparation, 60 µL of precipitant mix was added to the plate reservoir. Crystals were grown by sitting drop vapour diffusion by mixing 200 nL of protein with 200 nL of reservoir buffer using a Mosquito robot (Labtech). Crystal trays were incubated at 18°C, with crystal formation observed after 2 days. Crystals were picked 11 days after setting up trays and flash frozen in liquid nitrogen. For the crystal that was used for data collection, no cryoprotectant was added during crystal picking as this was present in the buffer mix. Crystals were sent to Diamond Light Source for data collection (data collection parameters shown in **Table 2.3**). Data was collected on a crystal from a Morpheus HT-96 crystal screen (well E12; 0.12 M Ethylene Glycol, 0.1 M Buffer System 3 pH 8.5, 50% v/v Precipitant Mix 4) which diffracted to 1.7 Å, later truncated to 2.0 Å (**Table 2.3**). Data was processed using xia2-3dii [179] and the structure was solved by molecular replacement using a SPOP^{MATH} structure (PDB ID 7KPI) [63]. The structure then underwent several rounds of model building

in Coot-0.9.8.1 [180] followed by refinement in Phenix-1.2.1 [181]. Building of the MyD88 peptide was performed manually. Final model statistics are show in **Table 2.3**. Omit map was generated by setting the occupancy of the MyD88 peptide to 0, followed by a round of refinement. Map was displayed in Pymol-2.4.2 and contoured at σ1.

Data collection and processing parameters

Data collection:	
Beam line	I24 microfocus
Wavelength (Å)	0.99987
Space group	$P2_12_12_1$
Cell dimensions a, b, c (Å)	40.00, 57.29, 60.97
α, β, γ (°)	90.00, 90.00, 90.00
Resolution	33.44-2.00 (2.07-2.00)
Total reflections	58226
Unique reflections	9882 (939)
Completeness (%)	99.6 (98.9)
R_{merge} (%)	18.1 (62.0)
$R_{p.i.m.}$ (%)	8.1 (28.2)
I/σI	10.7 (5.5)
Multiplicity	5.9 (5.6)
$CC_{1/2}$	0.99 (0.86)
Refinement:	
R_{work}/R_{free}	0.187/0.232
B-factors (Å^2)	19.61
RMS bonds (Å)	0.006
RMS angles (°)	0.849
Ramachandran favoured (%)	97.33
Ramachandran allowed (%)	2.00
Ramachandran outliers (%)	0.67
Number of water molecules	26

Table 2.3 – X-ray crystallography data collection and refinement statistics of the SPOP^{MATH}-MyD88 peptide crystal structure
Highest resolution shell information is in parentheses. Unique reflections are the total number of reflections after symmetry averaging. Completeness

evaluates the coverage of all theoretically possible unique reflections. R_{merge} evaluates the precision of unique reflections of the unmerged data and is dependent on the multiplicity. $R_{p.i.m.}$ evaluates the precision of the merged data. $I/\sigma I$ is the signal-to-noise ratio. Multiplicity is the average number of measurements per unique reflection. $CC_{1/2}$ is the correlation coefficient between the intensities of two random halves of the total reflections. R_{work} measures the agreement between the model and the experimental data. R_{free} was calculated using a random 10% of data that was not used for refinement. Model was refined using Phenix [181] refine and statistics were calculated using MolProbity [182].

2.29 – X-ray crystal structure of the UFC1-UFM1 complex (by Dr Helge Magnussen)

UFC1-UFM1 crystals were obtained using the sitting drop vapour diffusion technique where UFC1-UFM1 (22.8 mg/mL) was 1:1 mixed with 30% *v*/*v* PEG 400, 0.1 M Tris pH 8.5, 0.2 M Na citrate and incubated at 19 °C. Single crystals appeared within 2-3 days. Crystals were flash-frozen in crystallisation buffer containing 30% *v*/*v* ethylene glycol. Data sets were collected at Diamond Light Source (DLS), beamline I04, and processed with Xia2 [179] and DIALS [195]. The crystal structure was solved by Molecular replacement (PHASER) [196] using the crystal structures of UFC1 (PDB ID 3evx) [133] and UFM1 (PDB 5ia7) [197] as starting model. Refinement and model building was carried out with REFMAC [198] and Coot, respectively. The resolution was 1.78 Å and the final model has an R_{work}/R_{free} of 0.182/0.226.

2.30 - Negative stain electron microscopy of $SPOP^{28-359}$ and $SPOP^{18-359}$

Grid making

For negative stain electron microscopy (EM) grid making, formvar/carbon coated 300 mesh copper grids (Agar Scientific) were placed on a parafilm coated slide and were glow discharged using a PELCO easiGlow system (Ted Pella) at a pressure of 0.38 mBar for 30 seconds at 10 mA. Prior to use, the negative stain solution was centrifuged at 13,000 x g for 10 minutes to pellet any aggregates. For $SPOP^{28-359}$, a 1% uranyl acetate stain was used. For $SPOP^{18-359}$, a 1% uranyl formate stain was used.

First, 5 µL of protein sample was applied to the grid and incubated for 30 seconds. The sample was blotted by touching the side of the grid with a piece of Whatman blotting paper, leaving a thin layer of liquid on the surface of the grid. The grid was then washed by applying 5 µL of milli-Q water and immediately blotted, following by a second water wash and blot. 5 µL of negative stain (uranyl acetate or uranyl formate) was applied to the grid and immediately blotted, followed by a second application of stain and blot. 5 µL of stain was applied again and incubated for 3 minutes. The grid was blotted and left to dry prior to imaging.

Data collection

Negative stain-EM grids were imaged on a Tecnai F20 transmission electron microscope (FEI) operated at 200 kV with a CETA (CMOS CCD) camera (FEI). For SPOP$^{18-359}$ data collection, 101 micrograph images in TIF format were acquired at a nominal magnification of 25,000x (actual magnification 33,492x), giving a pixel size of 3.51.

Image processing

An overview of a typical EM image processing pipeline is shown in **Fig. 2.3**. Micrographs were initially image processed using Relion-3.1.2 [183]. First, the image contrast transform function (CTF) was estimated using CTFFind-4.1.13 [184]. 1,176 particles were manually picked with a particle diameter size of 200 Å by selecting regions along a SPOP$^{18-359}$ oligomer. Particles were extracted with a box size of 90 pixels. Extracted particles were used for reference-free 2D classification, asking for 20 classes. Particles from selected classes were used as a reference for particle autopicking, with a picking threshold of 0.1 and a minimum inter-particle distance of 200 Å. 68,739 autopicked particles were extracted as stated previously. Particles were imported into CryoSPARC-3.2 [185] for further processing. The particles were used to generate two initial 3D models with C1 symmetry. The best initial model containing 27,815 particles was taken forward for non-uniform refinement with C1 symmetry to generate the final map.

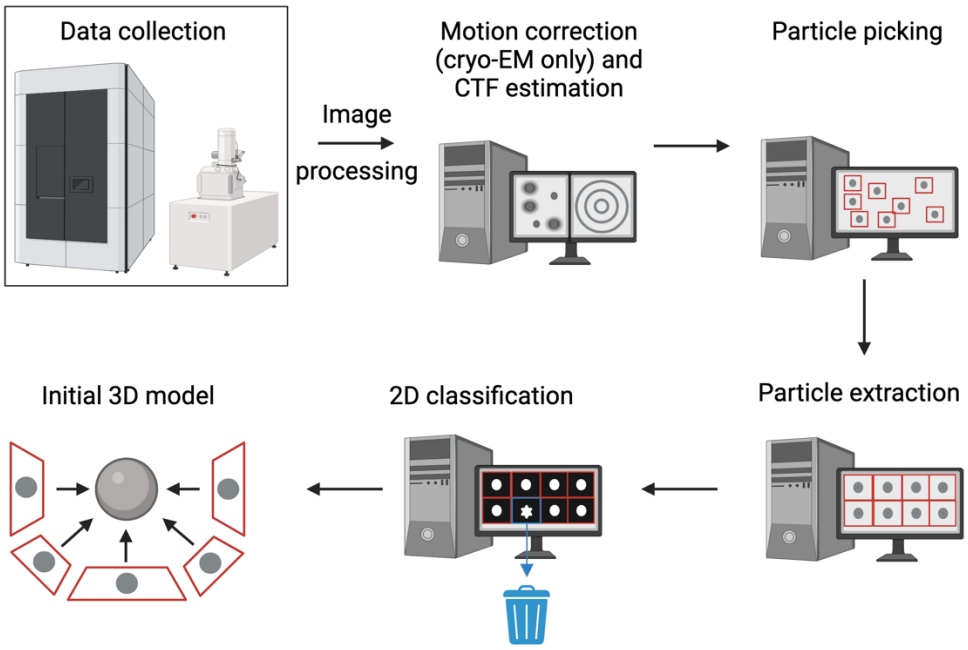


Figure 2.3 – Overview of electron microscopy image possessing steps
Negative stain-EM (micrograph) and cryo-EM (movie) data collected from a transmission electron microscope is transferred to a high-performance computing machine for image processing. Cryo-EM movies show significant blurring due to motion of the sample caused by the electron beam during data collection. To correct for this blurring, the movies are motion corrected using MotionCor2 [186]. During this step the movie is gain-corrected and dose-weighted, where high resolution information from later frames of the movie (which may be radiation damaged) are down-weighted compared to earlier movie frames. Motion correction first corrects the global motion of each frame, then splits each frame into 5x5 patches to correct the local motion in these patches. The motion between frames is corrected by identifying identical features in the frame and merging the frames of a movie into a micrograph image. All of the following processing steps are identical for negative stain and cryo-EM data processing. Next, the contrast transfer function (CTF) parameters of the microscope are estimated using CTFFIND4 [184]. This considers the defocus of the collected images, the accelerating voltage the microscope was operated at and the spherical aberration of the microscope to estimate and correct for the loss of contrast during imaging. Then, single molecules of interest (particles) are picked from the micrographs. This particle picking designates the coordinates within a micrograph which contains a particle. The picked particles then undergo particle extraction, where boxed images are saved at the position of the particle coordinates, selected during the previous steps. These extracted particles are then subjected to 2D classification, where similar 2D images are grouped and their signal is averaged together. This particle averaging increases the signal-to-noise and can help identify unwanted classes, such as contamination of an unwanted protein in the sample. These unwanted classes can be removed before continuing with downstream image processing. The selected 2D images represent random views of the 3D molecule of interest. These 2D images are used to reconstruct an initial 3D model by estimating the orientation of the 2D images in relation to one another. The data can then undergo further processing to refine the initial 3D model.

2.31 - Cryo-electron microscopy (cryo-EM) of SPOP$^{18-359}$

Grid preparation

Quantifoil R1.2/1.3 copper 300 mesh holey grids were plasma cleaned using a Tergio plasma cleaner (PIE Scientific) using the remote cleaning mode with an RF power of 15 Watt for 1 min. Cryo-EM grids were prepared using a Vitrobot MK IV (FEI) with a chamber humidity of 100% and a chamber temperature of 4°C (**Fig. 2.4**). 3 μL of 0.5 mg/mL SPOP$^{18-359}$ was applied to the glow discharged grid and immediately blotted for 6 seconds with a blot force of 0, followed by rapid plunge freezing in liquid nitrogen cooled liquid ethane (**Fig. 2.4**).

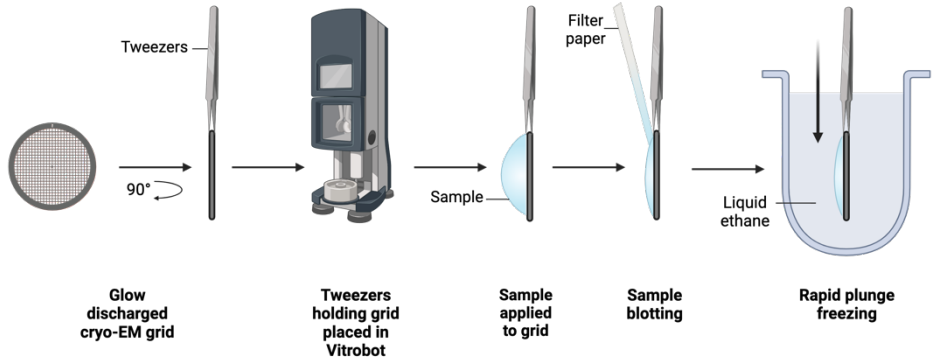


Figure 2.4 – Overview of cryo-EM grid preparation
Schematic showing the steps taken during cryo-EM grid preparation. Once a cryo-EM grid is glow discharged, it is placed in a humidity and temperature-controlled chamber in a Vitrobot. Sample is pipetted onto the grid through a small side door of the chamber. Automated sample blotting and plunge freezing steps are performed by the Vitrobot using a user specified blotting duration time.

Data collection

Single particle cryo-EM data was collected on a Titan Krios G2 transmission electron microscope (Thermo Fisher) with a Falcon 4 direct electron detector (Thermo Fisher). Data was collected with an accelerating voltage of 300 kV and a nominal magnification of 96,000x, which corresponds to a pixel size of 0.82 Å/pixel. A total of 2,916 cryo-EM movies were collected in EER format. Full data collection parameters are shown in **Table 2.4**.

Image processing

Cryo-EM movies were imported, motion corrected (MOTIONCOR2) and the CTF parameters estimated (CTFFIND-4.1) using Relion. 145 particles were manually picked from the micrographs with a particle picking diameter of 300 Å. Picked particles were extracted with a box size of 400 pixels. Particles where then subjected to reference-free 2D classification with a mask diameter of 380 Å, asking for 5 classes. Selected particles were then used as a reference for particle autopicking, with a picking threshold of 0.2 and a minimum inter-particle distance of 200 Å, resulting in 229,919 picked particles. Particles were extracted as previous and imported into cryoSPARC for further processing. Several rounds of 2D classification were performed, asking for 20 classes with an initial classification uncertainty factor of 3, number of online-EM iterations of 40 and batchsize per class of 200. This resulted in 30,899 particles. To generate an initial 3D model, the predicted structural model of the SPOP oligomer [100] was converted into a density map using ChimeraX [187] and the box and pixel size was resized to 400 pixels and 0.82 Å/pixel respectively using Relion image handler. The map was imported into cryoSPARC and low pass filtered to 60 Å before being used as an initial 3D model along with the selected particles for homogenous 3D refinement with C1 symmetry.

Image re-processing using FL SPOP as a reference model

To see whether using the published FL SPOP oligomer model [188] as an initial 3D reference would improve the resulting refined 3D model, the FL SPOP model was converted to a density map and resized as stated previously. This new initial model was used for homogenous 3D refinement with this studies' particles to generate a 3D refined density map.

Image processing using cryoDRGN and 3D Flexible Refinement (3DFlex)

To try to resolve any remaining heterogeneity in the data, a cryoDRGN [189] model was trained using the particles and particle poses that were

generated when using the SPOP oligomer predicted model [100] as an initial 3D model reference. First, the particles were downsampled to 128 pixels and a cryoDRGN model was trained with 8 latent dimensions and 50 training iterations. The trained model was then used to generate 20 reconstructions from the downsampled particle data. Particles contributing to compositional heterogeneity were remove using cryoDRGN particle filtering, which resulted in a final particle stack of 16,111 particles. The particles were re-extracted at their full box size and were re-imported into cryoSPARC for 3DFlex. 3DFlex was used to model the flexibility of the SPOP oligomer within the data. The particles were downsampled to 128 pixels and a training mesh mask was generated with a mask dilation of 4 Å. The downsampled particles and training mesh mask were used to train a 3DFlex model with 2 latent dimensions, 32 hidden units and a rigidity prior of 3. Two volume series were generated, which were used to generate morph movies showing the flexible motion of the $SPOP^{18-359}$ oligomer.

2.32 - Cryo-EM grid making and screening of $MyD88^{DD-ID}$

QuantiFoil R1.2/1.3 300 mesh copper grids were glow discharged using a PELCO easiGlow glow discharge system (Ted Pella), as previously stated. Cryo-grids were prepared using a Vitrobot MK IV (Thermo Fisher) with a chamber temperature of 4 °C and 100% humidity. 3 μL of $MyD88^{DD-ID}$ at 0.67 mg/mL was applied to the grid and immediately blotted for 6 seconds with a blot force of 1, followed by rapid plunge freezing in liquid ethane. Cryo-EM grids were screened on a ThermoFisher Titan Krios G2 transmission electron microscope with a ThermoFisher Falcon 4 direct electron detector, operating at an accelerating voltage of 300 kV. Micrographs were collected at nominal magnification of 96,000x, which corresponds to a pixel size of 0.82 Å.

2.33 - Cryo-EM of the MyD88^{DD-ID} and SPOP$^{18-359}$ complex

Complex formation

The MyD88^{DD-ID}-SPOP$^{18-359}$ complex was formed using purified SPOP$^{18-359}$ and a high concentration pre-size-exclusion chromatography stock of MyD88^{DD-ID}. SPOP$^{18-359}$ (19 μM) and MyD88^{DD-ID} (114 μM) were mixed in either a 1:1, 1:3, 3:1 volumetric ratio.

Grid preparation

Quantifoil R1.2/1.3 300 mesh copper grids were glow discharged using an PELCO easiGlow glow discharge system (Ted Pella) as previous. Cryo-EM grids were prepared using a Vitrobot as previous, but with a blot force of 1.

Data collection

A single particle cryo-EM data set was collected on a cryo-EM grid made with a 3:1 ratio of SPOP$^{18-359}$ and MyD88^{DD-ID}. Data was collected on a ThermoFisher Titan Krios G2 transmission electron microscope with a ThermoFisher Falcon 4i direct electron detector and SelectrisX energy filter. Data was collected with an accelerating voltage of 300 kV and nominal magnification of 165,000x, which corresponds to a pixel size of 0.74 Å (full data acquisition settings shown in **Table 2.4**). A total of 5,624 cryo-EM movies were acquired.

Image processing

Movies were imported into Relion for motion correction using MotionCor2. The CTF parameters were estimated using CTFFind4. 734 particles were manually picked with a particle diameter size of 200 Å. Picked particles were extracted with a box size of 560 pixels. Extracted particles were used for 2D classification, asking for 10 classes. Selected classes resulted in 679 remaining particles, which were used for particle autopicking with a picking threshold of 0.3 and a minimum inter-particle distance of 200 Å. The resulting 350,738 particles were extracted with a box size of 560 pixels and imported into cryoSPARC for further

processing. The particles were subjected to several rounds of 2D classification with 40 online-EM iterations and a batchsize per class of 200. 154,223 particles were used to generate an initial 3D model, which was refined using the non-uniform refinement algorithm to generate the final map.

2.34 - Cryo-EM of the UFL1 ligase-60S ribosome complex

Grid preparation

Cryo-grids were prepared with 0.05% glutaraldehyde crosslinked 60S-UFL1-UFBP1-CDK5RAP3-UFC1-UFM1 complex at 7.7 mg/mL in 25 mM HEPES pH 7.5, 50 mM KCl, 5 mM $MgCl_2$, 2 mM DTT (prepared by Joshua Peter, Yogesh Kulathu lab). Quantifoil R3.5/1 copper 200 mesh holey grids were glow discharged using a PELCO easiGlow glow discharge unit at 15 mAmp for 30 s. Cryo-grids were prepared using a Vitrobot MK IV (ThermoFisher) with a chamber temperature of 4 °C and 100% humidity. 3 μL of protein was applied to the grid and immediately blotted for 6 s with a blot force of 1, followed by rapid plunge freezing in cooled liquid ethane.

Data collection

Single particle cryo-EM data was collected on a Titan Krios G2 transmission electron microscope with a Falcon 4i direct electron detector and SelectrisX energy filter (Thermo Fisher). Data was collected with an accelerating voltage of 300 kV and nominal magnification of 165,000x, which corresponds to a pixel size of 0.74 Å (full data acquisition settings shown in **Table 2.4**). A total of 59,394 cryo-EM movies were acquired.

Image processing

The cryo-EM movies were imported, beam-induced motion corrected (MOTIONCOR2) and the CTF parameters estimated (CTFFIND4.1) using Relion. Approximately 2.2 million particles were picked from motion

corrected micrographs using crYOLO-1.6.1 [190] un-trained particle picking (2019 general model) with a particle box size of 400 pixels and a picking confidence threshold of 0.2. Picked particles were extracted in Relion with a particle box size of 588 pixels, re-scaled to 128 pixels (re-scaled pixel size ~3.4 Å). Extracted particles were imported into cryoSPARC-3.2 for processing. Seven rounds of reference-free 2D class averages were generated with initial classification uncertainty factor set between 2-7, number of online-EM iterations set to 40 and batchsize per class set to 200 and all ribosome-like particles were taken forward. The selected ~1.6 million particles were used to generate an initial 3D model with C1 symmetry. The initial 3D model was further refined using the non-uniform refinement algorithm with the dynamic masking start resolution set to a value below the resolution of the data (i.e., 1 Å), to generate a refined 3D model and a mask that encompasses the entire box size. The mask and model were input for 3D variability analysis asking for three classes. Particles from the class containing ligase bound 60S ribosomes were taken forward for another round of 3D refinement, this time with dynamic masking start resolution set to default (12 Å) and the dynamic mask threshold set to 0.1. This was then followed by several rounds of 3D variability analysis, asking for two classes to separate ligase bound 60S ribosomes from unbound 60S ribosomes, which resulted in 356,394 ligase-bound ribosome particles. Particles were then down sampled to 128 pixels and a cryoDRGN-3.2 [189] model was trained with 8 latent dimensions and 50 training iterations. CryoDRGN particle filtering removed 57,386 junk particles, resulting in a final particle stack of 299,008 particles. The homogenous particle population containing ligase-bound ribosomes were re-extracted in Relion at the full box size. A 3D model was generated with C1 symmetry, followed by non-uniform refinement with per particle defocus optimization, Ewald sphere correction and CTF refinement in cryoSPARC-4.2.1 to generate the ligase-bound 60S ribosome map.

To further refine the density for the ligase complex, two masks were created from the final 3D refinement volume using UCSF ChimeraX-

1.2.5 software: one which encompasses the ligase complex and another which encompasses the 60S ribosome. The ribosome mask was used for particle signal subtraction. Signal subtracted particles were then used for local refinement of the ligase complex, using the ligase mask, to generate a ligase only map. A cryoSPARC-4.2.1 3DFlex [191] training model was generated for the ligase with 6 latent dimensions and a rigidity prior of 2. The resulting 3DFlex model was used for 3DFlex reconstruction with 40 Max BFGS iterations to generate the final ligase only map. The ligase-bound 60S map was sharpened using Phenix-1.2.1 autosharpen map job and the ligase only map was sharpened using DeepEMhancer [192] tight target sharpening protocol.

Model building

Atomic models were built using Coot-0.9.8.1 [180]. For the ligase-bound 60S ribosome map, PDB ID 7QWR [193] was used as a starting model by rigid body fitting the model into the density map, followed by re-building in Coot. No ligase components were built into the ligase bound ribosome map except for the UFL1 loop. For the ligase complex, AlphaFold 2.0 [96] models of the individual proteins were separated into smaller segments and then rigid body fitted into the density map, followed by manual re-building in Coot. The UFL1 CTD (residues 515-786), CDK5RAP3 UUBD (residues 15-116) and UFM1 displayed poor side chain density and so the side chains of these regions were set to an occupancy of 0. Atomic models were refined using Phenix real space refinement [194] and validated using MolProbity [182]. All 3D density maps were visually inspected in UCSF ChimeraX-1.2.5 software [187].

2.35 - Cryo-EM of the UFMylated 60S ribosome

Data collection and image processing

Cryo-EM grids were prepared as previous with 1.5 mg/mL of sample. Single particle cryo-EM data was collected on a Titan Krios G2 transmission electron microscope with a Falcon 4 direct electron

detector. Data was collected with an accelerating voltage of 300 kV and nominal magnification of 96,000x, which corresponds to a pixel size of 0.82 Å (full data acquisition settings shown in **Table 2.4**). A total of 3,028 cryo-EM movies were acquired. The data was processed as previous, with the final map being generated from the particles after several rounds of 3D variability analysis.

Data collection parameters	**Ligase bound 60S ribosome**	**Ligase complex only**	**UFMylated 60S ribosome**	**SPOP$^{18-359}$**	**SPOP$^{18-359}$-MyD88^{DD-ID}**
Microscope	FEI Titan Krios G2				
Detector	Falcon4i		Falcon4		Falcon4i
Energy filter	SelectrisX		-	-	SelectrisX
Energy filter slit (eV)	10		-	-	10
Accelerating Voltage (kV)	300				
Magnification	x165,000		x96,000		x165,000
Spot size	8			6	8
Illuminated area (mm)	0.54		0.56	0.64	0.53
Pixel size (Å)	0.74		0.82	0.82	0.75
Defocus range (mm)	-0.2 to -2.0		-0.8 to -2.9	-1.5 to -3.0	-1.3 to -2.5
Total electron dose ($e^-/Å^2$)	33.4		39.8	33.2	40.8
Exposure (s)	2.67		6.66	4.01	3.25
Number of frames	42		32	41	51
Dose per frame ($e^-/Å^2$)	0.8				
Number of movies	59,394		3,028	2,916	5,624
Acquisition mode	Counting				
AFIS	Yes		No	No	Yes
AFIS range (mm)	8		-	-	8
EPU software version	3.2.0		2.14	3.0	3.4.0
Data possessing parameters	**Ligase bound 60S ribosome**	**Ligase complex only**	**UFMylated 60S ribosome**	**SPOP$^{18-359}$**	**SPOP$^{18-359}$-MyD88^{DD-ID}**

Symmetry point group	C1				
Final particle number	299,008		5,402	30,899	154,223
Map resolution (Å)	2.2	3.2	7.0	13.5	7.6
FSC threshold	0.143				
Map sharpening B factor ($Å^2$)	59.28	-	-	-	-

Model statistics	**Ligase bound 60S ribosome**	**Ligase complex only**
Bond length (Å) (# > 4σ)	0.004	0.004
Bond angles (°) (# > 4σ)	0.595	0.565
MolProbity score	1.65	1.82
Clash score	2.04	9.25
Ramachandran outliers (%)	0.02	0.00
Ramachandran allowed (%)	1.56	4.77
Ramachandran favoured (%)	98.43	95.23
Rotamer outliers (%)	0.95	0.09
Cβ outliers (%)	0.00	0.00
CaBLAM outliers (%)	1.12	2.63
Number of water molecules	2,164	-

Table 2.4 - Cryo-EM data collection, processing and model validation statistics

Chapter 3 – Characterising SPOP binding to MyD88

3.1 - Chapter introduction

The Myddosome-complex protein, MyD88, was recently identified as a substrate of the ubiquitin E3 ligase adaptor protein, SPOP [62]. This interaction was implicated as a mode of regulating Myddosome signalling through the ubiquitylation and degradation of MyD88. MyD88 contains a single SPOP-binding consensus (SBC) motif; however, the molecular details and mechanism of binding remains unknown. In this chapter, fluorescence polarization (FP) binding assays and X-ray crystallography were used to characterise the interaction between SPOP and MyD88.

3.2 - Measuring the binding affinity between MyD88 and SPOP

3.2.1 - Purification of SPOP MATH domain from *E. coli* cells

In order to investigate the binding interactions between SPOP and MyD88, a truncation of SPOP which included just the substrate-binding MATH domain ($SPOP^{MATH}$; residues 28-166; plasmid cloned by the Dr Stephanie Wright lab) was expressed and purified from *E. coli* cells (**Fig. 3.1 a**).

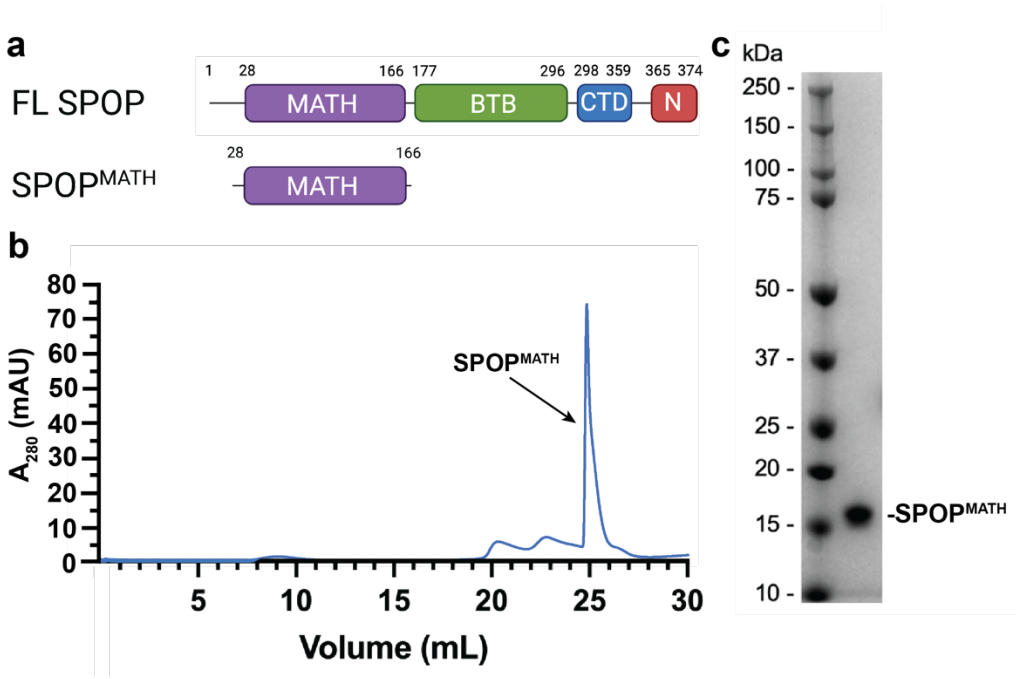


Figure 3.1 – Purification of the SPOP^{MATH} domain
a Domain organisation of full-length (FL) SPOP protein and the SPOP^{MATH} truncation protein used for *Escherichia coli* expression. Amino acid numbers are indicated. MATH is meprin and TRAF homology domain. BTB is Broad-complex, Tramtrack, and Bric-à-brac domain. CTD is C-terminal domain. N is nuclear localisation sequence.
b Size-exclusion chromatography chromatogram of SPOP^{MATH} purified using a Superose6 Increase 10/300 gl column.
c Representative SDS-PAGE gel of SPOP^{MATH} after size-exclusion chromatography purification.

This SPOP^{MATH} protein contained an N-terminal His^{6}-tag followed by an MBP tag to improve solubility. A HRV 3C protease cleavage site was included to remove the His-MBP tag after affinity purification. The His-MBP-tagged protein was initially purified via the His-tag using nickel-affinity purification. The tags were cleaved and the cleaved protein was isolated by reverse nickel-affinity purification. This protein was further purified by size-exclusion chromatography to obtain protein used for further experiments (**Fig. 3.1 b, c**). SPOP^{MATH} eluted at the predicted monomer position (~25 mL; 17 kDa) from the size-exclusion chromatography column (**Fig. 3.1 b**).

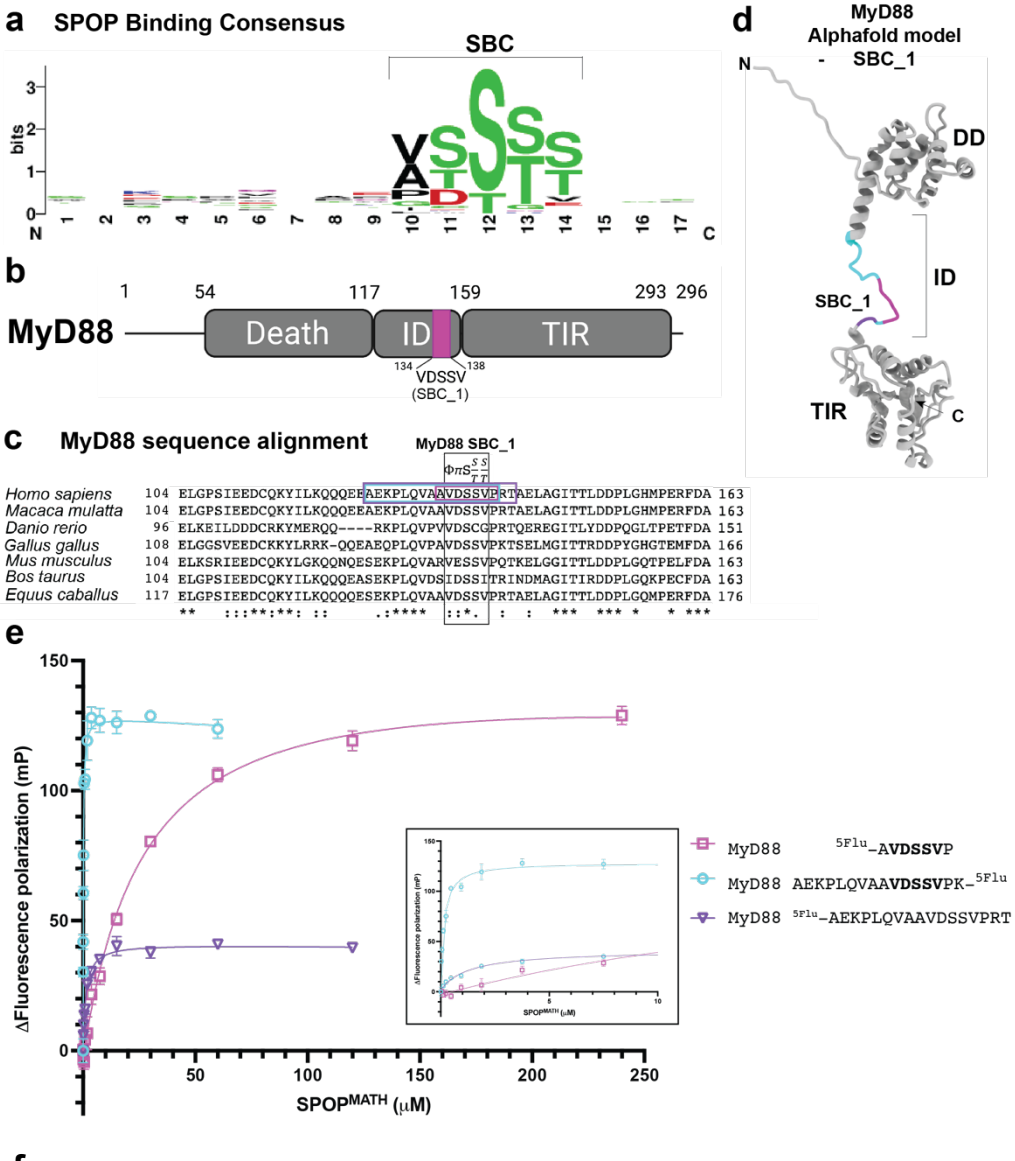
3.2.2 - Measuring SPOP^{MATH} binding to MyD88 peptides using a fluorescence polarization assay

In 2009 it had been shown that SPOP substrates contain a five-residue SBC which was identified as [**non-polar**]-[**polar**]-[**S**]-[**S/T**]-[**S/T**], with SPOP substrates typically containing more than one SBC in their protein sequence [55]. Since then, many new SPOP substrates have been identified and their SBCs reported (**Table 1.1**). Not all of the reported substrate SBCs have a S/T rich SBC and so a multiple sequence alignment using all currently known SPOP substrate SBC sequences was performed using WebLogo software (University of California, Berkeley) [195] to generate an updated SBC motif (**Table 1.1**, **Fig. 3.2 a**). The resulting motif agrees well with the original proposed SBC, but highlights the occurrence of unusual residues, particularly at position 5 of the SBC (**Fig. 3.2 a**).

It has been reported that MyD88 contains a single SBC (^{134}**VDSSV**^{138}) that is located in its intermediate domain (ID) (**Fig. 3.2 b, c, d**) [62]. This SBC is unusual as it contains a valine at position 5, which was the third most occurrent amino acid in the motif search at this position (**Fig. 3.2 a**). This unusual SBC is further confounded by the fact that MyD88 only has one known SBC, meaning that the potentially weaker SBC is not compensated for by multivalent binding to SPOP using several weaker SBCs. To investigate this further, the binding affinity between MyD88 and SPOP was measured using an FP assay.

5-carboxyfluorescein (^{5Flu})-labelled MyD88 peptides (either attached to the N-terminus or linked through an added C-terminal lysine) were used for binding affinity measurements with SPOP^{MATH}. The 7-mer peptide ^{5Flu}-A**VDSSV**P (**Fig. 3.2 c** magenta box; **Fig. 3.2 d**) includes the SBC along with an amino acid flanking either side. Analysis of the MyD88 protein sequence near the SBC showed a well-conserved region just N-terminal to the SBC (**Fig. 3.2 c**). Importantly, this conserved region is still within the unstructured intermediate domain, as predicted by AlphaFold (**Fig.**

3.2 d). To test whether these conserved MyD88 residues are also involved in SPOP^{MATH} binding, longer MyD88 peptides were also used for SPOP binding analysis (**Fig. 3.2 c**; **Fig. 3.2 d**). These fluorescently labelled MyD88 peptides were used to measure the binding affinity to the substate binding domain of SPOP, SPOP^{MATH}, using an FP assay. The SPOP^{MATH} truncation was chosen over the FL protein for this assay, as this monomeric form contains only one binding site per molecule, simplifying protein concentration measurements and K_d calculations.



f

Protein	Residue range	Peptide sequence	K_d (μM)
MyD88	133-139	^{5Flu}–A**VDSSV**P	32.25 ± 2.08
MyD88	125-141	^{5Flu}-AEKPLQVAA**VDSSV**PRT	1.94 ± 0.47

MyD88	125-139	AEKPLQVAA**VDSSV**PK^{-5Flu}	0.18 ± 0.02

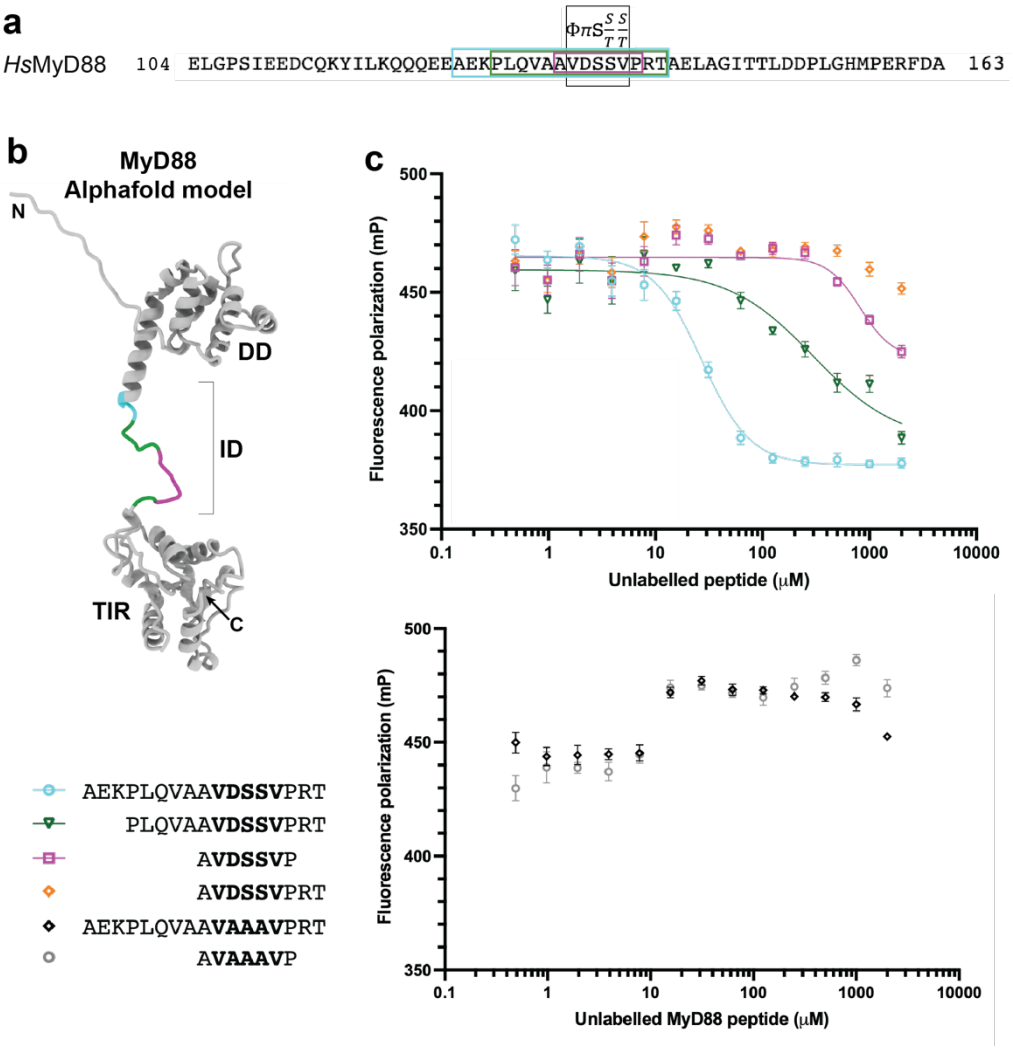
Figure 3.2 – Measuring SPOP^{MATH} binding to MyD88 peptides
a Motif search of known SPOP substrates using WebLogo software (University of California, Berkeley) [195]. Amino acids are coloured as follows. Black is hydrophobic. Green is polar. Red is acidic. Blue is basic. Purple is neutral. A list of proteins used to derive this consensus is shown in **Table 1.1**.
b Domain organisation of full-length MyD88. Numbers are residue numbers. Death is Death domain. ID is intermediate domain. TIR is Toll/interleukin-1 receptor domain. ^{134}VDSSV^{138} is the SPOP binding consensus (SBC).
c Multiple sequence alignment of MyD88 from various species aligned using the UniProt Clustal Omega sequence alignment tool [196]. SBC shown in black box. $\phi\pi S\frac{S}{T}\frac{S}{T}$ is the SBC where ϕ represents a non-polar amino acid and π represents a polar amino acid. ***** represents identical sequence. **:** represents conserved sequence. **.** represents semi-conserved sequence. MyD88 peptides used for fluorescence polarization assays shown as coloured boxes.
d AlphaFold2 [96] predicted structural model of MyD88, displayed as a cartoon figure in ChimeraX [187]. MyD88 peptide sequences used for fluorescent polarization assays are mapped onto the structure in their respective colours as shown in **c**.
e Fluorescence polarization assay measuring binding of SPOP^{MATH} and various 5-carboxyfluorescein (^{5Flu})-labelled MyD88 SBC peptides. SBC shown in bold. ΔFluorescence polarization is the normalised fluorescence polarization in milli-polarization (mP). Lines show the non-linear fit using Prism9's one site total binding analysis (GraphPad Prism version 9, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com). Error bars are standard error of the mean. Boxed is an enlarged view at lower SPOP^{MATH} concentrations.
f Table showing the dissociation constants (K_d) obtained in **e**. Bold residues are the SBC. Standard error of the mean is shown, calculated from 3 independent experiments.

Initially, peptides with either an N-terminal or C-terminal 5Flu label were tested for SPOP^{MATH} binding. However, some of the peptides showed visible precipitation in the assay buffer; therefore, the binding affinity could not be accurately measured due to potential concentration inaccuracies. Ultimately, the C-terminally labelled minimal SBC peptide could not be used for binding measurements with SPOP^{MATH}.

Strikingly, the MyD88 AEKPLQVAA**VDSSV**PK^{-5Flu} peptide showed a much higher affinity towards SPOP^{MATH} compared to the minimal MyD88 $^{5Flu-}$A**VDSSV**P peptide, with over a 100-fold difference in affinities of 0.18 ± 0.02 μM and 32.25 ± 2.08 μM respectively (**Fig 3.2 e, f**). This extended MyD88 peptide also represents the highest affinity SPOP binder measured so far. However, the N-terminally labelled peptide ^{5Flu}-AEKPLQVAAVDSSVPRT bound with an affinity of 1.94 ± 0.47 μM, which

is weaker affinity compared to the C-terminally labelled longer peptide but higher affinity compared to the shorter peptide (**Fig 3.2 e, f**). The ^{5Flu}-AEKPLQVAAVDSSVPRT peptide also showed a lower maximum FP value of ~40 mP compared to ~130 mP for the other two peptides. This could indicate the presence of a fluorophore-bound species which is lower molecular weight than the expected SPOP^{MATH}-peptide complex, such as peptide aggregates. Regardless, these data strongly suggests that the region N-terminal to the SBC is also involved in binding to SPOP^{MATH}.

To further investigate this extended MyD88 binding sequence, an FP peptide competition assay was performed. For this assay, the labelled MyD88 ^{5Flu}-A**VDSSV**P peptide competed for binding to SPOP^{MATH} with unlabelled MyD88 peptides of varying lengths (**Fig. 3.3 a, b**).



d

Protein	Residue range	Peptide sequence	IC_{50} (μM)
MyD88	133-139	A**VDSSV**P	~936.9 ± 119.1
MyD88	133-139	A**VAAAV**P	nb
MyD88	133-141	A**VDSSV**PRT	nb
MyD88	128-141	PLQVAA**VDSSV**PRT	~206.1 ± 90.7
MyD88	125-141	AEKPLQVAA**VDSSV**PRT	22.9 ± 9.2
MyD88	125-141	AEKPLQVAA**VAAAV**PRT	nb

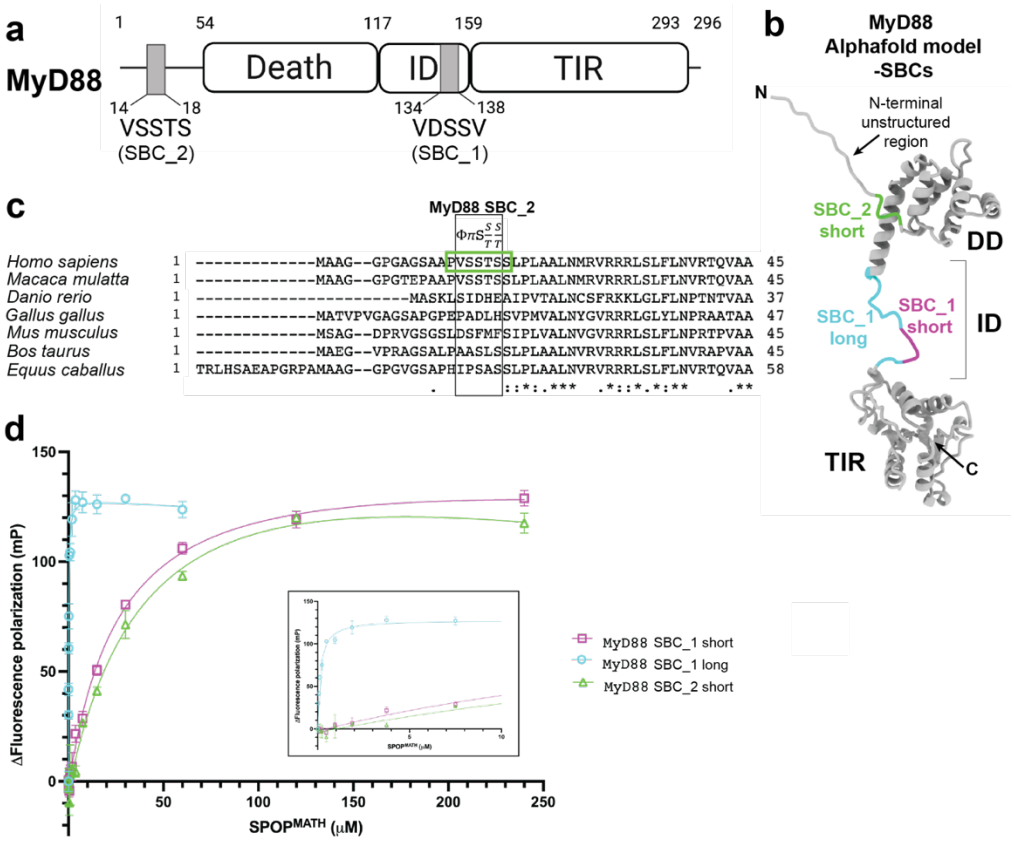
Figure 3.3 – Measuring binding between MyD88 peptides and SPOP^{MATH} using a fluorescence polarization peptide competition assay
a Protein sequence of *Homo sapiens* MyD88. SPOP-binding consensus (SBC) motif shown in black box. $\phi\pi S\frac{S}{T}\frac{S}{T}$ is the SBC where ϕ represents a non-polar amino acid and π represents a polar amino acid. Numbers are amino acid numbers. Unlabelled MyD88 peptide competitors used for fluorescence polarization assays shown as coloured boxes.
b AlphaFold2-predicted structural model of MyD88 displayed as a cartoon figure in ChimeraX. N is N-terminus. C is C-terminus. MyD88 peptide sequences used for fluorescent polarization assays are mapped onto the structure in their respective colours as shown in **a**.
c Peptide competition fluorescence polarization assay probing competitive binding of various unlabelled MyD88 peptides against a labelled MyD88 peptide (^{5Flu}-A**VDSSV**P) bound to SPOP^{MATH}. Top are wild-type peptides. Bottom are DSS to AAA mutant peptides. Lines show the non-linear fit using Prism-9's [inhibitor] vs response - variable slope analysis. Error bars are standard error of the mean.
d Table showing the half maximal inhibitory concentration (IC_{50}) values obtained in **c**. Bold residues are the SBC. nb is no measurable binding at the concentrations tested. Standard error of the mean is shown, calculated from 3 independent experiments.

In support of the previous FP data, the peptide competition assay showed that the longer unlabelled MyD88 peptide (^{125}AEKPLQVAA**VDSSV**PRT^{141}) was able to compete out the labelled peptide better than the other shorter peptides tested (**Fig. 3.3 c, d**). An intermediate-length peptide containing residues ^{128}PLQVAA**VDSSV**PRT^{141} also showed increased competitive binding compared to the minimal SBC peptide (^{133}A**VDSSV**P^{129}), but not to the degree of the AEKPLQVAA**VDSSV**PRT peptide. This suggests that residues within both the ^{125}AEK^{127} and the ^{128}PLQVA^{132} segments contribute to SPOP^{MATH} binding (**Fig. 3.3 c, d**). Interestingly, the

AEKPLQVAA**VDSSV**PRT peptide is predicted to encompass the entire unstructured ID of MyD88 (**Fig. 3.3 b**). In addition, two unlabelled MyD88 peptides with ^{135}DSS^{137} within the SBC mutated to AAA were also tested. Both of these peptides showed no binding to SPOP^{MATH} at the concentrations tested, suggesting that the SBC is critical for SPOP binding, which is not compensated for by additional binding residues (**Fig. 3.3 c, d**). Altogether, these analyses suggest that MyD88 interacts with SPOP^{MATH} using an extended binding motif that also includes residues N-terminal to the known five-residue SBC.

3.2.3 - Does MyD88 contain a second SBC?

Within the sequence of the unstructured N-terminus of MyD88 there is a sequence that resembles an SBC (^{14}**VSSTS**^{17}; SBC_2) (**Fig. 3.4 a, b**). Sequence alignment of MyD88 shows that this SBC_2 is only conserved in primates (**Fig. 3.4 c**). In previous SPOP-MyD88 studies, deletion of this SBC_2 motif and subsequent co-purifications and ubiquitylation assays showed that SPOP was still able to bind to and ubiquitylate MyD88$^{\Delta SBC_2}$ at levels similar to WT MyD88 [119]. In the same study, deletion of the SBC_1 (^{134}**VDSSV**^{138}) motif prevented any detectable SPOP binding and ubiquitylation of MyD88, suggesting that SBC_1 was a genuine SBC whereas SBC_2 was not [119]. To study this second putative MyD88 SBC_2 using a more minimal approach, the interaction of SPOP^{MATH} with a 7-mer MyD88 P**VSSTS**SK^{-5Flu} peptide was tested.



e

Protein	Residue range	Peptide sequence	K_d (μM)
MyD88 (*SBC_1 short*)	133-139	^{5Flu}–A**VDSSV**P	32.25 ± 2.08
MyD88 (*SBC_1 long*)	125-139	AEKPLQVAA**VDSSV**PK^{-5Flu}	0.18 ± 0.02
MyD88 (*SBC_2 short*)	13-18	^{5Flu}–P**VSSTS**S	30.27 ± 14.47

Figure 3.4 – Measuring MyD88 SBC_2 peptide binding to SPOP^{MATH}
a Domain organisation of full-length MyD88. Numbers are residue numbers. Death is Death domain. ID is intermediate domain. TIR is Toll/interleukin-1 receptor domain. ^{134}VDSSV^{138} is the known SPOP binding consensus (SBC). ^{14}VSSTS^{17} is the putative second SBC.
b Alphafold2-predicted structural model of MyD88 displayed as a cartoon figure in ChimeraX. MyD88 peptide sequence used for fluorescent polarization assays is mapped onto the structure, as shown in **c** and **d**.
c Multiple sequence alignment of MyD88 from various species aligned using the UniProt sequence alignment tool. SBC shown in black box. $\phi\pi S\frac{S}{T}\frac{S}{T}$ is the SBC where ϕ represents a non-polar amino acid and π represents a polar amino acid. ***** represents identical sequence. **:** represents conserved sequence. **.** represents semi-conserved sequence. MyD88 SBC_2 peptide used for fluorescence polarization assays boxed in green.

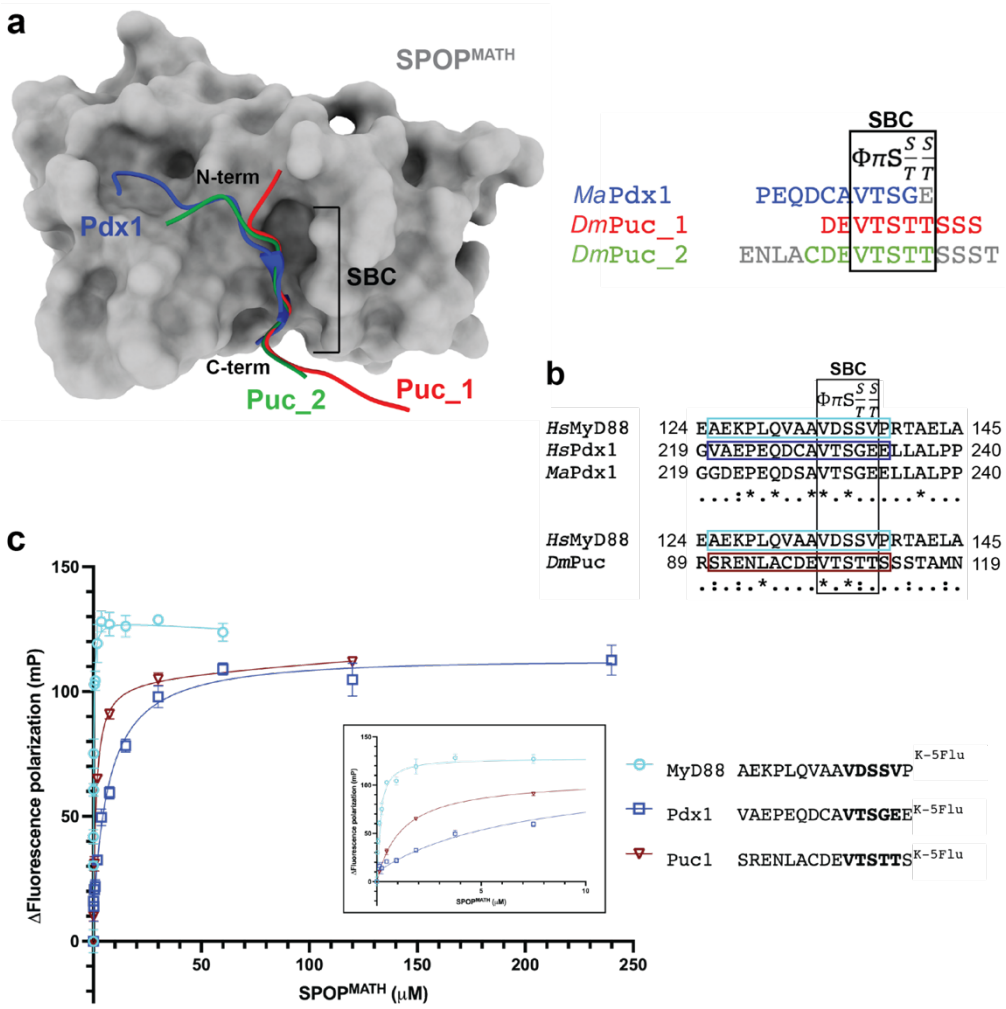
d Fluorescence polarization assay probing binding between a range of $SPOP^{MATH}$ concentrations and various 5-carboxyfluorescein-labelled (^{5Flu}) MyD88 SBC peptides. Sequences of peptides are shown in **e**. ΔFluorescence polarization is the normalised fluorescence polarization in milli-polarization (mP). Lines show the non-linear fit using Prism9's one site total binding analysis. Error bars are standard error of the mean. Boxed is a zoomed in view at lower $SPOP^{MATH}$ concentrations.
e Table showing the dissociation constants (K_d) obtained in **d**. Bold residues are the SBC. Standard error of the mean is shown, calculated from 3 independent experiments. SBC_2 short was calculated from 2 independent experiments.

The MyD88 SBC_2 short peptide (^{5Flu}-P**VSSTS**S) and the MyD88 SBC_1 short peptide (^{5Flu}-A**VDSSV**P) showed comparable binding to $SPOP^{MATH}$, with K_ds of 30.27 ± 14.47 and 32.25 ± 2.08 respectively. This is in contrast with previous reports that MyD88 SBC_2 was dispensable for SPOP binding [119]. It is therefore feasible that this ^{13}P**VSSTS**S^{19} region of MyD88 may act as an additional SBC, at least *in vitro*. Further experiments would be needed to validate whether MyD88 uses both SBCs within a single protomer to engage a SPOP oligomer.

3.2.4 - Comparison of extended MyD88 SBC_1 to other SPOP substrate SBCs

To date, Pdx1 is the only SPOP substrate reported to have an extended SPOP binding site, with additional residues N-terminal to the SBC able to interact with $SPOP^{MATH}$ [63, 99] (**Fig. 3.5 a**). Comparison of the Pdx1 and MyD88 protein sequences showed some degree of similarity in the region N-terminal to the SBC (**Fig. 3.5 b**). Another SPOP substrate, Puc, is a SPOP substrate with the highest known affinity of binding to SPOP. The binding affinity of a Puc SBC peptide (^{93}LACDE**VTSTT**SSSTA^{107}) with $SPOP^{MATH}$ was measured with a K_d of 3.70 ± 0.03 μM, whereas the majority of SBCs have affinities in the mid to high μM range [55]. Notably, a crystal structure of Puc bound to $SPOP^{MATH}$ did not exhibit any extended binding N-terminal to the SBC (**Fig. 3.5 a**). To measure how the binding affinity of MyD88 compares to the extended binder Pdx1 and the high affinity binder Puc, peptides for Pdx1 and Puc encompassing

the same region around the SBC as the MyD88 SBC_1 extended peptide were tested using an FP assay (**Fig. 3.5 b**).



d

Protein	Residue range	Peptide sequence	K_d (μM)
MyD88	125-139	AEKPLQVAA**VDSSV**PK^{-5Flu}	0.18 ± 0.02
Pdx1	220-234	VAEPEQDCA**VTSGE**EK^{-5Flu}	6.15 ± 1.05
Puc	90-104	SRENLACDE**VTSTT**SK^{-5Flu}	1.190 ± 0.003

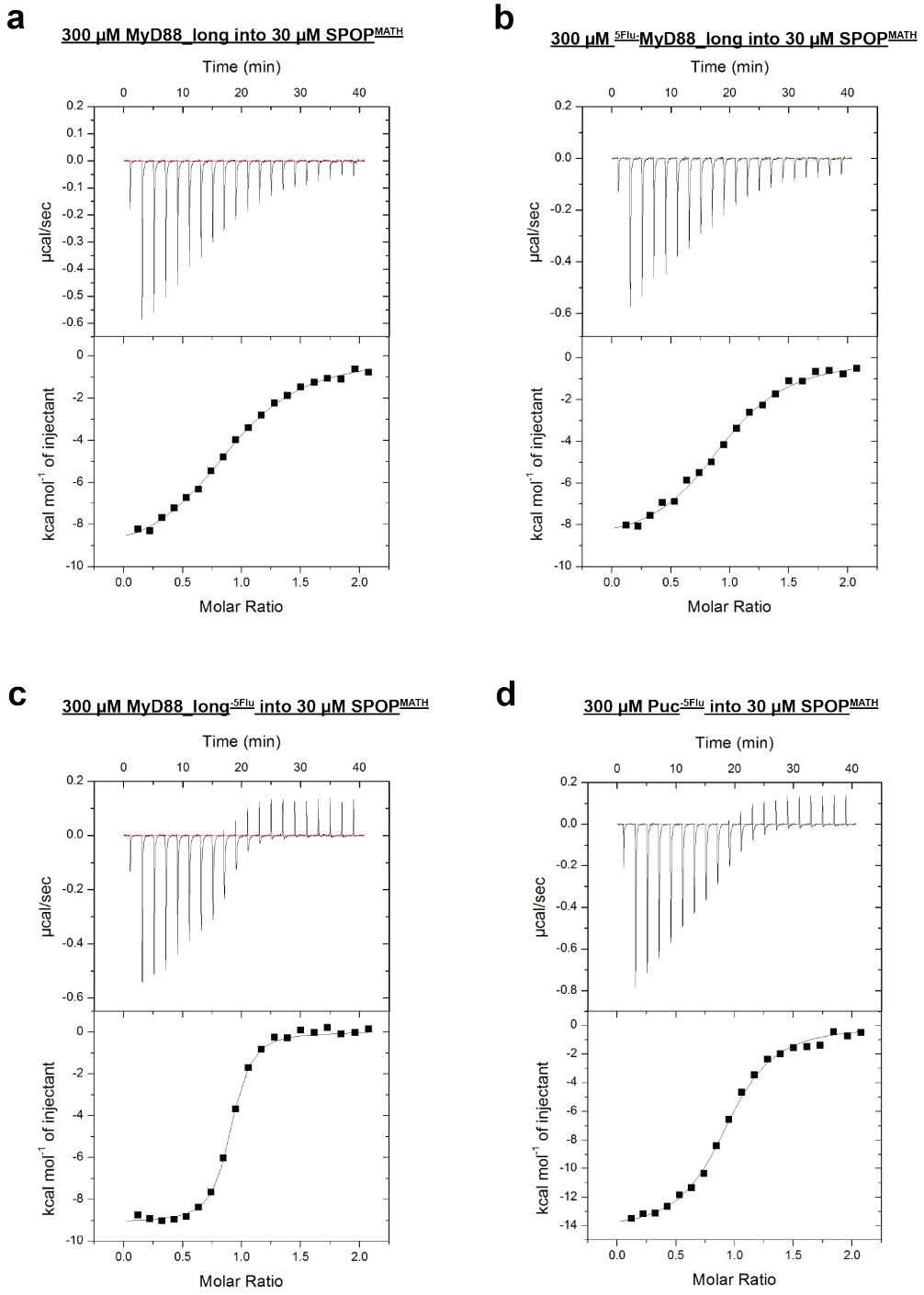
Figure 3.5 – Measuring SPOP^{MATH} binding to SPOP substrate peptides
a *Left* - comparison of substrate peptide-bound SPOP structures. SPOP^{MATH} shown in grey. SPOP substrate peptides shown as colours. Pdx1 PDB ID 6F8F [99]. Puc_1 PDB ID 3HQL[55]. Puc_2 PDB ID 3HQL [55]. *Right* – Protein sequences of SPOP binding peptides coloured as shown in left. Grey amino acids were included in the peptide but not built into model. Black box denotes SPOP binding consensus (SBC) sequence. $\phi\pi S\frac{S}{T}\frac{S}{T}$ is the SBC where ϕ represents a non-polar amino acid and π represents a polar amino acid.

b Sequence alignment of the SBC regions of *Homo sapiens* (Human; *Hs*) MyD88, *Hs* and *Mesocricetus auratus* (Golden hamster; *Ma*) Pdx1 and *Drosophila melanogaster* (Fruit fly; *Dm*) Puckered (Puc). Alignments generated using EMBOSS Stretcher [197]. SBC shown as black box. ***** represents identical sequence. **:** represents conserved sequence. **.** represents semi-conserved sequence.
c Fluorescence polarization assay probing binding between a range of $SPOP^{MATH}$ concentrations and various 5-carboxyfluorescein-labelled (^{5Flu}) peptides. SBC shown in bold. ΔFluorescence polarization is the normalised fluorescence polarization in milli-polarization (mP). Lines show the non-linear fit using Prism9's one site total binding analysis. Error bars are standard error of the mean.
d Table showing the average dissociation constants (K_d) obtained in **c**. Bold residues are the SBC. Standard error of the mean is shown, calculated from 2 independent experiments.

The Pdx1 peptide displayed a higher binding affinity to $SPOP^{MATH}$ compared to most known SPOP substrates, with a K_d of 6.15 ± 1.05 μM (**Fig. 3.5 c**, **d**). The Puc peptide also displayed relatively tight binding, with a K_d of 1.190 ± 0.003 μM (**Fig. 3.5 c**, **d**). This K_d is similar to that reported by Zhuang *et al.* [55], although different peptide boundaries were used in this experiment, which may explain the slight difference in K_d. Notably, the C-terminally labelled MyD88 peptide showed a much higher binding affinity than both Pdx1 and Puc (**Fig. 3.5 c, d**). Meanwhile the N-terminally labelled MyD88 peptide showed a comparable K_d as Puc (**Fig. 3.2 e, f**). Despite the discrepancy between the K_ds of the N-terminally and C-terminally labelled MyD88 peptides, both values are at least as high affinity as Puc.

3.2.5 - Isothermal titration calorimetry measurements of MyD88 and Puc binding to $SPOP^{MATH}$

To investigate which MyD88 peptide binding measurement represents a more real K_d measurement, isothermal titration calorimetry (ITC) was used. The binding to $SPOP^{MATH}$ was measured by titrating the N-terminal labelled, C-terminal labelled and unlabelled MyD88 SBC_1 long peptides. Additionally, binding of the Puc-^{5Flu} peptide was also measured for comparison.


a
300 µM MyD88_long into 30 µM SPOP^MATH
b
300 µM 5Flu-MyD88_long into 30 µM SPOP^MATH
c
300 µM MyD88_long-5Flu into 30 µM SPOP^MATH
d
300 µM Puc-5Flu into 30 µM SPOP^MATH
Time (min)
µcal/sec
kcal mol-1 of injectant
Molar Ratio

e

	MyD88_long	**^{5Flu}-MyD88_long**	**MyD88_long^{-5Flu}**	**Puc^{-5Flu}**
Peptide sequence	AEKPLQV AAVDSSVPRT	^{5Flu}-AEKPLQV AAVDSSVPRT	AEKP LQVAAV DSSVPK^{-5Flu}	SRENLAC DEVTSTT SK^{-5Flu}
N (sites)	0.94 ± 0.01	0.96 ± 0.02	0.87 ± 0.01	0.93 ± 0.01
K_a (M-1)	$2.51x10^5$ ± $1.58x10^4$	$3.41x10^5$ ± $3.44x10^4$	$4.21x10^6$ ± $4.40x10^5$	$7.30x10^5$ ± $6.48x10^4$
K_d (μM)	3.98	2.93	0.24	1.37
ΔH (cal/mol)	-9761.00 ± 159.80	-8998.00 ± 197.90	-9080.00 ±75.87	$-1.44x10^4$ ± 198.40
ΔS (cal/mol/°)	-8.03	-4.86	-0.15	-21.30

Figure 3.6 – Isothermal titration calorimetry (ITC) measurements of MyD88 and Puc peptides binding to SPOP^{MATH}
For **a**-**d**, peptide sequences are shown in **e**. Top panel is the raw ITC data with the baseline shown as a red line. The bottom panel is the integrated heat peaks plotted against the molar ratio.
a Binding isotherm of 300 μM MyD88_long peptide titrated into 30 μM SPOP^{MATH}.
b Binding isotherm of 300 μM 5-carboxyfluorescein (^{5Flu})-MyD88_long peptide (N-terminally labelled with 5Flu) titrated into 30 μM SPOP^{MATH}.
c Binding isotherm of 300 μM MyD88_long-^{5Flu} peptide (C-terminally labelled with 5Flu) titrated into 30 μM SPOP^{MATH}.
d Binding isotherm of 300 μM Puc-^{5Flu} peptide (C-terminally labelled with 5Flu) titrated into 30 μM SPOP^{MATH}.
e Table showing ITC parameters from **a**-**d**. N is the stoichiometry of binding. K_a is the association constant. K_d is the dissociation constant calculated by taking the inverse of K_a. ΔH is the enthalpy change. ΔS is the entropy change. Values are from a single experiment. Error values are the fitting error.

Overall, the K_d values measured in the ITC experiment agree well with those measured using the FP assay (**Fig. 3.5 d**, **Fig. 3.6 e**). The use of

ITC allowed for the K_d measurement of the unlabelled MyD88 peptide binding to SPOP^{MATH}, which was 3.98 μM. This value more closely resembles the value obtained for the N-terminal labelled MyD88 peptide (2.93 μM), rather than the C-terminal labelled peptide (240 nM) (**Fig. 3.6 e**). It is therefore likely that the addition of the fluorophore at the C-terminus of the MyD88 peptide is contributing to binding to SPOP, resulting in a 10-fold increase in affinity. Although, this phenomenon does not seem to be observed in the case of Puc-^{5Flu} binding to SPOP^{MATH}, which reported K_d values that were consistent with previous studies [55]. Overall, the binding affinity of 3.98 μM for the MyD88 SBC_1 long peptide binding to SPOP^{MATH} still represents a relatively high binding affinity, compared to other SPOP substrates.

3.3 - Determining the X-ray crystallography structure of a MyD88 peptide bound to SPOP^{MATH}

To visualise the MyD88 peptide-SPOP^{MATH} interaction and infer the structural basis of MyD88's extended binding, crystallization trays were set up. SPOP^{MATH} was mixed with a 3x molar excess of MyD88 peptide (^{125}AEKPLQVAA**VDSSV**PRT^{141}). Plate-shaped crystals appeared after 11 days, and X-ray diffraction data was collected using the Diamond (Oxfordshire, U.K.) I24 microfocus beamline (full data collection and processing parameters in **Table 2.3**). The crystal diffracted to 1.7 Å and data was later truncated to 2.0 Å (**Fig. 3.7 a, b**). The crystal had space group $P2_1\ 2_1\ 2_1$ with one copy in the asymmetric unit, assuming 40% solvent content. The structure was solved by molecular replacement using an apo SPOP^{MATH} structure (PDB ID: 7KPI) [63]. The resulting electron density showed a good fit with the known SPOP^{MATH} structure, with well resolved additional electron density in the substrate binding region of SPOP, which represented the MyD88 peptide (**Fig. 3.7 c, d**). The MyD88 peptide was built into the density *de novo* and after several rounds of model building and refinement, the structure had an R_{work}/R_{free} of 0.187/0.232 (**Table 2.3, Fig. 3.7 d, e**).

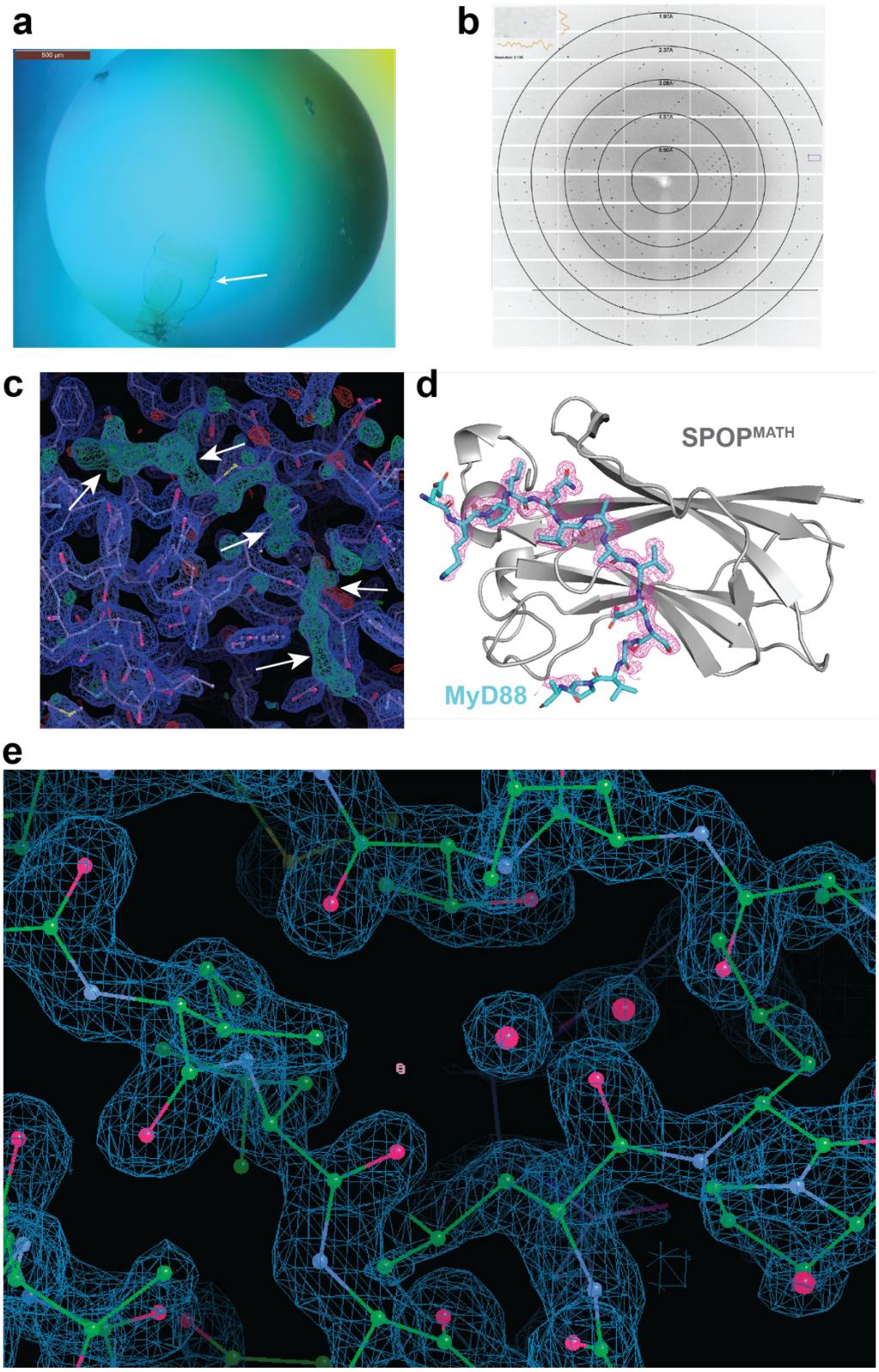


Figure 3.7 – X-ray crystallography data collection and model building of the SPOP^{MATH}-MyD88 peptide crystal structure

a Image of the crystal used for X-ray crystallography data collection. White arrow points to crystal.

b Diffraction pattern from the X-ray crystallography data collection. Resolution rings are shown.

c Initial electron density after molecular replacement with a $SPOP^{MATH}$ structure (PDB ID: 7KPI [63]; purple sticks). White arrows highlight F_o-F_c density contoured at 2.5σ (green mesh), which corresponds to the MyD88 peptide.
d Omit map showing density of the MyD88 peptide. F_o-F_c density contoured at 1.0σ.
e Representative image of final 2F_o-F_c electron density contoured at 1.0σ (blue mesh) and the model shown as green sticks.

The crystal structure shows that almost the full 17-mer MyD88 peptide is resolved and involved in SPOP binding, consistent with the FP data (**Fig. 3.8 a**, **b).** Only the very N-terminal A-125 and C-terminal T-141 are missing from the structure, likely due to the flexibility of these regions (**Fig. 3.8 a**). The MyD88 peptide SBC binds to β-7 of $SPOP^{MATH}$, as seen in other $SPOP^{MATH}$-substrate peptide structures however, there are additional interactions directly N- and C-terminal to the SBC that are unique to MyD88.

The crystal structure recapitulates the importance of the SBC, with six hydrogen bonds observed between the side chains of MyD88 ^{135}DSS^{137} and SPOP residues Y-87, K-129 and D-130, with additional MyD88 main chain or water-mediated hydrogen bonding in this region (**Fig. 3.8 c-i**). In addition, SPOP W-131 and F-133 are proximal to MyD88 V134 and are the main hydrophobics in this region (**Fig. 3.8 c-i**). These interactions are similar to what is observed with other SPOP-substrate structures.

Directly N-terminal to the SBC is the linker region containing the small hydrophobic MyD88 residues ^{131}VAAV^{134} (**Fig. 3.8 b**). This region contains a vast network of water-mediated main chain hydrogen bonding, which is not observed in other $SPOP^{MATH}$-substrate peptide structures, including the extended binder Pdx1 (**Fig. 3.8 c-ii**).

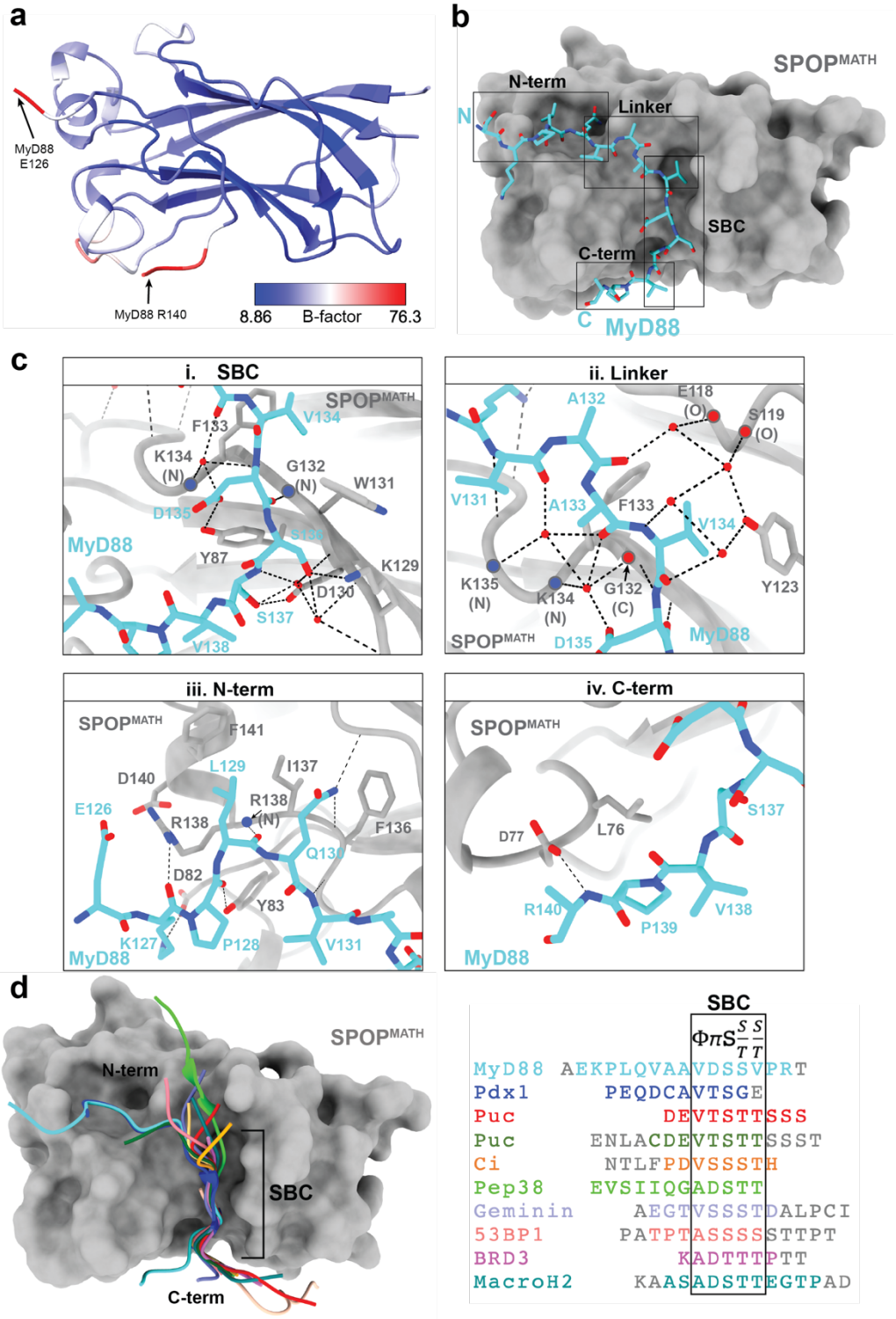


Figure 3.8 – MyD88 bound SPOP^{MATH} structure
a 2.0 Å SPOP^{MATH} and MyD88 peptide (^{125}AEKPLQVAAVDSSVPRT^{141}) X-ray crystal structure cartoon coloured by B-factor.
b X-ray crystal structure as in **a**. The MATH domain of SPOP is shown as a surface representation in grey. The MyD88 peptide is shown as sticks in cyan. Black boxes denote regions of the MyD88 peptide. N-term is N-terminus of MyD88 peptide. SBC is SPOP Binding Consensus. C-term is C-terminus of MyD88 peptide.
c Close up view of the structure as shown in **b**. Hydrogen bonds shown as black dashed lines.

d *Left* - comparison of substrate peptide-bound SPOP^{MATH} structures. SPOP^{MATH} shown as grey surface representation. SPOP substrate peptides shown as colours. *Right* – Protein sequences of SPOP binding peptides coloured as shown in left. Grey amino acids were included in the peptide but not built into model. MyD88 is from this study. Pdx1 PDB ID: 6F8G [99]. Puc_1 PDB ID: 3IVV [55]. Puc_2 PDB ID: 3HQL [55]. Ci PDB ID: 3HQM [55]. Pep38 PDB ID: 7D3D [198]. Geminin PDB ID: 7KLZ [65]. 53BP1 PDB ID: 7LIN [66]. BRD3 PDB ID: 6I7A [199]. MacroH2A PDB ID: 3HQH [55]. Black box is SPOP binding consensus (SBC) sequence. $\phi\pi S\frac{S}{T}\frac{S}{T}$ is the SBC where ϕ represents a non-polar amino acid and π represents a polar amino acid.

At the N-terminal region of the MyD88 peptide there are various electrostatic and hydrophobic interactions (**Fig. 3.8 c-iii**). Of note, hydrogen bonding is observed between MyD88 Q-130 with the carbonyl groups of SPOP K-115 and F-136, as well as between MyD88 K-127 and SPOP D-82 (**Fig. 3.8 c-iii**). Additionally, MyD88 L-129 and Q-130 sit within a hydrophobic pocket of SPOP (F-136, I-137 and F-141) with the adjacent MyD88 P-128 and V-131 likely helping to anchor this region of the peptide to SPOP^{MATH} (**Fig. 3.8 c-iii**). Interestingly, a SPOP D140H mutant showed a reduced SPOP-MyD88 interaction and reduced the ubiquitylation of MyD88 [119]. SPOP D-140 sits adjacent to the binding of this N-terminal region of MyD88, perhaps explaining the loss of SPOP-MyD88 interaction when mutated from D140H (**Fig. 3.8 c-iii**). The *Ma*Pdx1-SPOP^{MATH} crystal structure also shows that residues N-terminal to the Pdx1 SBC are involved in SPOP binding (**Fig. 1.8**, **Fig. 3.8 d**) [99]. Comparison of two structures shows the MyD88 peptide superimposes fairly well with the *Ma*Pdx1 peptide structure, with an RMSD of 3.46 (**Fig. 3.8 d**). Interestingly, Pdx1 also has a glutamine bound in the same position as MyD88 Q-130, which also has two hydrogen bonds to SPOP (**Fig. 1.8**).

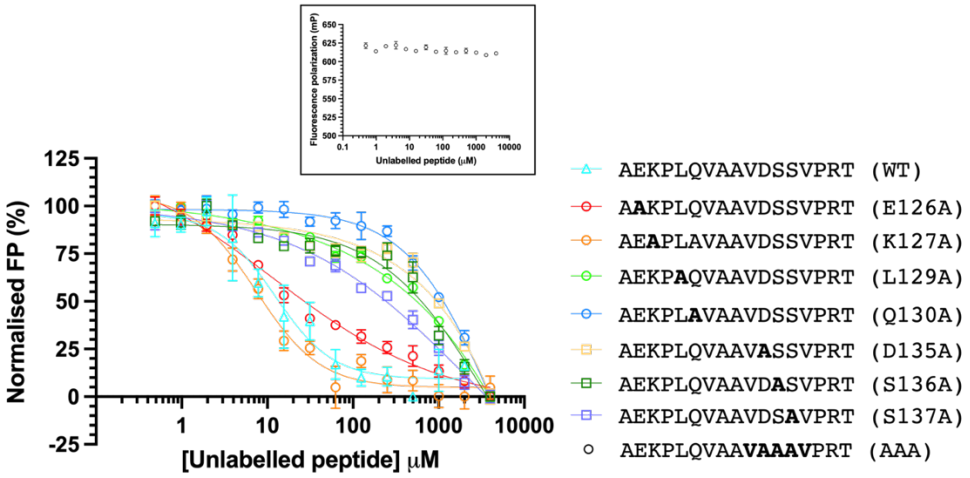
The C-terminus of the MyD88 peptide is unusual in comparison to other SPOP^{MATH}-substrate structures. It curves back towards SPOP^{MATH} and engages in a different conformation in respect to other substrate peptides, giving the MyD88 peptide a horse shoe-like shape (**Fig. 3.8 b, d**). In other SPOP-peptides structures, the polar residue of SBC position

five typically hydrogen bonds with SPOP K-129 or D-130, directing the subsequent residues C-terminal to the SBC in this direction (**Fig. 1.7**). In MyD88, this polar residue is replaced with a valine (V-138), preventing this hydrogen bonding (**Fig. 3.8 c-iv**). C-terminal to V-138, the N^1 of MyD88 R-140 hydrogen bonds to SPOP D-77, anchoring the C-terminus of the MyD88 peptide away from SPOP K-129 or D-130 (**Fig. 3.08 c-iv**). However, the B-factor for the C-terminal region of the MyD88 peptide is relatively high, likely indicating a lower occupancy or high flexibility of this region (**Fig. 3.08 a**).

3.4 - Structure-guided mutations of MyD88

To validate the interactions observed in the crystal structure, and to probe which MyD88 residues are critical for SPOP binding, point mutant MyD88 peptides were tested for their binding to $SPOP^{MATH}$ using an FP peptide competition assay.

a



b

MyD88 peptides[125-141]	Mutation	IC_{50} (μM)
AEKPLQVAA**VDSSV**PRT	WT	12.32
A**A**KPLQVAA**VDSSV**PRT	E126A	11.59
AE**A**PLQVAA**VDSSV**PRT	K127A	7.37
AEKP**A**QVAA**VDSSV**PRT	L129A	-
AEKPL**A**VAA**VDSSV**PRT	Q130A	~4,078
AEKPLQVAA**V**A**SSV**PRT	D135A	-
AEKPLQVAA**VD**A**SV**PRT	S136A	~3,538
AEKPLQVAA**VDS**A**V**PRT	S137A	~2,118
AEKPLQVAA**VAAAV**PRT	[135]AAA[137]	nb

c

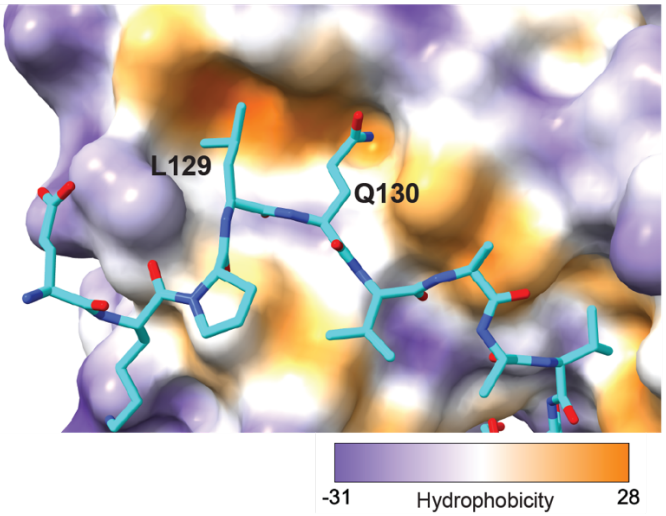


Figure 3.9 – Testing structure guided MyD88 mutants for binding to SPOP[MATH]

a Peptide competition fluorescence polarization assay probing competitive binding of various unlabelled MyD88 mutant peptides against a labelled MyD88 peptide (^{5Flu}-A**VDSSV**P) bound to SPOP^{MATH}. Lines show the non-linear fit using Prism-9's [inhibitor] vs response - variable slope analysis. Error bars are standard error of the mean. Data is normalised so that the maximum mP value represents 100% and the minimum mP value represents 0%.
b Table showing IC_{50} values measured for each peptide, as in **a**. Values are from a single experiment. nb is no measurable binding at the concentrations tested. IC_{50}s with a (-) could not be confidently calculated.
c SPOP^{MATH} shown as surface representation, coloured by hydrophobicity. MyD88 peptide shown in cyan. The sidechains of MyD88 leucine-129 and glutamic acid-130 sit in hydrophobic pocket of SPOP.

As expected, point mutation along SBC residues ^{135}DSS^{137} impairs binding to SPOP^{MATH}, with D135A and S136A showing the most marked effects (**Fig. 3.9 a, b**). This is likely due to the disruption of the hydrogen bonding network with SPOP^{MATH}, as observed in the crystal structure (**Fig. 3.9 c-i**). Interestingly, L129A and Q130A peptides also exhibited a reduced binding, at similar levels to the ^{135}DSS^{137} point mutants (**Fig. 3.9 a, b**). When looking at the crystal structure, L-129 and Q-130 sit within a hydrophobic pocket of SPOP^{MATH} (F-136, I-137 and F-141), with Q-130 also forming many charged interactions (**Fig. 3.9 c; Fig. 3.9 c-iii**). The mutation from the large hydrophobic leucine to the smaller hydrophobic alanine would likely lessen the hydrophobic interactions with SPOP, greatly reducing the IC_{50}, despite still maintaining a WT SBC (**Fig. 3.9 a, b**). Similarly, mutation from the large polar Q-130 to alanine showed the largest decrease in binding out of all of the point mutants tested, with an IC_{50} of ~4,078 μM (**Fig. 3.9 a, b**). This mutation would disrupt the two hydrogen bonds between MyD88 Q-130 and SPOP^{MATH}, as observed in the crystal structure (**Fig. 3.8 c-iii**). It is therefore proposed that MyD88 ^{129}LQ^{130} acts as a secondary binding site in addition to the SBC that contributes to high affinity SPOP^{MATH} binding. However, the affinity of this second binding site alone is not enough to rescue binding of the SBC ^{135}DSS^{137} to AAA mutant peptide, at the concentrations measured (**Fig. 3.9 a**).

3.4.1 - Testing the effect of MyD88 ^{129}LQ^{130} mutants in cells

With the identification of MyD88 ^{129}LQ^{130} as critical SPOP binding residues by binding assays and an X-ray crystal structure, these residues were then tested for SPOP binding in a cellular context using the full-length proteins by Dr Mukul Mishra, Dr Luca Busino lab, University of Pennsylvania.

Single and double mutation of MyD88 ^{129}LQ^{130} to alanine showed a complete inhibition of SPOP binding by co-immunoprecipitation from HEK293T cells (**Fig. 3.10 a**). Conversely, mutation of SPOP residues (Y83A, Y83F, R138E) proximal to the MyD88 ^{129}LQ^{130} binding site also showed a reduced binding of SPOP to MyD88 (**Fig. 3.8 c iii**, **Fig. 3.10 b**). Treatment of HEK293T cells expressing either wildtype MyD88 or LQ to AA MyD88 mutant with the translation blocking drug cycloheximide showed a stabilisation of MyD88 LQ to AA mutant, compared to wildtype (**Fig. 3.10 c**). This suggests that SPOP binding and SPOP-mediated ubiquitylation and degradation of MyD88 LQ to AA was unable to take place due to these mutations. Together these data suggest that MyD88 ^{129}LQ^{130} is important for SPOP binding and that mutation of these residues protects MyD88 from SPOP-mediated ubiquitylation and degradation.

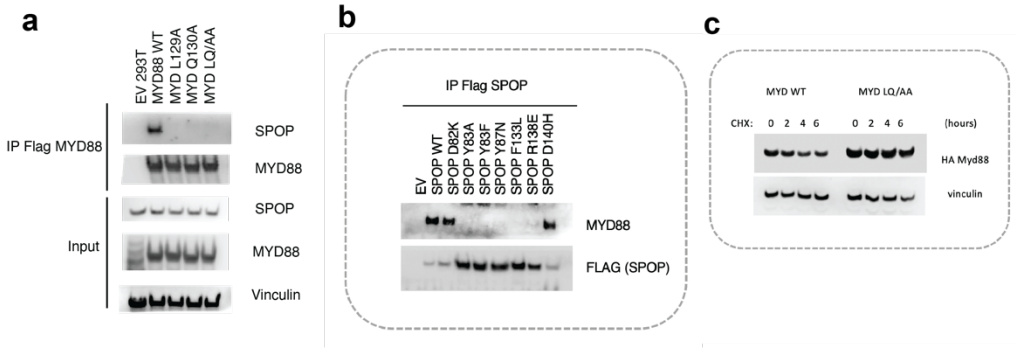


Figure 3.10 – Testing MyD88 and SPOP extended binding region mutants in cells

Data generated by Dr Mukul Mishra, Dr Luca Busino lab.

a Immunoprecipitation of FLAG-MyD88 from HEK293T cells expressing either wildtype (WT) FLAG-MyD88, FLAG-MyD88 mutants. EV is an empty vector control. Vinculin is a loading control.

b Immunoprecipitation of FLAG-SPOP from HEK293T cells expressing either WT SPOP or SPOP mutants. SPOP^{D82}, SPOP^{Y83}, SPOP^{R138}, SPOP^{D140} are

residues proximal to the MyD88 extended binding region. $SPOP^{Y87}$ and $SPOP^{F133}$ are proximal to the MyD88 SBC.
c HEK293T cells expressing either WT MyD88 or MyD88 $^{129}LQ^{130}$ to AA mutant were treated with cycloheximide (CHX). Samples are from 0, 2, 4 and 6 hours after CHX treatment.

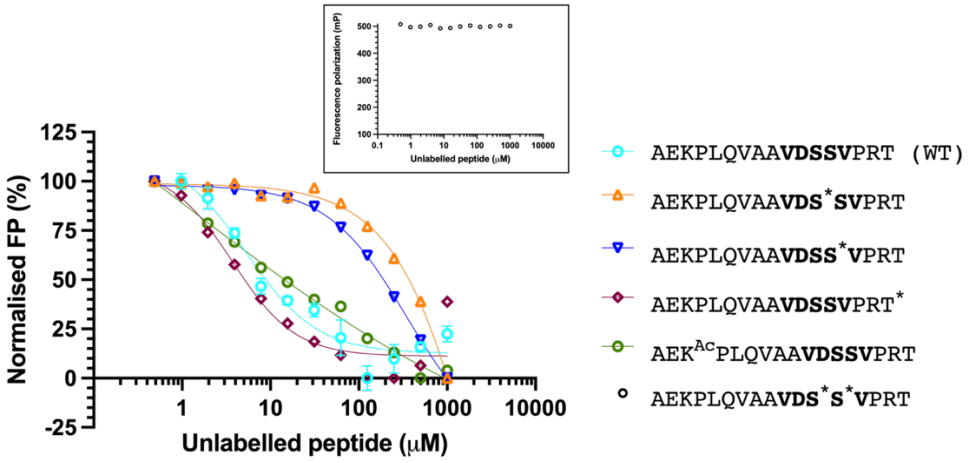
MyD88 is frequently mutated in Diffuse Large B-Cell Lymphoma (DLBCL), with the most prevalent mutation in this cancer, L252P (frequently referred to as L265P), occurring in the TIR domain [200]. This mutation is associated with the constitutive assembly of the Myddosome in the absence of TLR activation, resulting in continuous Myddosome signalling. In addition to the L252P mutation, there have been reports of other DLBCL-associated MyD88 mutations of unknown significance. Within the ID of MyD88, mutations P128T and Q130L (reported as P141T and Q143L) [201] as well as P128S, V131L and S136I (reported as P141S, V144L and S149I) [202] have been identified in DLBCL. While S-136 lies within the SBC, the other mutations are all located in the extended binding region, that was also shown to also be important for SPOP binding. It is tempting to speculate that these DLBCL-associated MyD88 mutants cause disease by preventing SPOP binding, thereby preventing MyD88 ubiquitylation and degradation. This inability to clear MyD88, either at the free MyD88 or Myddosome level, may promote uncontrolled Myddosome signalling, leading to cancer progression.

3.4.2 - Post-translational modifications of the MyD88 ID

An analysis of MyD88 post-translational modifications by mass spectrometry highlighted several modifications occurring within the ID of MyD88 (analysis by the Dr Luca Busino lab). First, S-136 and S-137 of the SBC were shown to be phosphorylated. Previous studies had shown that phospho-mimicking mutations of these two residues prevented SPOP binding [62]. In addition, the mass spectrometry analysis also identified acetylated K-127 and phosphorylated T-141. To explore the effect of these MyD88 post-translation modifications on $SPOP^{MATH}$

binding, MyD88 peptides containing these post-translational modifications were used in an FP peptide competition assay.

a



b

MyD88 peptides$^{125\text{-}141}$	Modification	IC_{50} (μM)
AEKPLQVAA**VDSSV**PRT	WT	5.26
AAKPLQVAA**VDS*SV**PRT	phospho S-136	~3,701.00
AEAPLQVAA**VDSS*V**PRT	phospho S-137	~349.30
AEKGLQVAA**VDSSV**PRT*	phospho T-141	3.72
AEK^{Ac}PAQVAA**VDSSV**PRT	acetyl K-127	1.20
AEKPLAVAA**VDS*S*V**PRT	phospho S-136 and S-137	nb

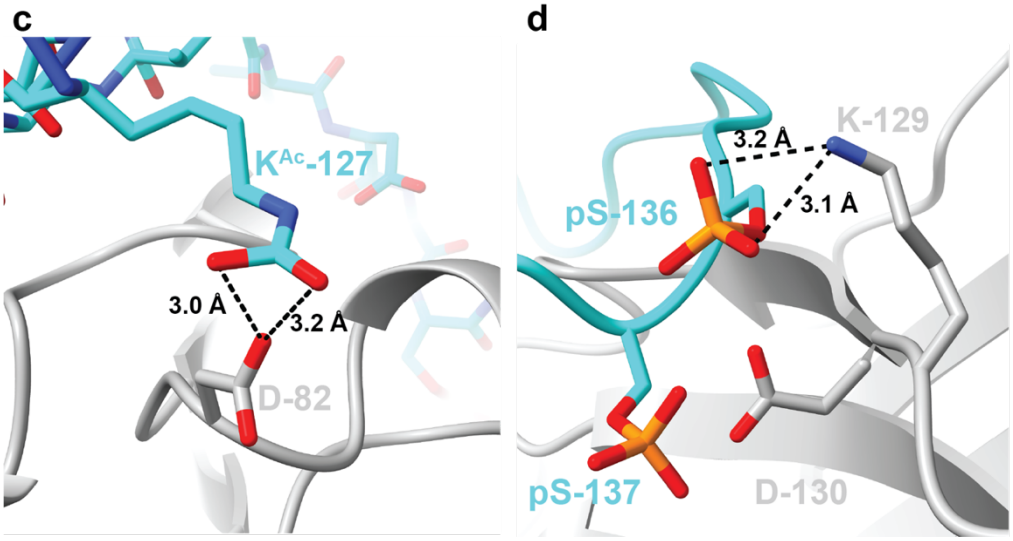


Figure 3.11 – Testing post-translational modifications of the MyD88 SPOP-binding region

a Peptide competition fluorescence polarization assay probing competitive binding of various unlabelled MyD88 post-translational modification peptides against a labelled MyD88 peptide (^{5Flu}-A**VDSSV**P) bound to SPOP^{MATH}. Lines show the non-linear fit using Prism-9's [inhibitor] vs response - variable slope analysis. Error bars are standard error of the mean. Data is normalised so that the maximum mP value represents 100% and the minimum mP value represents 0%.
b Table of IC_{50} values calculated as in **a**. nb is no binding. Data is from one independent experiment.
c MyD88 acetyl-lysine (K^{Ac})-127 built into the SPOP^{MATH}-MyD88 peptide crystal structure. MyD88 in cyan. SPOP^{MATH} in grey. Dashed lines are hydrogen bonds.
d MyD88 phospho-serine (pS)-136 and pS-137 were built into the SPOP^{MATH}-MyD88 peptide crystal structure. MyD88 is in cyan. SPOP^{MATH} is in grey. Dashed lines are hydrogen bonds.

As expected, the double phosphorylation of S-136 and S-137 showed no measurable binding to SPOP^{MATH} at the peptide concentrations tested (**Fig. 3.11 a**). The single phospho mutants of these residues also showed a marked decrease in SPOP^{MATH} binding compared to the WT peptide, with phospho S-136 having the most effect (**Fig. 3.11 a, b**). Phosphorylation of MyD88, Puc and MacroH2A SBC serine or threonine residues has been shown to inhibit SPOP binding, and the phosphorylation of these residues was proposed to be mechanism to regulate substrate ubiquitylation by SPOP [55, 62].

To rationalise the FP data with the SPOP^{MATH}-MyD88 peptide crystal structure, the tested post-translational modifications were built into the crystal structure in an attempt to structurally interpret the FP data. The addition of phosphoryl groups to MyD88 S-136 and S-137 likely introduces steric clashes with SPOP residues K-129 and D-130. The position of the main chain of SPOP substrate peptides near their SBCs is consistent across the many structures solved (**Fig. 3.8 d**). The main chain positioning of the substrate SBC is needed to bring the small, negatively charged residues at positions three and four of the SBC close to SPOP D-130 and K-129 for hydrogen bonding. Any changes to this highly structurally conserved region would likely prevent binding, as shown by the reduced binding on the MyD88 single phospho-S peptides

and the complete abrogation of binding with the MyD88 double phospho-S peptide (**Fig. 3.11 a**, **b**).

Acetylation of MyD88 K-127 showed a modest decrease in IC_{50} from 5.26 μM to 1.20 μM, suggesting acetylation may increase MyD88 binding to $SPOP^{MATH}$ (**Fig. 3.11 a, b**). Building of the acetyl-K-127 into the crystal structure shows the acetyl group would form two hydrogen bonds with SPOP D-82, increasing the hydrogen bonding from one hydrogen bond in the WT crystal structure to two (**Fig. 3.8 c iii, Fig. 3.11 c**). Yet, in the MyD88 point mutant peptide competition assay, mutation from K127A did not show much of a change in IC_{50}, suggesting that this residue isn't critical for SPOP binding (**Fig. 3.9 a**). In addition, a SPOP D82K mutation did not show any marked effect on MyD88 binding (**Fig. 3.10 b**). This suggests that while MyD88 K-127 is involved in SPOP binding, as shown by the crystal structure, it is not critical for binding (**Fig. 3.8 c iii**).

Interestingly, phospho-T-141 also showed a modest decrease in IC_{50}, suggesting an increase in $SPOP^{MATH}$ binding (**Fig. 3.11 a, b**). This residue is not present in the $SPOP^{MATH}$-$MyD88^{peptide}$ crystal structure for the structural interpretation of T-141 phosphorylation. Together this may suggest that acetylation of MyD88 K-127 and phosphorylation of T-141 may slightly increase binding to SPOP; although, further experimentation would be needed to prove this as the data are from one independent experiment.

3.5 – Search for other SPOP substrates with an extended SBC motif

The identification of an extended SPOP binding site within the ID of MyD88 prompted a search as to whether other known SPOP substrates also possess an extended SPOP binding site. To explore this, first, an extended SPOP binding consensus motif was generated from the sequences of MyD88 and Pdx1, the two known extended binders (**Fig.**

3.12 a). This resulted in an extended SBC motif where a large amino acid could bind at the position of MyD88 L-129 and a glutamine could bind at the position of MyD88 Q-130 (**Fig. 3.12 a**). Additionally, MyD88 P-128 and A-133 are conserved in both proteins (**Fig. 3.12 a**). Lastly, positions 4 and 5 of the SBC could be relaxed to include any amino acid (**Fig. 3.12 a**). This extended SBC motif was then expanded by analysing the respective crystal structures for residues which could potentially bind in the same manner, with a particular focus on residues at the MyD88 ^{129}LQ^{130} position (**Fig. 3.9 c, Fig. 3.12 b**). This expanded consensus motif was then compared to the sequences of known SPOP substrates and six potential extended binder candidates were selected for testing (**Fig. 3.12 c, d**).

a

```
          Ext.    SBC
Ext. SBC  XXXPλQXXAφπSXXXXX
HsMyD88   AEKPLQVAAVDSSVPRT
HsPdx1    VAEPEQDCAVTSGEELL
MaPdx1    GDEPEQDSAVTSGEELL
          ..:*.*..*.*.*...
```

b

XXXX[LVIEDQN][QNE]XXXφπSXXXXX

c

Potential SPOP extended binders	
MyD88	AEKPLQVAAVDSSVPRT
PDX1	VAEPEQDCAVTSGEELL
GLI2	NNMPVQWNEVSSGTVDA
DEK	SVKSANVKKADSSTTKK
SRC3	NDDDVQKADVSSTGQGV
SENP7	DTDNLQSEQLSSSSDGS
SETD2	DKGSVQAPEISSNSIKD
CAPRIN1	QPEATQVPLVSSTSEGY

d

Protein	Sequence	Amino acid range	UniProt ID	Ref.
MyD88_1	GPGAGSAAPVSSTSSLP	5-21	Q99836	[62]
MyD88_2	AEKPLQVAAVDSSVPRT	125-141	"	[62]
*Hs*Pdx1	VAEPEQDCAVTSGEELL	220-236	P52945	[63]
*Ma*Pdx1	GDEPEQDSAVTSGEELL	220-236	P70118	[99]
Puc	SRENLACDEVTSTTSSS	89-105	O46122	[55]

CI_1	PFKDVVPEQPSSTSGGV	362-377	P19538	[64]
CI_2	PVNNTLFPDVSSSTHPY	1353-1369	“	[64]
GEMININ	EIGTCAEGTVSSSTDAK	190-206	O75496	[65]
TP53BP1	NVSSPATPTASSSSSTT	1632-1648	Q12888	[66]
BRD3	VKKKGVKRKADTTTPTT	237-252	Q15059	[67]
MacroH2A1	QGEVSKAASADSTTEGT	162-178	O75367	[55]
DAXX_2	TPSSVVETDISSSRKQS	578-593	Q9UER7	[61]
DAXX_3	TVLENGAGMVSSTSFNG	599-615	“	[55]
DAXX_4	SPLASLAPVADSSTRVD	671-687	“	[55]
DAXX_5	RVDSPSHGLVTSSLCIP	685-701	“	[61]
GLI2_1	NNMPVQWNEVSSGTVDA	1140-1156	P10070	[68]
GLI2_2(?)	QNKQSSESAVSSTVNPV	353-369	“	[68]
GLI2_3(?)	EPGGPESTEASSTSQAV	621-637	“	[68]
GLI2_4(?)	PLPSPGVNQVSSTVDSQ	1487-1503	“	[68]
GLI3_1	ELLSPGANQVTSTVDSL	1478-1494	P10071	[69]
GLI3_2	RTDVSEKAVASSTTSNE	26-42	“	[69]
GLI3_3	QGLSKVSEEPSTSSDER	67-82	“	[69]
DEK	SVKSANVKKADSSTTKK	277-292	P35659	[70]
BRD4	KTKKGVKRKADTTTPTT	283-299	O60885	[71]

NCOA3 (SRC-3/AIB1)	NDDDVQKADVSSTG QGV	91-107	Q9Y6Q9	[59]
DDIT3	TEEEPEPAEVTSTSQ SP	64-80	P35638	[72]
ERG	VAPHPPALPVTSSSF FA	431-447	P11308	[73]
SENP7	DTDNLQSEQLSSSSD GS	192-208	Q9BQF6	[74]
SETD2	DKGSVQAPEISSNSIK D	1363-1379	Q9BYW2	[75]
CDC20	AGRTPGRTPGKSSS KVQ	52-67	Q12834	[76]
SIRT2(?)	RVICLVGAGISTSAGI P	78-94 (PREDICTED)	Q8IXJ6	[77]
EGLN2_1	LSQALPQLPGSSSEP LE	10-26	Q96KS0	[78]
EGLN2_2	AGSGTPRATATSTTA SP	59-65	“	[78]
C-MYC_1	SPNPARGHSVCSTS SLY	176-192	P01106	[79]
C-MYC_2	PLVLHEETPPTTSSD SE	252-268	“	[79]
INF2	SLLGVLQAEADSTSE GL	1136-1152	Q27J81	[80]
HDAC6	VMKVEDREGPSSSK LVT	836-852	Q9UBN7	[81]
AR	NLKLQEEGEASSTTS PT	637-653	P10275	[82]
ERα	TDQSHLATAGSTSSH SL	563-579	P03372	[83]
HIPK2_1	STGHIVVTSASSTSVT G	89-105	Q9H2X6	[84]
HIPK2_2	SVTCGWGDVASSTT RER	854-870	“	[84]
BRD2	AKKKGVKRKADTTTP TP	274-290	P25440	[71]
NANOG	SMDLLIQDSPDSSTS PK	57-73	Q9H9S0	[85]

ATF2_1	VIIQQAVPSPTSSTVIT	183-199	P15336	[86]
ATF2_2	SRPQSLQQPATSTTETP	310-326	"	[86]
FASN_1	RGPSIALDTACSSSLMA	150-167	A0A0U1RQF0	[87]
FASN_2	AYLQARFPQLDSTSFAN	1704-1720	"	[87]
CAPRIN1	QPEATQVPLVSSTSEGY	426-442	Q14444	[88]
CDCA5	TSTPVPNPEAESSSKEG	113-129	Q96FF9	[89]
PDPK1	LLSQFGCMQVSSSSSSH	379-395	O15530	[90]
DHX9	WSPPQSNWNPWTSSNID	331-348	Q08211	[91]
BRMS1	WWDDKLHARGSSRSWDS	180-196	Q9HCU9	[92]
FLI-1	VPPHPSSMPVTSSSFFG	401-417	Q01543	[93]

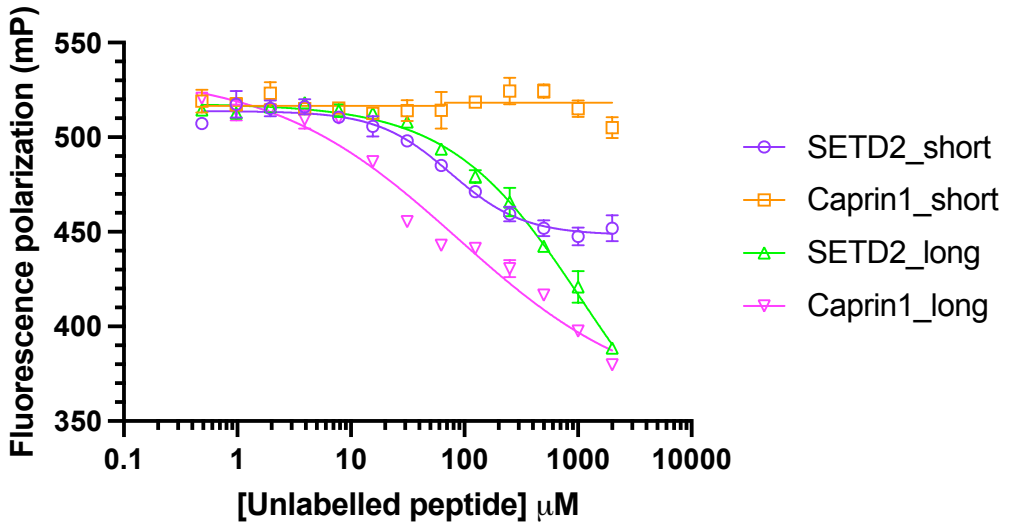
Figure 3.12 – Search for SPOP substrates with an extended SPOP binding consensus (SBC) motif
a Sequence alignment of MyD88 and Pdx1 to generate an extended SBC motif. Residues in blue are the SBC. Residues in yellow are the extended binding residues (Ext.). λ represents a large amino acid, X represents any amino acid, ϕ represents a non-polar amino acid and π represents a polar amino acid. ***** represents identical sequence. **:** represents conserved sequence. **.** represents semi-conserved sequence.
b Potential residues which could correspond to the extended SBC motif, as in **a**. Residues were also selected based the $SPOP^{MATH}$-MyD88 peptide crystal structure.
c List of SPOP substrates which may have an extended SBC motif with their sequences shown. Residues in blue are the SBC. Residues in yellow are the extended binding residues.
d Table of known SPOP substrates. Residues in blue are the SBC. Residues in yellow are the extended binding residues. Ref. is references.

3.5.1 – Testing potential SPOP substrates containing an extended SPOP binding consensus motif

To test the six SPOP substrates with suspected extended binding to SPOP, unlabelled peptides containing the extended SBC motif were

tested for these six candidate proteins and their binding to SPOP^{MATH} measured by competition FP assay (**Fig. 3.12 c, Fig. 3.13**).

a



b

Peptide	Peptide sequence	IC_{50} (μM)
SETD2_short	E**ISSNS**IKD	75.23
Caprin1_short	L**VSSTS**EGY	nb
SETD2_long	DKGSVQAPE**ISSNS**IKD	~1079.00
Caprin1_long	QPEATQVPL**VSSTS**EGY	~86.70

Figure 3.13 – Testing Caprin1 and SETD2 for extended binding to SPOP^{MATH}

a Peptide competition fluorescence polarization assay probing competitive binding of various unlabelled SPOP substrate peptides against a labelled MyD88 peptide (^{5Flu}-A**VDSSV**P) bound to SPOP^{MATH}. Lines show the non-linear fit using Prism-9's [inhibitor] vs response - variable slope analysis. Error bars are standard error of the mean.

b Table showing the half maximal inhibitory concentration (IC_{50}) values obtained in **a**. Bold residues are the SBC. nb is no measurable binding at the concentrations tested. Values are from 1 independent experiment.

Interestingly, the Caprin1_long peptide displayed a much higher binding to SPOP^{MATH} compared to the Caprin1_short peptide, suggesting that residues other than the SBC are involved in binding (**Fig. 3.13 a, b**). Despite reporting largely different IC_{50} values, the difference in binding between the long and short SETD2 peptides is less clear (**Fig. 3.13 a,**

b). The SETD2_short peptide reached saturation at the peptide concentrations tested, whereas the SETD2_long peptide had not, resulting in a less reliable IC_{50} value being reported for the long peptide (**Fig. 3.13 a, b**). While further repeats are needed, this preliminary screen has identified Caprin1 as a potential SPOP extended binder.

3.6 - Chapter summary

In this chapter it was identified that MyD88 binds to SPOP using an extended SBC motif. A combination of binding assays and an X-ray crystal structure allowed the identification of key MyD88 residues involved in this extended binding and the discovery of a high-affinity degron sequence. Interestingly, these newly identified residues are also implicated in DLBCL and potentially offers an explanation on how these previously uncharacterised MyD88 mutants cause disease. The post-translational modifications of MyD88 within the extended SPOP binding region was also explored and it is speculated that these modifications may help to regulate SPOP-mediated ubiquitylation and degradation of MyD88. Lastly, it was explored whether other SPOP substrates also contain an extended SBC motif. Overall, the SPOP-MyD88 binding interaction has been characterised, which may also allow for designing targeted therapeutics to modulate this interaction.

Chapter 4 – Cryo-EM structure determination of the SPOP oligomer

4.1 - Chapter introduction

The ubiquitin E3 ligase adaptor, SPOP, can homo-oligomerise to form large assemblies, which allows for efficient binding and ubiquitylation of substrates. Previous structural efforts had revealed the structure of the individual protein domains of SPOP, with no structural information of what the organisation of the FL SPOP oligomer resembles. Here an attempt is made to obtain the structure of the high-order SPOP oligomer using cryo-EM.

4.2 - Expression and purification of the SPOP oligomer

Previous structural determination of SPOP was limited to monomeric and dimeric forms of the various SPOP domains [55, 100]. However, efficient SPOP-mediated ubiquitylation of substrates occurs when SPOP oligomerises, for which structural information is lacking [203]. Isolation of FL SPOP from recombinant sources has proved challenging in the past due to its low solubility and nuclear localisation. Recombinant oligomeric SPOP had been produced in previous studies using amino acid residues 28-359 ($SPOP^{28-359}$) [95]. To produce oligomeric SPOP for structural studies, a $SPOP^{28-359}$ construct was designed for recombinant *E. coli* expression by the Dr Stephanie Wright lab, University of Leeds. This truncation includes the substrate-binding MATH domain, the BTB dimerisation domain (which forms BTB-BTB dimers) and the CTD dimerisation domain (which forms CTD-CTD dimers) (**Fig. 4.1 a**). In addition to truncating the protein, an N-terminal His^6-tag was added for affinity purification, followed by an MBP-tag to improve solubility and a HRV 3C protease site for tag removal.

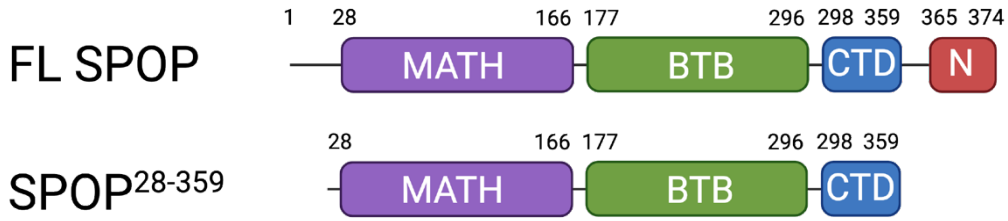


Figure 4.1 – Schematic of SPOP protein domains
Domain organisation of full-length (FL) SPOP and the truncation SPOP$^{28\text{-}359}$ used in experiments. Amino acid numbers are indicated. MATH is Meprin and TRAF homology domain. BTB is Broad complex, Tramtrack, and Bric-à-brac domain. CTD is C-terminal domain. N is nuclear localisation sequence.

This truncation of SPOP removes the very N-terminus and the C-terminal nuclear localisation sequence, which was not predicted to have any structural or functional significance aside from determining the subcellular localisation of SPOP (**Fig. 4.1 a**).

4.2.1 - Purification of SPOP$^{28\text{-}359}$

The His^{6}-MBP-3C-SPOP$^{28\text{-}359}$ fusion construct was expressed in *E. coli* BL21 DE3 cells and purified by nickel-affinity chromatography. The His^{6}-MBP-tag was cleaved using 3C protease and the cleaved tag was removed by reverse Nickel-chromatography (**Fig. 4.2 a**). SPOP$^{28\text{-}359}$ was further purified by size-exclusion chromatography to obtain protein used for experiments. The size-exclusion chromatogram showed that the majority of SPOP$^{28\text{-}359}$ eluted at ~40 mL with a peak “shoulder” at later elution volumes, which likely corresponds to SPOP oligomers of decreasing size (**Fig. 4.2 b**). Subsequent SDS-PAGE analysis of peak fractions showed a band at ~37 kDa, which is the predicted molecular weight of SPOP$^{28\text{-}359}$ (**Fig. 4.2 c**).

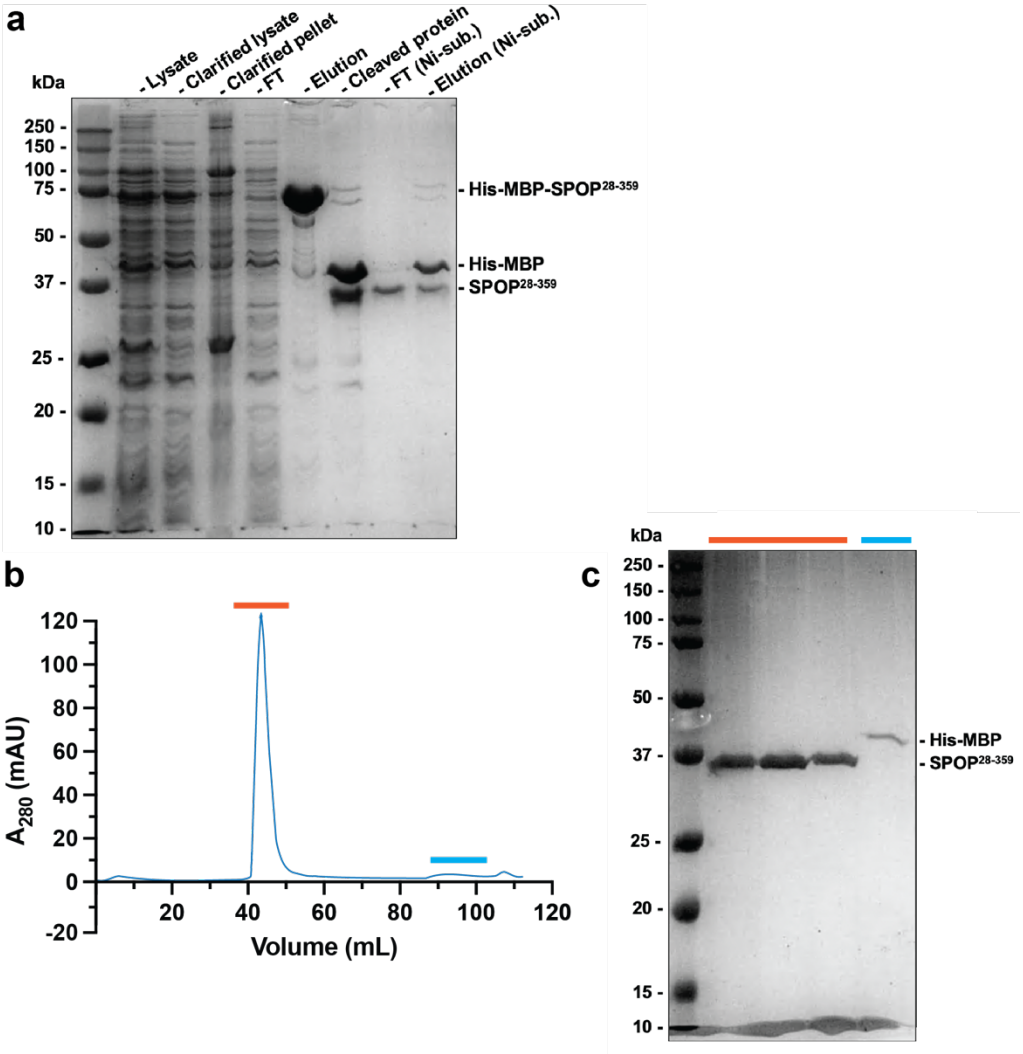


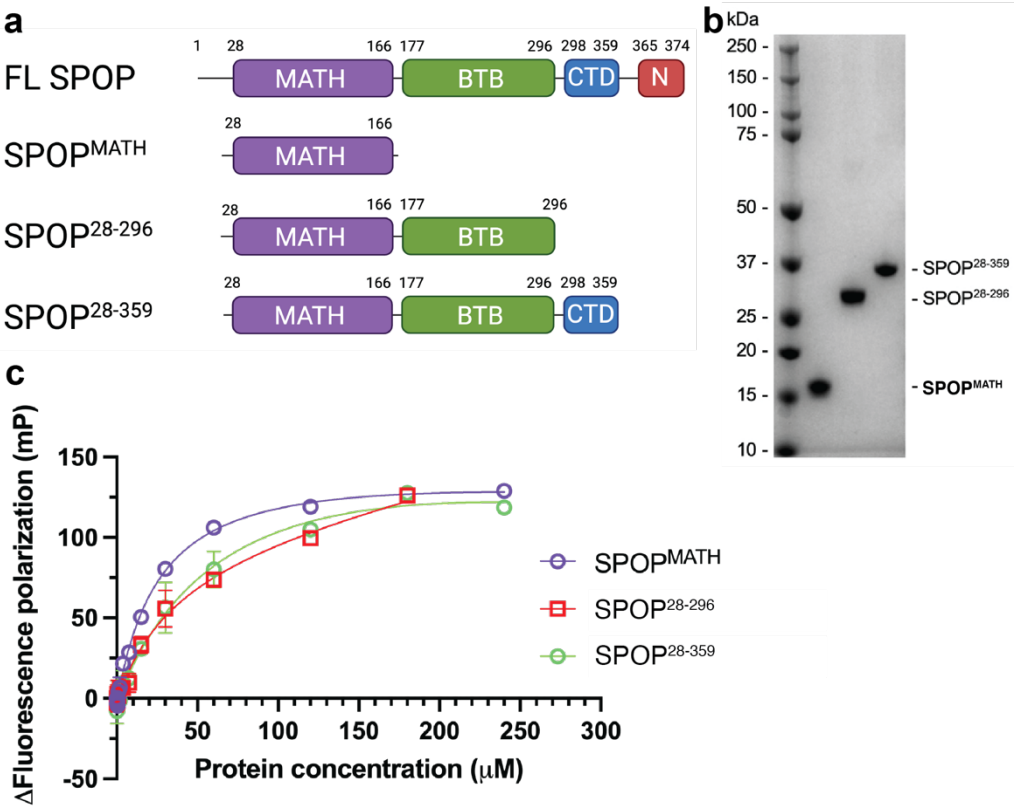
Figure 4.2 – Purification of SPOP$^{28-359}$
a SDS-PAGE gel of a Nickel-affinity purification of SPOP$^{28-359}$ from BL21 (DE3) cells. FT is flow-through. Ni-sub is nickel subtraction after cleaving the His-tag from SPOP$^{28-359}$.
b Size-exclusion chromatography chromatogram of SPOP$^{28-359}$. Column is HiLoad Superdex 200pg 16/600.
c SDS-PAGE gel of fractions as in **b**.

4.3 - Characterising the SPOP$^{28-359}$ oligomer

4.3.1 - SPOP$^{28-359}$ is able to bind to substrate

One concern with the protein preparation is that the majority of the protein eluted at an early volume of the size-exclusion chromatography column. While this was expected due to the large molecular weight and shape of the SPOP oligomer, it was also plausible that the protein was aggregated. To test whether SPOP$^{28-359}$ can bind to substrates, and

hence likely represents a folded protein, an FP assay was used to measure binding between various SPOP truncations, including SPOP$^{28-359}$, and a fluorescently labelled MyD88 peptide (^{5Flu}-^{133}A**VDSSV**P^{139}). The SPOP truncations tested also included SPOP^{MATH} which contains just the substrate binding MATH domain (monomeric), as well as SPOP$^{28-296}$, which also includes the BTB dimerisation domain (dimeric) (**Fig. 4.3 a, b**).



d

SPOP truncation	K_d (μM)
SPOP^{MATH}	32.25 ± 2.08
SPOP$^{28-296}$	22.7
SPOP$^{28-359}$	37.5

Figure 4.3 – Fluorescence polarization assay measuring binding between SPOP truncations and a MyD88 peptide
a Domain organisation of full length (FL) SPOP protein and various SPOP truncation constructs. Numbers indicate residue numbers. MATH is Meprin and TRAF homology domain. BTB is Broad-complex, Tramtrack, and Bric-à-brac domain. CTD is C-terminal domain. N is nuclear localisation sequence.
b SDS-PAGE gel of purified SPOP truncations. First lane is molecular weight standard with molecular weights (kDa) shown to the left.

c Fluorescence polarization assay measuring binding between a fluorescently labelled MyD88 peptide (^{5Flu}-AVDSSVP) and various SPOP truncation proteins. ΔFluorescence polarization is normalised fluorescence polarization in millipolarization units (mP).
d Dissociation constant (K_d) values as in **c**. Data are mean +/- SEM from 3 independent experiments for $SPOP^{MATH}$. $SPOP^{28-296}$ and $SPOP^{28-359}$ are from 1 independent experiment.

All three SPOP truncations bind to the MyD88 peptide with similar K_ds of 32.25 ± 2.08 μM for $SPOP^{MATH}$, 22.7 μM for $SPOP^{28-296}$ and 37.5 μM for $SPOP^{28-259}$ (**Fig. 4.3 c).** Since one MyD88 peptide is able to bind one MATH domain only, this assay does not distinguish for the multivalent substrate binding that is possible with SPOP oligomers binding to substrates with multiple SBCs. Altogether, oligomeric $SPOP^{28-359}$ is able to bind to substrate with a similar binding affinity as the minimal, monomeric $SPOP^{MATH}$ and the dimeric $SPOP^{28-296}$, suggesting that this protein is a folded oligomer rather than aggregated protein.

4.3.2 - Measuring the molecular weight of the $SPOP^{28-359}$ oligomer

A predicted structural model of the SPOP oligomer [100] based on published crystal structures shows that the oligomer is helical, with dimerisation interfaces formed across adjacent BTB-BTB domains and CTD-CTD domains (**Fig. 1.11 a**). This shape would give the oligomer a large apparent Stokes radius, which may cause the protein to elute from the size-exclusion chromatography column earlier than what is expected for its molecular weight. To explore the oligomeric state of $SPOP^{28-359}$ further, mass photometry was used to determine the molecular weight.

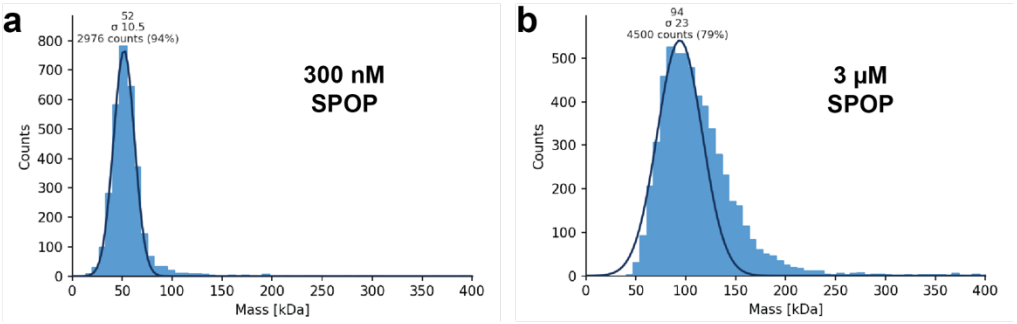


Figure 4.4 – Measuring the molecular weight of the SPOP$^{28-359}$ oligomer by mass photometry
a Molecular mass distribution histogram of 300 nM SPOP$^{28-359}$. Data are representative of a single experiment. Black line is the Gaussian fit of the peak. Median of the peak (kDa), standard deviation (σ; kDa) and number of counts shown above peak.
b Molecular mass distribution histogram of 3 μM SPOP$^{28-359}$. Data are representative of a single experiment. Black line is the Gaussian fit of the peak. Median of the peak (kDa), standard deviation (σ; kDa) and number of counts shown above peak.

Mass photometry measurements were made at 300 nM and 3 μM SPOP$^{28-359}$. Comparison of measurements at these two concentrations shows a clear shift in the mean molecular weight peaks from 52 kDa for 300 nM SPOP$^{28-359}$ to 94 kDa for 3 μM SPOP$^{28-359}$ (**Fig. 4.4 a, b**). In addition, the 3 μM SPOP$^{28-359}$ mass photometry measurements show a shoulder at higher molecular weights, indicating the presence oligomers of increasing size (**Fig. 4.4 b**). The minimum mass detection for this technique is ~50 kDa, therefore the mass measurement of 52 kDa for 300 nM SPOP likely represents an equilibrium between monomeric (37 kDa) and a dimeric (74 kDa) SPOP$^{28-359}$ (**Fig. 4.4 a**). The mass measurement of 94 kDa for 3 μM SPOP$^{28-359}$ likely represents an equilibrium between dimeric (74 kDa) and tetrameric (148 kDa) SPOP$^{28-259}$ (**Fig. 4.4 b**). The oligomerisation state of SPOP was shown to be highly concentration dependent and these mass measurements reflect this [203].

The high-order oligomerization of SPOP is mediated by both BTB-BTB and CTD-CTD interactions [55, 95, 100]. The SPOP BTB dimerization domains associate with low nanomolar affinity (≤1 nM), whereas the CTD

dimerisation domains have a relatively low dimerization affinity of 36 ± 3 μM [203]. Therefore, measuring the mass of longer SPOP oligomers would require increasing the protein concentration. Unfortunately, increasing the protein concentration further was not possible for mass photometry as higher concentrations saturated the detector due to molecular crowding effects. Therefore, the relative low concentrations needed for mass photometry experiments are too low to produce larger SPOP oligomers. However, it is clear from these preliminary experiments that SPOP$^{28-359}$ can self-associate in a concentration dependent manner, as has been reported previously, reaffirming that SPOP$^{28-359}$ forms ordered oligomers.

4.4 - Visualising SPOP$^{28-359}$ oligomers by electron microscopy

4.4.1 - Negative stain-EM of SPOP$^{28-359}$

With the aim of eventually obtaining the high-resolution cryo-EM structure of the SPOP oligomer, the SPOP$^{28-359}$ sample was first visualised by negative stain-EM. Negative stain grids were made with 0.75 mg/mL (20 μM), 1.5 mg/mL (40 μM) and 3 mg/mL (80 μM) SPOP$^{28-359}$. The grids showed that there was protein present on the grid; however, with the exception of the grids containing the highest protein concentration, very few oligomers were present (**Fig. 4.5**).

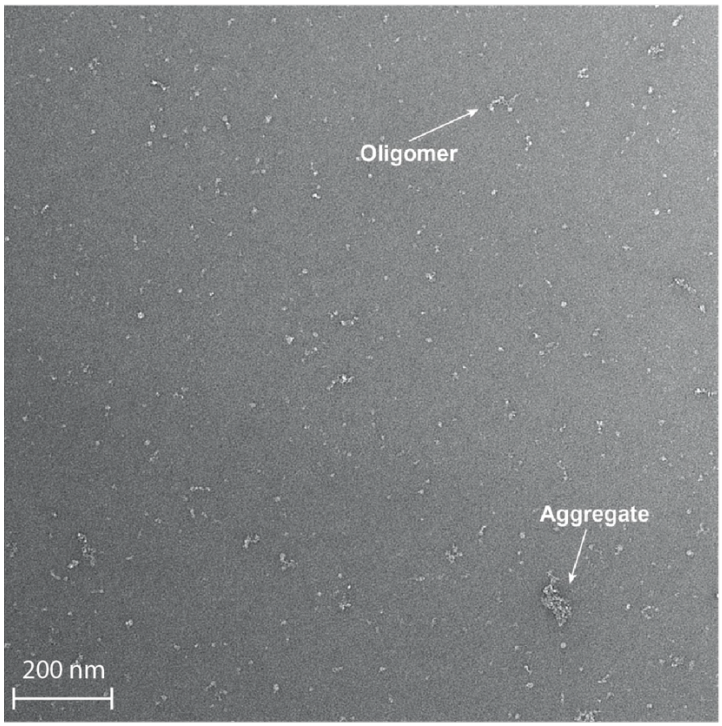


Figure 4.5 – Initial SPOP$^{28-359}$ negative stain-EM screening
Negative stain-EM micrograph of 80 μM SPOP$^{28-359}$ stained with 1% uranyl acetate. Imaged at a magnification of 39,886x. An example of a SPOP$^{28-359}$ oligomer and a protein aggregate are indicated.

The stain used to prepare these negative stain grids was 1% uranyl acetate, which has a pH of ~4.5. It is possible that this acidic stain caused the SPOP oligomers to dissociate or denature. Importantly, none of the negative stain grids displayed any large SPOP oligomers which would be amenable to negative stain-EM data collection and image processing (**Fig. 4.5**).

4.4.2 - Cryo-EM of SPOP$^{28-359}$

To check whether the acidic stain used for negative stain-EM was causing the SPOP oligomers to dissociate, SPOP$^{28-359}$ was imaged using cryo-EM. Cryo-EM grids were made with either 0.75 or 3 mg/mL (20 or 80 μM) SPOP$^{28-359}$. At 3 mg/mL, the grids appear to be overcrowded with protein, with clear oligomers difficult to pick out (**Fig. 4.6 a**).

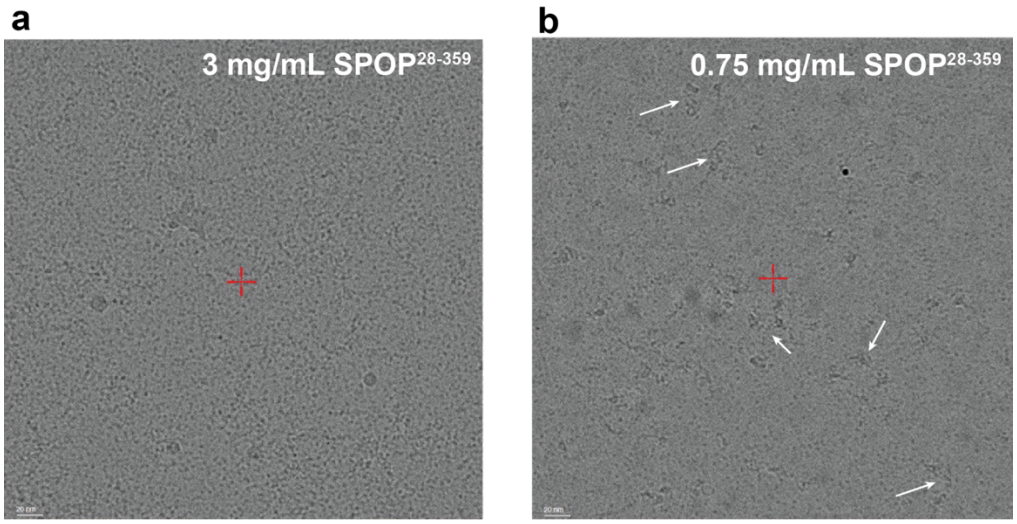


Figure 4.6 – SPOP$^{28-359}$ cryo-EM grid screening
a Micrograph of 3 mg/mL (80 μM) SPOP$^{28-359}$ cryo-EM grid. Imaged at a magnification of 96,000x.
b Micrograph of 0.75 mg/mL (20 μM) SPOP$^{28-359}$ cryo-EM grid. White arrows indicate potential SPOP$^{28-359}$ oligomers. Imaged at a magnification of 96,000x.

At 0.75 mg/mL, the protein seemed less crowded, where some potential SPOP oligomers could be seen (**Fig. 4.6 b**). These oligomers appear to be ~20-60 nm (200-600 Å) in length. However, the ice quality was poor and the number of clear SPOP oligomers was too low to justify collecting a cryo-EM dataset. From these images it was estimated that a protein concentration of approximately 0.75-2 mg/mL would be ideal for visualising SPOP oligomers by cryo-EM. This would be a high enough concentration to produce SPOP oligomers but would not be too concentrated so that the grid would be overcrowded with protein. However, it was clear from the initial EM screening that the SPOP$^{28-359}$ oligomers would need to be further stabilised.

4.5 - SPOP$^{28-359}$ negative stain-EM optimisation

With the intention of producing large, stable SPOP oligomers for cryo-EM analysis, the SPOP$^{28-359}$ sample was crosslinked using glutaraldehyde (GA). This would allow SPOP$^{28-359}$ oligomers to remain

associated, even after dilution of the protein sample, making it more amenable to cryo-EM structure determination.

4.5.1 - Optimisation of glutaraldehyde crosslinking

SPOP$^{28-359}$ was incubated with various concentrations of GA, on ice, for 30 minutes before analysing the reactions by SDS-PAGE under reducing and denaturing conditions (**Fig. 4.7 a**).

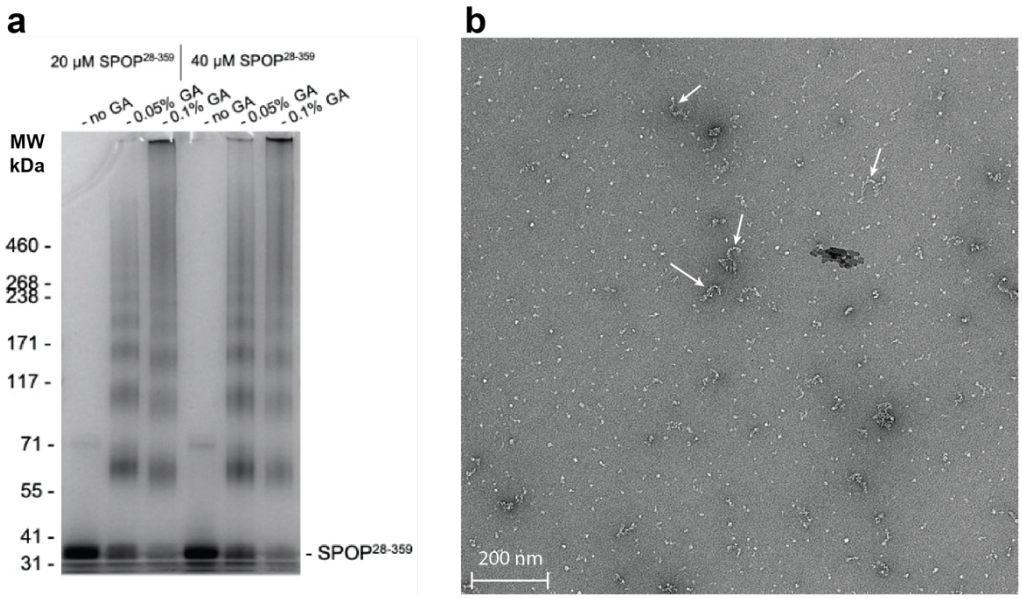


Figure 4.7 – SPOP$^{28-359}$ glutaraldehyde crosslinking optimisation
a 3-8% Tris-Acetate SDS-PAGE gel under reducing conditions of 20 µM and 40 µM SPOP$^{28-359}$ treated with either 0%, 0.05% or 0.1% glutaraldehyde.
b Representative image of 1% uranyl formate-stained, 0.05% glutaraldehyde crosslinked 2.5 µM SPOP$^{28-359}$. Imaged at a magnification of 39,886x. White arrows indicate long SPOP$^{28-359}$ oligomers.

The SDS-PAGE analysis shows that without crosslinker, the majority of the 20 µM or 40 µM SPOP$^{28-359}$ is monomeric (37 kDa) (**Fig. 4.7 a**). With the addition of 0.05% GA, several bands are observed, corresponding to high molecular weight species (**Fig. 4.7 a)**. These bands are separated by approximately 37 kDa, and therefore likely represent different oligomeric states of the SPOP$^{28-359}$. In addition, the band corresponding to the monomer is less intense in the 0.05% GA samples compared to the no GA samples, showing that SPOP is incorporated into larger SPOP

oligomers (**Fig. 4.7 a)**. This is increased further in the 0.1% GA samples, where higher molecular weight bands can be seen up to the well of the gel, likely reflecting larger SPOP oligomers or even non-specific aggregation (**Fig. 4.7 a)**.

4.5.2 - Negative stain-EM of glutaraldehyde-crosslinked SPOP$^{28-359}$

With the aim of observing large SPOP oligomers, the 0.05% GA-crosslinked 40 μM SPOP$^{28-359}$ sample was used directly after crosslinking for analysis by negative stain-EM. This time, 1% uranyl formate was used as a stain. Compared to uranyl acetate, uranyl formate has a smaller grain size which allows for finer staining. This is particularly advantageous for smaller proteins or proteins which have small dimensions on one of their axes, such as the long but thin shape of SPOP$^{28-359}$ oligomers. Negative stain grids made with protein concentrations from 0.094–1.5 mg/mL (2.5-40 μM) were screened.

The GA crosslinking seemed to improve the stability of the oligomers for negative stain-EM. Even at the lowest protein concentration (2.5 μM) imaged, convincing oligomers were seen (**Fig. 4.7 b**). There are still a variety of lengths of oligomers, with some curved shapes seen with the longer oligomers, which resemble the predicted SPOP oligomer structural model (**Fig. 4.4 a, Fig. 4.7 b**). Overall, GA crosslinking likely provides a more suitable sample for EM studies, where crosslinked SPOP oligomers remain, even when diluting the sample to lower protein concentrations. However, oligomeric species were still low and further optimisation to generate stable SPOP oligomers was required.

4.6 - SPOP$^{18-359}$ produces more stable oligomers

Concurrent with the above GA-crosslinked SPOP$^{28-359}$ negative stain-EM optimisation, it was noted that the AlphaFold2 [96] predicted model of FL

SPOP shows the N-terminus (residues 21-38) forming a β-sheet with a region immediately C-terminal to the MATH domain (residues 154-172) (**Fig. 4.8 a**). Additionally, generating a predicted structure of a SPOP dimer using the ColabFold multimer prediction server [204] showed that this additional β-sheet forms a larger β-sheet with the adjacent SPOP protomer, providing an additional interface that would likely stabilise the SPOP oligomer (**Fig. 4.8 b**).

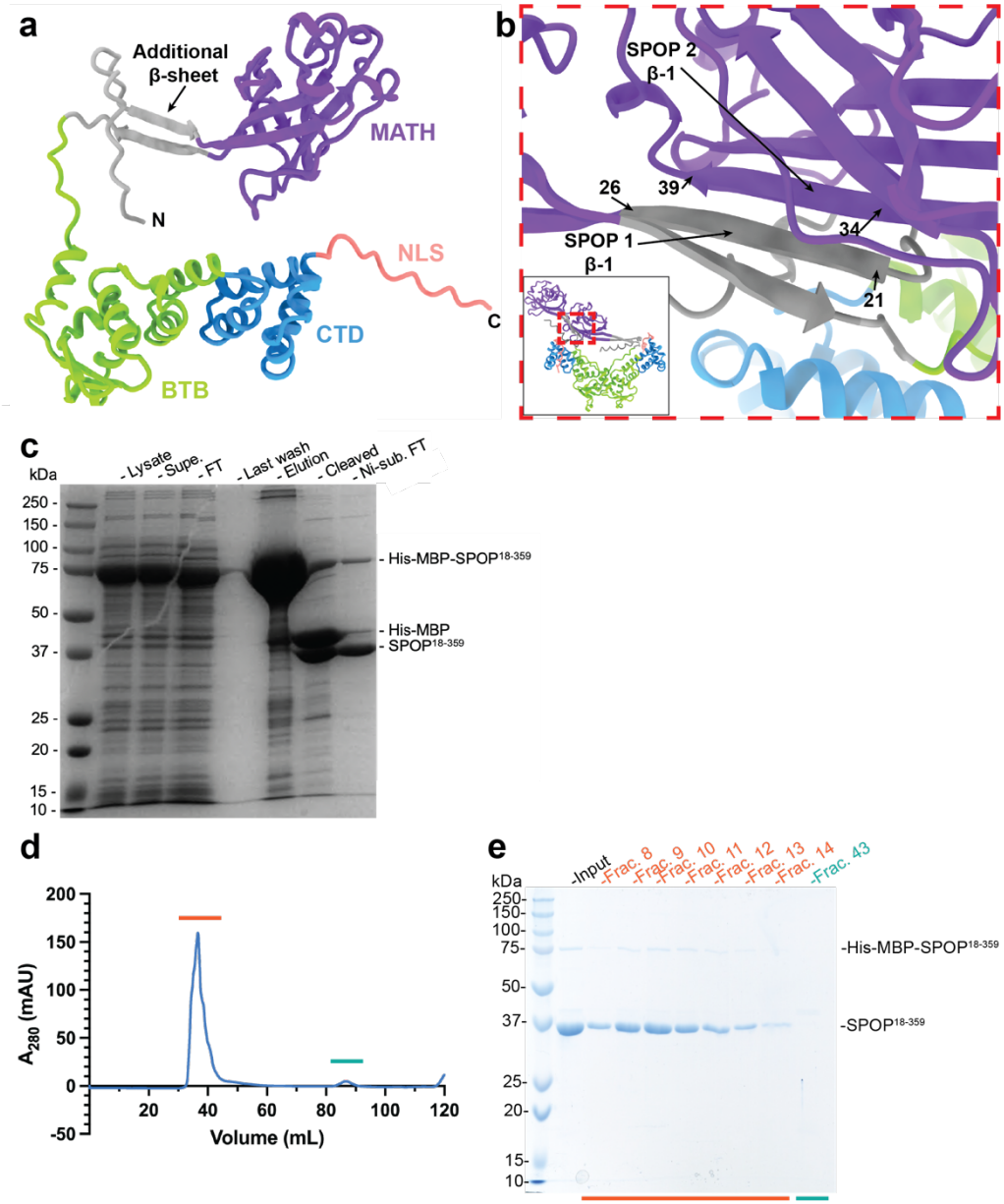


Figure 4.8 – AlphaFold predicted model of SPOP and SPOP$^{18-359}$ purification
a AlphaFold [96] predicted structure of full length SPOP. MATH is Meprin and TRAF homology domain. BTB is Broad-complex, Tramtrack, and Bric-à-brac domain. CTD is C-terminal domain. NLS is nuclear localisation sequence. N-

terminus (N) and C-terminus (C) of SPOP indicated. The N-terminal residues 21-38 form a β-sheet with a region immediately C-terminal to the MATH domain (residues 154-172).
b ColabFold [204] multimer predicted structure of a FL SPOP dimer, showing parallel β-sheet formed between β-1 of two adjacent SPOP MATH domains. Residue numbers are indicated. Zoomed out view is shown in the bottom left.
c SDS-PAGE gel of a Nickel-affinity purification of $SPOP^{18-359}$ from BL21 (DE3) cells. Supe is supernatant. FT is flow-through. Cleaved is after incubation with TEV protease. Ni-sub is nickel subtraction after TEV cleavage.
d Size-exclusion chromatogram of $SPOP^{18-359}$. Column is HiLoad Superdex 200pg 16/600. Coloured lines represent peaks from which fractions were analysed by SDS-PAGE, as shown in **e**.
e SDS-PAGE gel of $SPOP^{18-359}$ fractions, as shown in **d**.

In the $SPOP^{28-359}$ construct used in previous experiments, this β-sheet is incomplete due to truncating SPOP at residue 28. Therefore, a longer SPOP construct spanning amino acids 18-359 (His-MBP-3C-$SPOP^{18-359}$) was expressed in *E. coli* and purified in a similar manner to $SPOP^{28-359}$ (**Fig. 4.8 c**). $SPOP^{18-359}$ also eluted from the SEC column at ~40 mL; however, the elution peak was much sharper compared to $SPOP^{28-359}$, with less of a shoulder at later elution volumes (**Fig. 4.2 b, Fig. 4.8 d**).

During the purification of His-MBP-3C-$SPOP^{18-359}$, it proved difficult to fully cleave the His-MBP tag from the protein, despite several efforts to optimise this process (**Fig. 4.8 c, e**). Likely, the more stable oligomers of $SPOP^{18-359}$ hindered access to the 3C protease cleavage site. This could theoretically be circumvented by diluting the protein in order to depolymerise the oligomers. However, $SPOP^{18-359}$ was difficult to concentrate using centrifugal concentrators, as it would stick to the membrane and clog it; this was not overcome by changing membrane material or by adding glycerol to the purification buffer. With the intention of eventually using this protein for cryo-EM studies, maintaining a high enough protein concentration after affinity and size-exclusion purification steps without being able to concentrate the protein proved challenging. The decision was made to not optimise removal of all of the His-MBP tag, as further processing of the sample would decrease overall yield. The protein preparation was scaled-up and purified as previous, without

any protein concentration steps. Regardless, the remaining His-MBP tag did not seem to affect the oligomerisation ability of $SPOP^{18-359}$. Overall, the sharper elution peak, the relative inaccessibility of the His-MBP tag to protease and the tendency for the protein to stick to centrifugal concentrator hinted at $SPOP^{18-359}$ forming longer or more stable oligomers compared to $SPOP^{28-359}$.

4.7 - Negative stain-EM of $SPOP^{18-359}$

$SPOP^{18-359}$ was imaged by negative stain-EM without any chemical crosslinking. The $SPOP^{18-359}$ negative stain micrographs showed long, curved filaments at protein concentrations ranging from 0.5 mg/mL (13 μM) to as low as 0.03 mg/mL (0.8 μM) (**Fig. 4.9 a**). Some single particles resembled circles, which likely represent a view of the oligomer similar to looking down the top of a spring (**Fig. 4.9 a**).

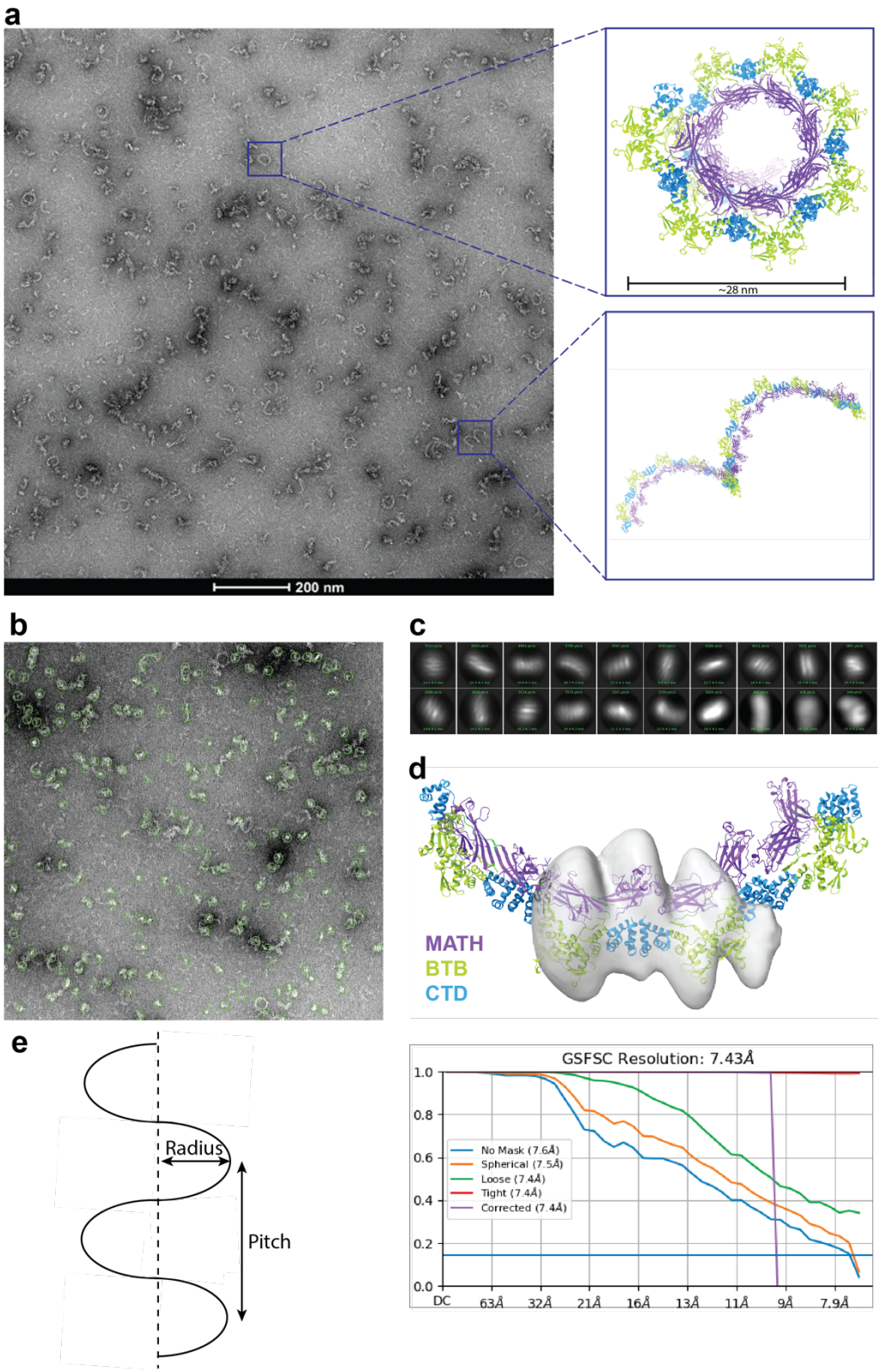


Figure 4.9 – Negative stain-EM of SPOP$^{18\text{-}359}$
a *Left* – Representative negative stain-EM micrograph of SPOP$^{18\text{-}359}$ oligomers at 0.125 mg/mL (3.3 μM). Imaged at magnification of 39,886x. *Right* – Selected single particle views relative to the predicted SPOP oligomer structure [100], coloured as follows: purple are MATH domains, green are BTB domains and blue are CTD domains.

b Representative micrograph from SPOP$^{18-359}$ negative stain-EM dataset showing picked particles (green circles).
c Reference-free 2D classes of the SPOP$^{18-359}$ negative stain-EM dataset.
d Negative stain-EM 3D refined map of SPOP$^{18-359}$ (transparent grey) fitted onto the predicted SPOP oligomer structure. FSC curve shown at the bottom right, calculated using the gold-standard FSC cut-off at 0.143.
e Schematic highlighting the radius and pitch of a helix.

A negative stain-EM dataset of 101 micrographs was collected and processed. Despite the SPOP oligomer predicted to be helical structure, a single particle data processing approach was taken, where boxes along part of an oligomer were picked and averaged together, without any helical symmetry imposed (**Fig. 4.9 b**). This approach was taken largely due to the helical oligomer having a large hollow helical radius rather than being a straight helical oligomer, which would complicate data processing when imposing helical symmetry (**Fig. 4.9 e**).

The picked single particles were used to generate 2D classes which show a long, thin filament with a slight curvature seen in some views (**Fig. 4.9 c**). Particles from selected 2D classes were used to generate a 3D density map. The 3D density map fitted over the predicted SPOP model shows a relatively good fit, although the resolution is low due to limitations of the negative stain (**Fig. 4.9 d**). Additionally, the reported resolution is likely an overestimation, with the resulting map not resembling a 7.43 Å map (**Fig. 4.9 d**). Overall, since the SPOP$^{18-359}$ sample was shown to form long oligomers by negative stain-EM and a small negative stain-EM dataset produced a reasonable 3D reconstruction, the sample was deemed suitable for cryo-EM (**Figure 4.9 a, d**).

4.8 - Cryo-EM dataset of SPOP$^{18-359}$

Next, SPOP$^{18-359}$ was visualised by cryo-EM. The cryo-EM micrographs show long, curved SPOP$^{18-359}$ oligomers, with good enough particle contrast to see the repeating SPOP units or domains of the oligomer

(**Fig. 4.10 a**). A dataset containing 2,916 cryo-EM movies was collected and processed using a single particle processing approach (full data collection and processing parameters in **Table 2.4**). “Single particles” along an oligomer were picked using a box size large enough to include several repeating SPOP units. Several rounds of 2D classification produced classes with enough signal to see individual protein domains. Two discrete layers along the filament can be seen in the classes. These layers correspond to the repeating BTB-BTB and CTD-CTD dimer layer and the MATH domain layer (**Fig. 4.10 b**).

The initial 3D model algorithm was unable to generate a convincing initial 3D map from the particles *de novo*. This is likely due to the use of a single particle processing approach, despite SPOP being a continuous helical filament rather than an actual single particle. Therefore, there are missing views of these picked “single particles”, which makes it difficult for the processing software to reconstruct a 3D map from the 2D images. To overcome this, the predicted structural model, low pass filtered to 60 Å, was used as an initial 3D model, along with the selected particles, for 3D model refinement (**Fig. 4.10 c**). The resulting 3D density map agrees well with the initial 3D model made from the predicted structural model and closely resembles the features observed in the reference-free 2D classes (**Fig. 4.10 b, c, d**).

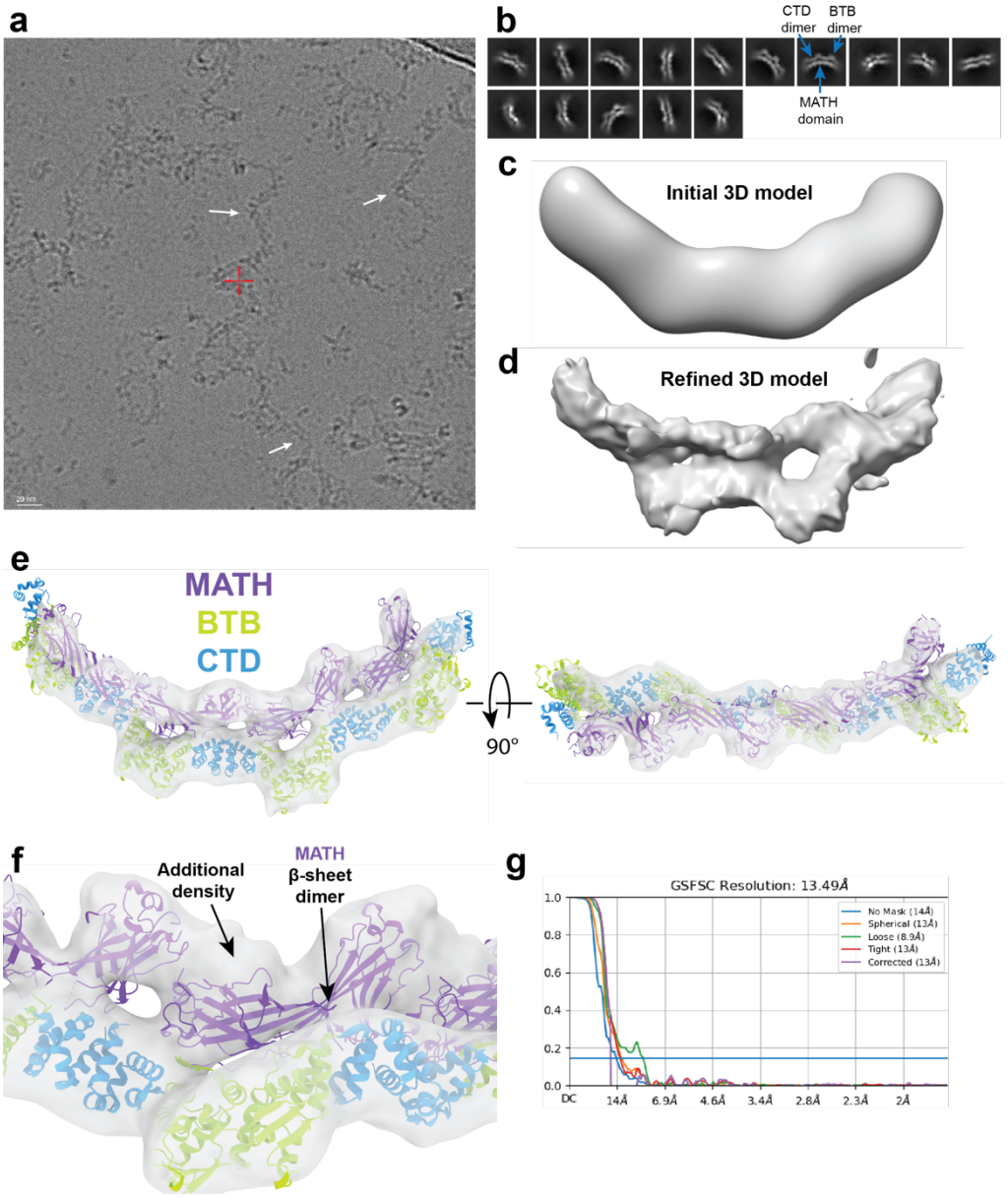


Figure 4.10 – Cryo-EM data processing of the SPOP$^{18-359}$ oligomer
a Cryo-EM micrograph of SPOP$^{18-359}$. White arrows indicate SPOP$^{18-359}$ oligomers. Imaged at a magnification of 96,000x.
b Selected reference-free 2D classes of picked single particles along SPOP$^{18-359}$ oligomers.
c 3D volume used as an initial density map for 3D model refinement. Volume was created from a predicted SPOP oligomer model [100] and low pass filtered to 60 Å.
d Cryo-EM 3D refined map of SPOP$^{18-359}$.
e 3D map in **d** fitted onto predicted SPOP oligomer model.
f View of **e** highlighting map to model fit near MATH domains.
g FSC curve of cryo-EM density map shown in **d**, calculated using the gold-standard FSC cut-off at 0.143.

A concern when refining an initial 3D density map not generated directly from the data is that the resulting refined 3D map will be biased. Although this is somewhat mitigated by low pass filtering the initial 3D map so that the detailed features of the model are erased, there still remains the risk of bias. Reassuringly, although the predicted structure fits well into the refined 3D density, there are some subtle differences which are clear, despite the limited resolution. The SPOP BTB-BTB and CTD-CTD domain dimers have been structurally characterised previously, and have dimerisation interfaces that are fairly rigid [55, 100]. This is reflected in the refined 3D map, showing that the BTB and CTD domains of the predicted oligomer model fit well into the density (**Fig. 4.10 e**). However, there are some differences in respect to the fit of the MATH domains.

The predicted SPOP oligomer structure was made by superimposing various crystal structures of SPOP domains (**Fig. 1.11 a**). To date, published crystal structures of SPOP which included the MATH domains began at residues 26-28, prematurely truncating the predicted MATH-MATH β-sheet dimer (**Fig. 4.8 a, b**). Only one structure has shown β-sheet interactions between adjacent MATH domains. A crystal structure of a SPOP MATH+BTB dimer (residues 26-329) showed β-sheet interactions between adjacent MATH domains from amino acids 26-31 [55]. While the $SPOP^{MATH}$ protein (residues 28-166) used in **Chapter 3** contains four residues of this β-strand, it is not sufficient to cause dimerisation, with the protein predominantly observed as monomeric. An extended $SPOP^{MATH}$ protein (residues 18-172) appears to form some dimers in a concentration dependent manner (Dr Stephanie Wright, personal communication).

The MATH domains used for the predicted SPOP oligomer structure used in this study begin at amino acid 26 [55]. When looking at the predicted structure fitted into the cryo-EM density map, it appears as though one MATH domain, across a MATH dimer, fits well into the density whereas the other MATH domain does not (**Fig. 4.10 f**). One

explanation for this is that completion of the β-sheet by extending the MATH domain boundaries to residue 18 rigidifies the MATH dimers, causing them to have a different positioning in respect to the BTB-CTD domains.

Regardless, this difference in MATH positioning in the SPOP$^{18-359}$ cryo-EM density map versus the predicted structural model is reassuring, as it suggests that the 3D density map is not completely biased by the starting model. Although, this interpretation would have to be confirmed with high-resolution structural data.

4.9 - Comparison of SPOP$^{18-359}$ cryo-EM data with the published cryo-EM structure of SPOP

During optimisation of the SPOP$^{18-359}$ sample for high-resolution cryo-EM structure determination, a study published the 3.6 Å cryo-EM structure of the FL SPOP oligomer (**Fig. 4.11 a**) [188]. Overall, this published structure resembles the SPOP$^{18-359}$ cryo-EM data generated in this thesis; however, there are some differences.

One notable difference is the orientation of the MATH domains. While the predicted model showed the MATH domains oriented in a head-to-head, tail-to-tail orientation, the FL SPOP structure shows the MATH domains oriented in a head-to-tail orientation (**Fig. 4.11 a, b**). This orientation of the MATH domains is surprising, as a previous crystal structure [55] from which the predicted SPOP oligomer structural model was made had shown two adjacent MATH domains to form antiparallel β-sheets across β-1 (residues 26-31) of one MATH domain and β-1 (residues 26-31) of another, with the MATH domains oriented in a head-to-head, tail-to-tail orientation. In addition, the ColabFold multimer prediction of the SPOP-SPOP dimer shows two adjacent MATH domains forming parallel β-sheets across β-1 (residues 34-38) of one MATH domain and β-1 (residues 26-31) of another, with the MATH domains

oriented in a head-to-tail orientation (**Fig. 4.8 b)**. The FL SPOP structure [188] shows the two adjacent MATH domains interacting between β-1 (residues 34-38) of one MATH domain and the unstructured N-terminus (residues 21-25) of another, with the MATH domains oriented in a head-to-tail orientation (**Fig. 4.8 b, Fig. 4.11 a**). Presumably due to the limited resolution of the map, there is little secondary structure present across this MATH-MATH interaction interface. Though, if this interface were to form β-sheets, then they would be parallel. Despite the differences in the MATH-MATH orientation and interacting residues between the studies, the FL SPOP model does seem to fit well with its map (**Fig. 4.11 c**).

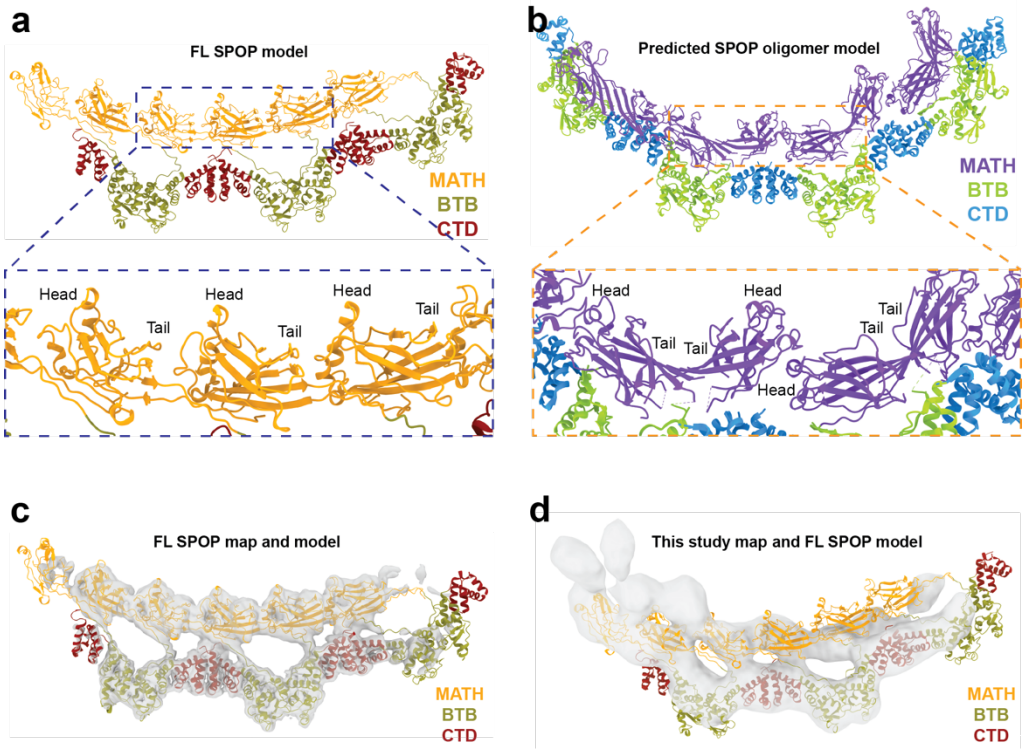


Figure 4.11 – Comparison of the SPOP predicted structural model with the published SPOP oligomer structure
a Cryo-EM structure of full length (FL) SPOP (PDB ID 8dwv) [188]. Dashed line box shows a zoomed in view of the MATH domains to show their orientation. α-helical end of the MATH domain labelled as head. β-sheet end of MATH domain labelled as tail.
b Predicted structural model of SPOP oligomer [100]. Dashed line box shows a zoomed in view of the MATH domains to show their orientation. α-helical end of the MATH domain labelled as head. β-sheet end of MATH domain labelled as tail.
c FL SPOP model fitted into the FL SPOP cryo-EM density map (transparent grey) [188].
d FL SPOP model [188] fitted into cryo-EM density map (transparent grey) generated in this study.

Fitting of the FL SPOP structure into the cryo-EM map generated in this study shows a poor fit between the map and the model (**Fig. 4.11 d**). Comparison of the two studies shows that the SPOP oligomer structures have different helical parameters. The FL SPOP structure reportedly has a helical pitch of 630 Å (twelve SPOP dimers per turn), whereas the predicted oligomer model had a helical pitch of 510 Å (eight SPOP dimers per turn) (**Fig. 4.9 e**) [100, 188]. This difference in helical pitch might suggest that the oligomer is quite flexible, perhaps allowing SPOP to engage various substrates of different shapes.

Superimposing the Cullin-3 ligase onto the predicted SPOP oligomer structure and the FL SPOP oligomer structure shows that both are able to accommodate Cullin-3 without any clashes, with FL SPOP able to accommodate more Cullin-3 per helical turn due to the larger helical pitch.

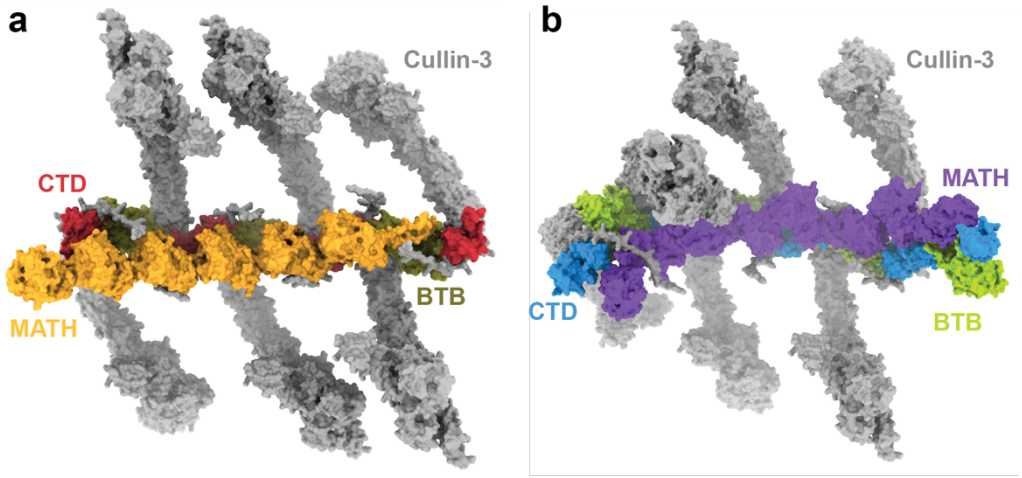


Figure 4.12 – Superimposing Cullin-3 ligase onto SPOP oligomer structures
a Superimposing the Cullin-3 AlphaFold [96] model onto the structure of full-length SPOP (PDB ID 8dwv) [188]. SPOP is coloured by its domains.
b Superimposing the Cullin-3 AlphaFold [96] model onto the predicted structure of SPOP. SPOP is coloured by its domains.

4.9.1 – Re-processing the SPOP$^{18-359}$ cryo-EM dataset

As a proof of concept, the SPOP$^{18-359}$ cryo-EM dataset was reprocessed using the FL SPOP model as an initial 3D reference model. As with the

predicted SPOP model, the FL SPOP model was converted to a 3D volume using ChimeraX and low pass filtered to 60 Å. This low pass filtered 3D volume was then used for 3D model refinement using the particles from this study. The resulting 3D map more closely matched the published FL SPOP map, with the cryo-EM density map showing the SPOP oligomer having a higher helical pitch compared to previously (**Fig. 4.13 a**). However, the cryo-EM map resolution was not improved, as would be expected from changing from an "incorrect" initial model to a "correct" initial model (**Fig. 4.13 a**). Fitting the predicted SPOP oligomer model into the re-processed 3D density also shows a poor fit (**Fig. 4.13 b**). Both the initial and re-processed cryo-EM density generated in this study are not high enough resolution to determine which one is correct; however, it is also possible that both densities are somewhat correct, and the SPOP oligomer undergoes significant flexible motion in solution that could account for the differences in helical pitch. Although residues 1-18 are present in the construct used to gain the structure of FL SPOP [188], cryo-EM density for these residues is lacking, suggesting that these are unlikely to contribute to the differences in oligomer structure.

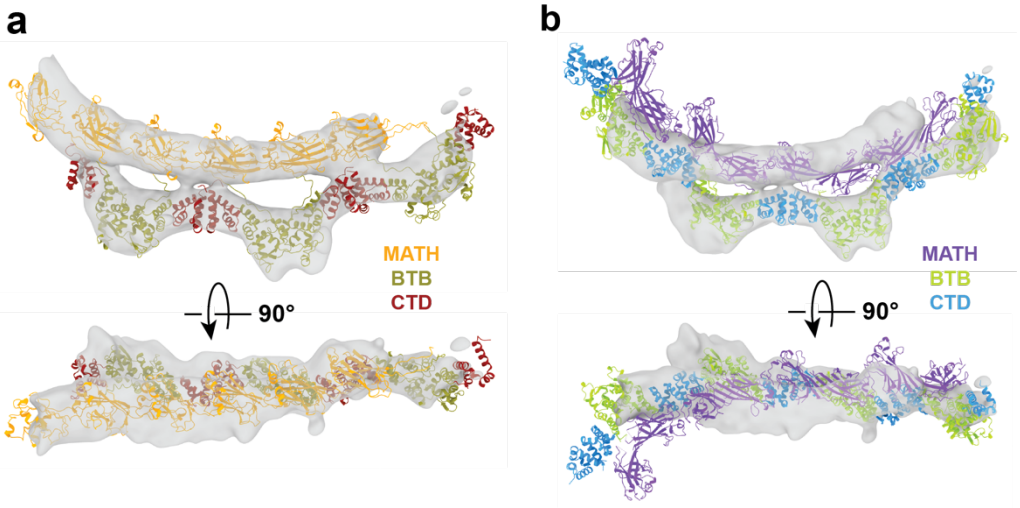


Figure 4.13 – Re-processing the cryo-EM dataset with the FL SPOP model as a reference model
a Re-processed cryo-EM map (transparent grey) with the FL SPOP model [188] rigid body fitted in.
b Re-processed cryo-EM map (transparent grey) with predicted SPOP oligomer model [100] rigid body fitted in.

4.10 - Using CryoSPARC 3DFlex to model motion of SPOP$^{18-359}$

To explore whether any flexible motion of the SPOP$^{18-359}$ oligomer could be modelled from the cryo-EM data, 3DFlex was used. 3DFlex can model the non-rigid motion or flexibility motion present in the cryo-EM data. It can then use this motion model to combine the signal from particles in different flexible conformations into a single consensus reconstruction to improve the resolution in these flexible regions.

The resulting 3DFlex model shows the SPOP oligomer undergoing large motions, particularly a twisting and relaxing motion (**Fig. 4.14**). The source of this flexibility appears to mostly occur at the CTD domains, where there is an expansion and contraction across the CTD dimerisation interface, resulting in the oligomer twisting and relaxing respectively. This also results in the MATH domains undergoing a positional shift where they twist and relax (**Fig. 4.14**). This twisting and relaxing motion would effectively change the helical pitch of the oligomer, reflecting the conformations seen in the FL SPOP model and the predicted SPOP model (**Fig. 4.9 e**, **Fig. 4.11 a, b**).

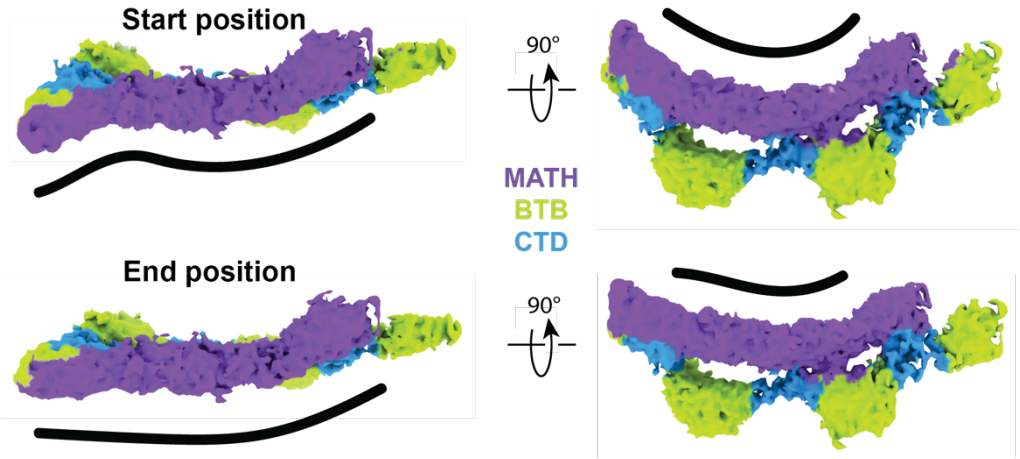


Figure 4.14 – SPOP$^{18-359}$ 3DFlex motion modelling
3DFlex volumes showing the SPOP$^{18-359}$ oligomer undergoing motion. Black line illustrates the twisting and relaxing motion of the oligomer, as shown by the start and end positioning of the MATH domains.

This flexible motion of the SPOP oligomer makes it a challenging sample for cryo-EM structure determination. Future attempts to structurally characterise the SPOP oligomer by cryo-EM would require a much larger dataset than that collected in this study, followed by extensive 3D classification steps to isolate SPOP oligomer classes in different conformations, or the use of flexible motion models such as 3DFlex to generate a consensus 3D map.

4.11 - Chapter summary

Here it is shown that the SPOP protein domain boundaries starting at amino acid 28 prematurely truncated the MATH domain, and extending the boundaries to begin at amino acid 18 produced more stable SPOP oligomers. $SPOP^{18-359}$ was imaged using cryo-EM and a low-resolution map was generated. The $SPOP^{18-359}$ cryo-EM dataset was also compared to the recently published FL SPOP oligomer structure and some key differences were highlighted. Since the structure of oligomeric FL SPOP was published, efforts were focussed on gaining the cryo-EM structure of the $SPOP^{18-359}$ oligomer bound to its substrate, MyD88.

Chapter 5 – Cryo-EM structure determination of the SPOP-MyD88 high-order oligomer complex

5.1 - Chapter introduction

In the previous chapter, the structure of the high-order SPOP oligomer was explored using cryo-EM. How this SPOP oligomer engages oligomeric MyD88 to cause its ubiquitylation is unknown. Here I aim to produce oligomeric MyD88 and reconstitute the $SPOP^{18-359}$ oligomer with MyD88 for cryo-EM structure determination.

5.2 - Producing MyD88 protein for structural studies

5.2.1 - MyD88 protein expression

In order to produce MyD88 protein, two MyD88 plasmids were cloned for *E. coli* protein expression, both with an N-terminal His-tag and a TEV protease cleavage site. The first plasmid was for the expression of FL MyD88 and the second was a truncated version containing amino acids 1-158 ($MyD88^{DD-ID}$), which includes the Death domain (DD) and both SBC-containing unstructured regions (**Fig. 5.1**).

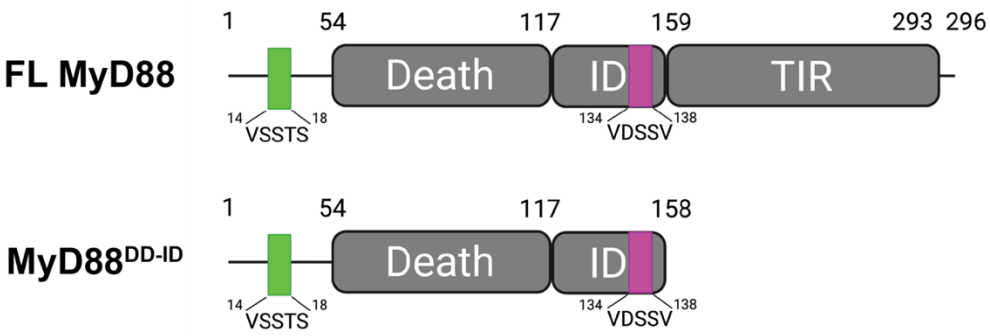


Figure 5.1 – Schematic of the domain organisation of the MyD88 constructs used for protein expression in *E. coli*
Amino acid numbers are indicated. FL is full-length. Death is the Death domain. ID is the intermediate domain. TIR is the Toll/interleukin-1 receptor domain.

SPOP binding consensus motifs highlighted as green and magenta boxes, with the amino acid sequences shown.

5.2.2 - Purifying FL MyD88

Expression trials and Ni^{2+} affinity purification of His^6-tagged FL MyD88 showed it to be largely insoluble when expressed in BL21 (DE3) cells (**Fig. 5.2 a, b**).

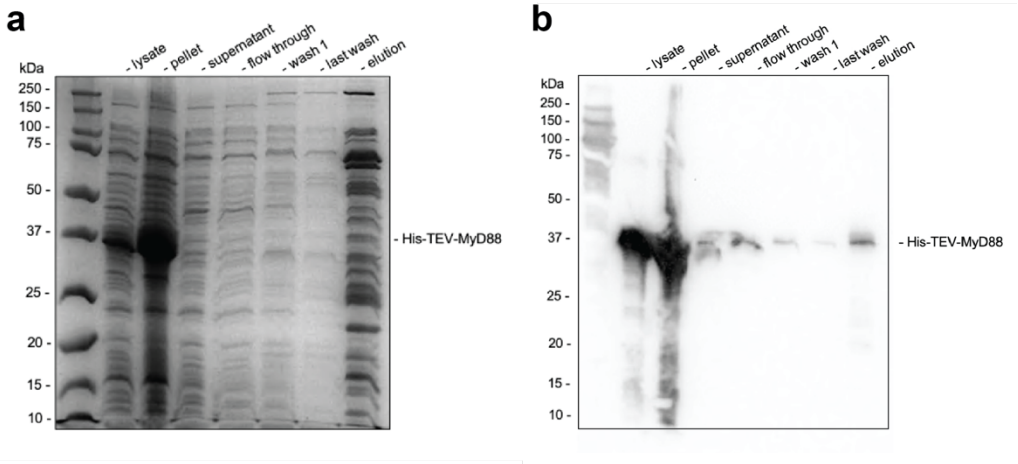


Figure 5.2 – Ni^{2+} affinity purification of FL MyD88 from *E. coli* cells
a SDS-PAGE gel of a nickel-affinity purification of His-TEV-MyD88 from BL21 (DE3) cells.
b anti-His Western blot of the samples shown in **a**.

MyD88 proteins containing either just the MyD88 Death domain or just the MyD88 TIR domain have previously been expressed and purified from bacteria, with both domains able to form long filamentous helical homo-oligomers, as shown by their respective cryo-EM structures [205, 206]. This suggests that both of these domains can be folded by bacterial chaperones but perhaps the presence of both domains leads to the formation of inclusion bodies. Efforts were therefore focussed on producing the MyD88 truncation, $MyD88^{DD\text{-}ID}$, in the hopes that this truncation would be more soluble while still containing the SPOP interaction motif.

5.2.3 - Initial purification of MyD88^{DD-ID}

A truncation of MyD88 which includes the Death oligomerisation domain and the SPOP binding ID, MyD88^{DD-ID}, was cloned for expression in *E. coli* cells. The protein was purified via its N-terminal His-tag, followed by tag cleavage using TEV protease. Following a second Ni^{2+} purification to remove any His-tag and uncleaved protein, the MyD88^{DD-ID} protein was further purified by SEC. A study which purified a similar construct to the MyD88^{DD-ID} showed that the protein oligomerises at lower salt concentrations (20 mM NaCl) in a concentration-dependent manner [205]. To explore this dependency of oligomerisation on salt condition, the protein was subjected to SEC under two buffer conditions: one with 150 mM NaCl and one with 20 mM NaCl.

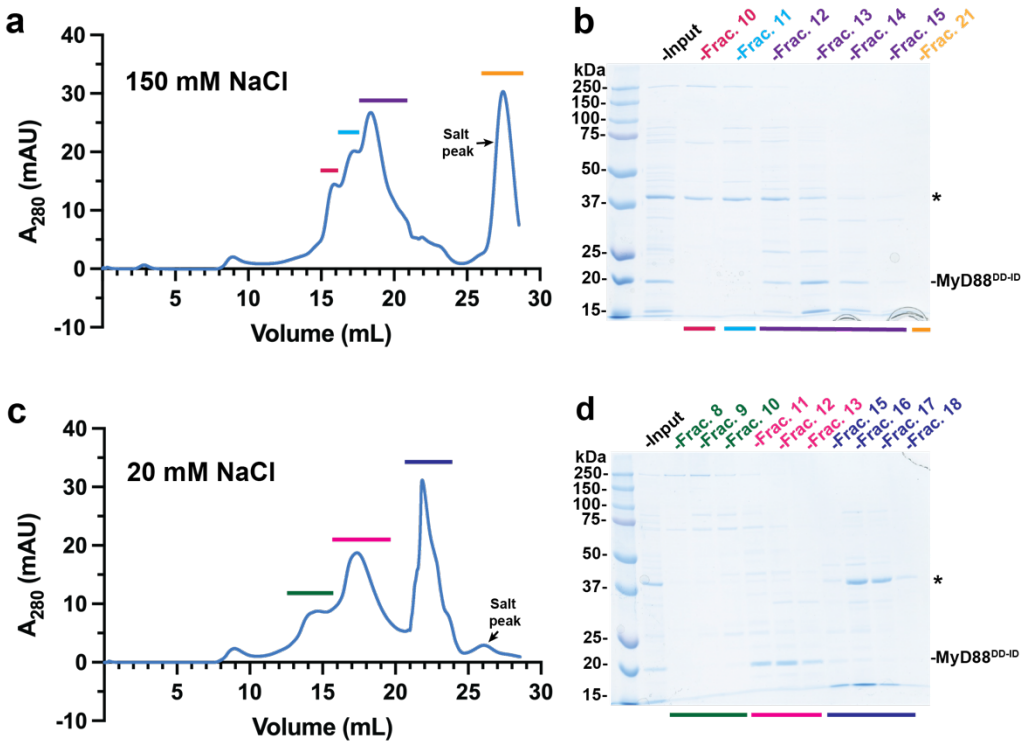


Figure 5.3 – Size-exclusion chromatography purification of MyD88^{DD-ID} under different salt conditions
a Size-exclusion chromatogram of MyD88^{DD-ID} purified in buffer containing 150 mM NaCl. Column was Superose6 Increase 10/300 gl.
b SDS-PAGE gel of fractions from **a**. * is a contaminant band.
c Size-exclusion chromatogram of MyD88^{DD-ID} purified in buffer containing 20 mM NaCl. Column was Superose6 Increase 10/300 gl.
d SDS-PAGE gel of fractions from **c**. * is a contaminant band.

Under both salt conditions, MyD88^{DD-ID} eluted in a peak at ~18 mL (**Fig. 5.3 a, c**). SDS-PAGE analysis of samples from both conditions shows a MyD88 band just below the 20 kDa molecular weight marker (**Fig. 5.3 b, d**). Proteins from the 20 mM NaCl condition eluted slightly earlier compared to the 150 mM salt condition, likely due to the different ionic strength affecting how the proteins interact with the column resin, as well as the different viscosity of the buffers. There is a major contaminant at ~37 kDa which, under higher salt conditions, eluted before MyD88 (**Fig. 5.3 a, b**). However, under lower salt conditions, this protein eluted in a single peak after MyD88 (**Fig. 5.3 c, d**). This could be due to the contaminant interacting with the chromatography column resin under low salt conditions, increasing the retention time. Alternatively, the contaminant might oligomerise under higher salt conditions causing it to elute earlier. Between the two salt conditions, there is no clear formation of MyD88^{DD-ID} oligomers during this purification, likely due to the low protein concentration that was injected onto the column (<0.7 mg/mL; 40 μM), which is exacerbated by dilution during SEC.

5.2.4 - Co-purifying MyD88^{DD-ID} and SPOP^{MATH}

Next, the ability of MyD88^{DD-ID} to bind to SPOP^{MATH} was tested. A small-scale affinity co-purification of His-MBP-SPOP^{MATH} with MyD88^{DD-ID} was attempted. 300 µg of MyD88^{DD-ID} and 100 µg of His-MBP-SPOP^{MATH} were incubated with 100 µL of Ni-NTA beads. After washing the beads of any unbound protein, the beads were boiled in SDS-PAGE buffer to elute any bound protein and the samples were analysed by SDS-PAGE.

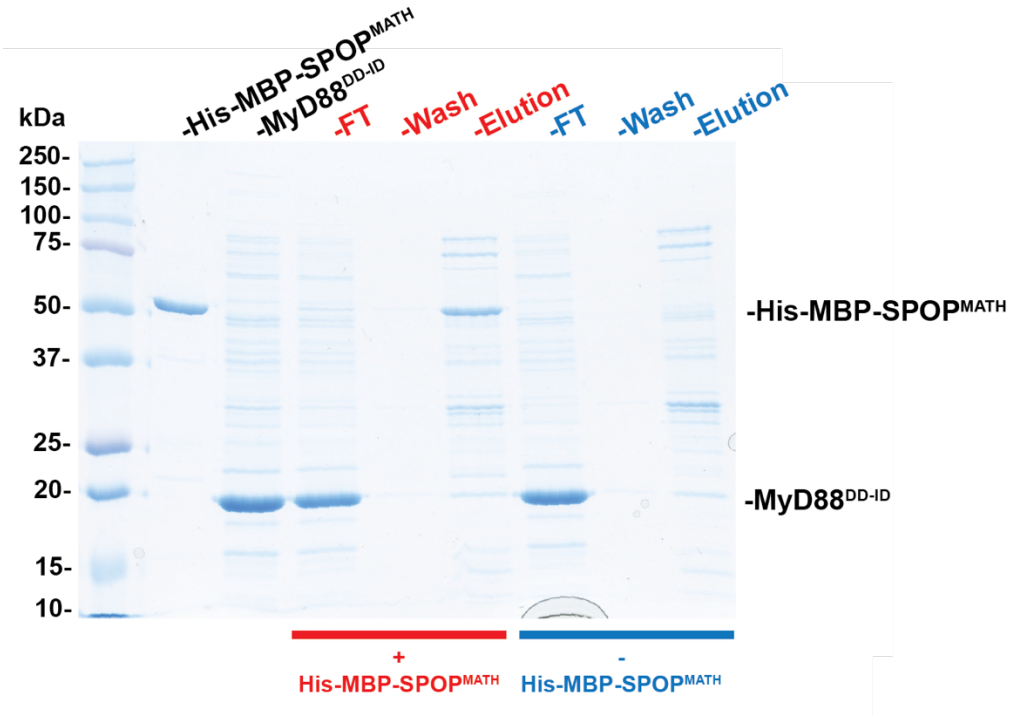


Figure 5.4 – Co-purification of MyD88^{DD-ID} and SPOP^{MATH}
SDS-PAGE gel showing fractions from a co-purification between His-MBP-SPOP^{MATH} and MyD88^{DD-ID}. First lane is the molecular weight standard. Lanes 2 and 3 are protein inputs. Lanes labelled in red include both MyD88 and His-SPOP. Lanes labelled in blue are a MyD88 only control. FT is the flow-through after incubating the protein with Ni-NTA beads.

The elution from the boiled beads shows no noticeable difference in pulling down MyD88^{DD-ID} with or without SPOP^{MATH}, suggesting that there was no binding under these conditions (**Fig. 5.4**). The concentration of MyD88 (17 μM) may have been too low to enable binding to SPOP (2.7 μM). The binding affinity between MyD88 SBC_1 and SPOP^{MATH} was ~2-4 μM, likely explaining why no binding was observed during this experiment (**Fig. 3.2 e, f, Fig. 3.6 a, e**). To date, all studies showing the interaction between SPOP and MyD88 have used either recombinant protein immunoprecipitations or isolation from cell extracts, all of which were visualised by western blotting. Perhaps this reliance on western blotting alludes to the relatively weak affinity of the SPOP-MyD88 interaction under standard purification conditions, at reasonable protein concentrations.

5.2.5 - Initial purification of SPOP$^{18-359}$ and MyD88^{DD-ID} complexes

MyD88^{DD-ID} contains two SBCs in its protein sequence. It was rationalised that the SPOP$^{18-359}$ oligomer, rather than the monomeric SPOP^{MATH}, may allow multivalent MyD88^{DD-ID} binding, leading to increased complex formation. In addition, this multivalent binding to SPOP by MyD88 may increase the local MyD88 concentration at the SPOP oligomer, helping MyD88 itself oligomerise. To explore whether SPOP oligomers were able to bind MyD88^{DD-ID}, SPOP$^{18-359}$ was incubated with MyD88^{DD-ID} and the protein mixture was separated by SEC. Since SPOP$^{18-359}$ oligomers typically elute in the early volumes of the SEC column due to their large size, it was expected that binding of MyD88^{DD-ID} to SPOP would cause it to co-elute at earlier volumes, rather than eluting later on, as was shown previously (**Fig. 5.3**).

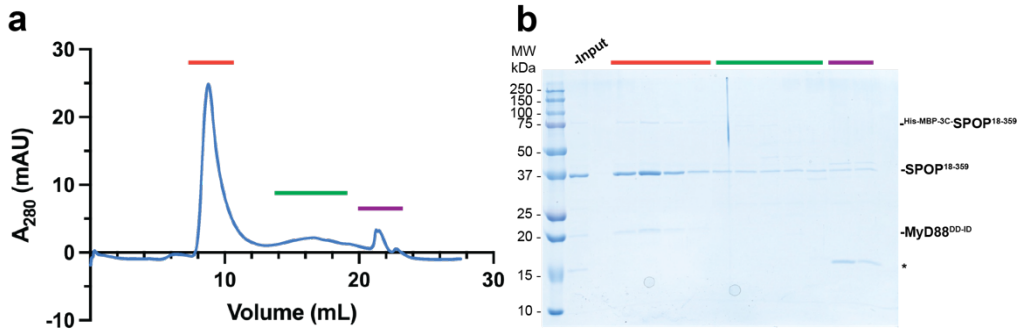


Figure 5.5 – Size-exclusion chromatography of the SPOP$^{18-359}$ and MyD88^{DD-ID} complex
a Size-exclusion chromatogram of the SPOP$^{18-359}$ and MyD88^{DD-ID} complex. Column was Superose6 Increase 10/300 gl.
b SDS-PAGE gel of fractions as in **a**. * indicates contaminant band.

The SEC chromatogram showed a large peak which eluted at ~10 mL, with much smaller peaks eluting later on (**Fig. 5.5 a**). SDS-PAGE analysis of the SEC fractions showed that the majority of SPOP$^{18-359}$ eluted in the high-molecular weight peak, with less SPOP eluting in the later fractions (**Fig. 5.5 b**). Faint bands corresponding to MyD88 are present in this high-molecular weight peak, suggesting that MyD88^{DD-ID}

was bound to SPOP$^{18-359}$ (**Fig. 5.5 b**). Fractions from the peak at ~21 mL showed a contaminant band between 15-20 kDa (**Fig. 5.5 a, b**).

5.3 – Purifying MyD88^{DD-ID} oligomers

As individual proteins, both SPOP and MyD88 have been shown to oligomerise in a protein concentration-dependent manner. Efforts were therefore focussed on increasing the protein concentration of MyD88^{DD-ID}, in the hopes that oligomeric MyD88 would be able to bind to SPOP at high enough quantities for cryo-EM studies. This attempt to obtain higher MyD88^{DD-ID} concentrations is hindered by the inability to concentrate the protein during any of the purification steps, as the centrifugal concentrator membrane clogs once concentrations reach >0.4 mg/mL (23 μM).

Since MyD88^{DD-ID} cannot be concentrated during any stage of the purification, the *E. coli* cell pellet starting material was scaled up and care was taken during the affinity purification to avoid protein dilution (**Fig. 5.6 a**). This resulted in an 114 μM MyD88^{DD-ID} sample for purification by SEC. Injection onto the SEC column and separation in low salt (20 mM NaCl) buffer showed a peak at ~8 mL, with a smaller peak at ~18 mL, which had previously corresponded to monomeric MyD88^{DD-ID} (**Fig. 5.6 b**). The presence of MyD88^{DD-ID} in this peak was confirmed by SDS-PAGE analysis (**Fig. 5.6 c**). The SDS-PAGE gel showed four bands across all fractions from this peak (**Fig. 5.6 c**). The top band likely corresponds to remaining His-TEV-MyD88^{DD-ID} (**Fig. 5.6 c**). During MyD88^{DD-ID} purifications where the protein was more concentrated (>23 μM), it proved difficult to cleave all of the His-tag from the protein. This is in contrast to previous purifications at lower MyD88 concentrations, where the resulting protein was monomeric. Therefore, it is likely that the oligomeric state of MyD88^{DD-ID} hinders access to the TEV protease cleavage site, resulting in heterogenous MyD88 oligomers where some of the protomers still contain their His-tags. Directly below the His-TEV-

MyD88$^{DD\text{-}ID}$ band is the untagged MyD88$^{DD\text{-}ID}$ band (**Fig. 5.6 c**). In addition, there are also two lower molecular weight bands that co-purify with the MyD88 oligomer (**Fig. 5.6 c**).

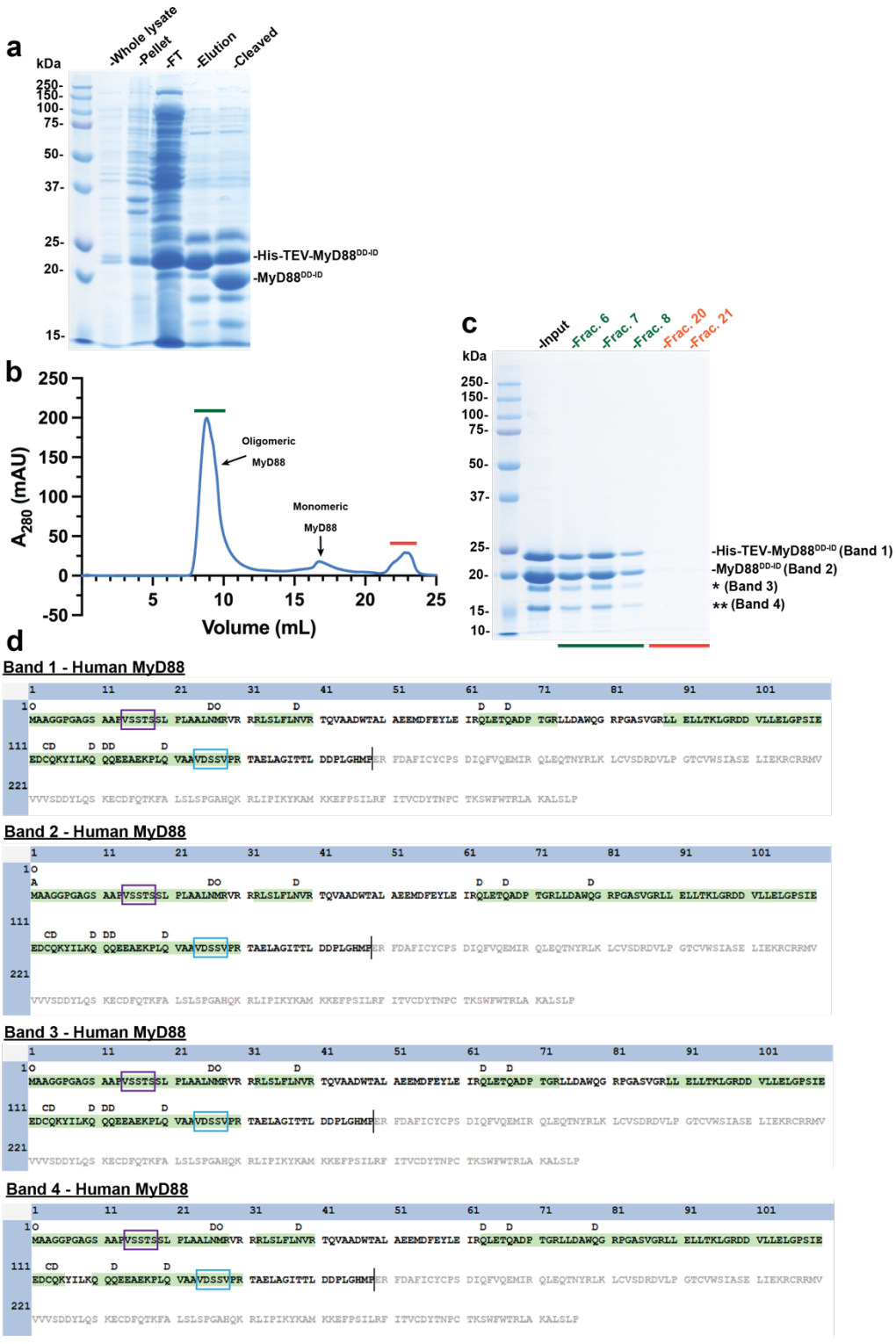


Figure 5.6 – Purification of the MyD88$^{DD\text{-}ID}$ oligomer and mass spectrometry analysis

a SDS-PAGE gel of a nickel-affinity purification of MyD88^{DD-ID} from BL21 (DE3) cells. FT is flow-through. Cleaved is after incubation with TEV protease.
b Size-exclusion chromatogram of MyD88^{DD-ID} purification in buffer containing 20 mM NaCl. Column was Superose6 Increase 10/300 gl.
c SDS-PAGE gel of fractions from **b**.
d In-gel digestion mass spectrometry identification of four protein bands shown in **c**. Performed by Dr Ranjani Ganji. Greyed out protein sequence of human full-length MyD88 is not present in the MyD88^{DD-ID} protein. Peptide hits are highlighted in green. Two MyD88 SPOP binding consensus motifs are boxed in purple and blue.

To confirm the identity of these four bands, in-gel digestion mass spectrometry was used (performed by Dr Ranjani Ganji). The mass spectrometry results indicated that all four bands correspond to *H. sapiens* MyD88 (**Fig. 5.6 d**). Importantly, all four bands have sequence coverage for the two SBCs meaning that MyD88^{DD-ID} should still be able to bind to SPOP$^{18-359}$ (**Fig. 5.6 d**). All four bands also show coverage at the very N-terminus of MyD88 and so MyD88^{DD-ID} likely is not suffering from N-terminal degradation during protein purification (**Fig. 5.6 d**). At the C-terminus of MyD88^{DD-ID}, all four bands are missing peptide coverage from amino acids 141-158 (**Fig. 5.6 d**). While this could occur due to the peptide not “flying” well, it could also indicate C-terminal degradation of the protein. Additionally, bands 1 and 3 are also missing peptide coverage from amino acids 74-88, which would account for a 1.5 kDa reduction in molecular weight (**Fig. 5.6 d**).

The mass spectrometry analysis does not however explain the molecular weight differences observed in the SDS-PAGE gel, as all four bands show a similar peptide coverage (**Fig. 5.6 c, d**). This is further confused by the fact that MyD88^{DD-ID} runs on an SDS-PAGE at a higher molecular weight than predicted. Upon overnight dialysis with TEV protease, a clear shift in molecular weight is observed from a predominant band at ~23 kDa (His-TEV-MyD88^{DD-ID} predicted molecular weight 20.5 kDa) to ~20 kDa (MyD88^{DD-ID} predicted molecular weight 17.5 kDa) (**Fig. 5.6 a**).

The molecular weight difference between bands 1 and 2 could be attributed to cleavage of the His-TEV tag, with both bands containing intact MyD88 (**Fig. 5.6 c**). Perhaps bands 3 and 4 could also represent tagged and untagged MyD88^{DD-ID} however, with C-terminally degraded MyD88^{DD-ID} (**Fig. 5.6 c**). Alternately, bands 1 and 2, which are higher molecular weight than expected, could also represent post-translationally modified MyD88, although the mass spectrometry analysis did not detect any additional modifications. Overall, the mass spectrometry analysis is unable to clarify what the four bands visualised by SDS-PAGE represent (**Fig. 5.6 c**).

5.4 - Cryo-EM of the MyD88^{DD-ID} oligomer

The oligomeric MyD88^{DD-ID} sample was next visualised by cryo-EM. Cryo-EM grids of MyD88^{DD-ID} oligomers at 0.6 mg/mL (34 μM) were prepared and imaged (full data collection and processing parameters in **Table 2.4**). The micrographs show long straight filaments of various lengths, ranging from approximately 20-160 nm (200-1,600 Å) (**Fig. 5.7 a**). The filaments in the micrographs resemble that of the MyD88 Death domain oligomer structure reported by Moncrieffe *et al*. [205], which showed the Death domains assembled as a tower-like structure with a helical arrangement of MyD88 protomers (**Fig. 5.7 b**).

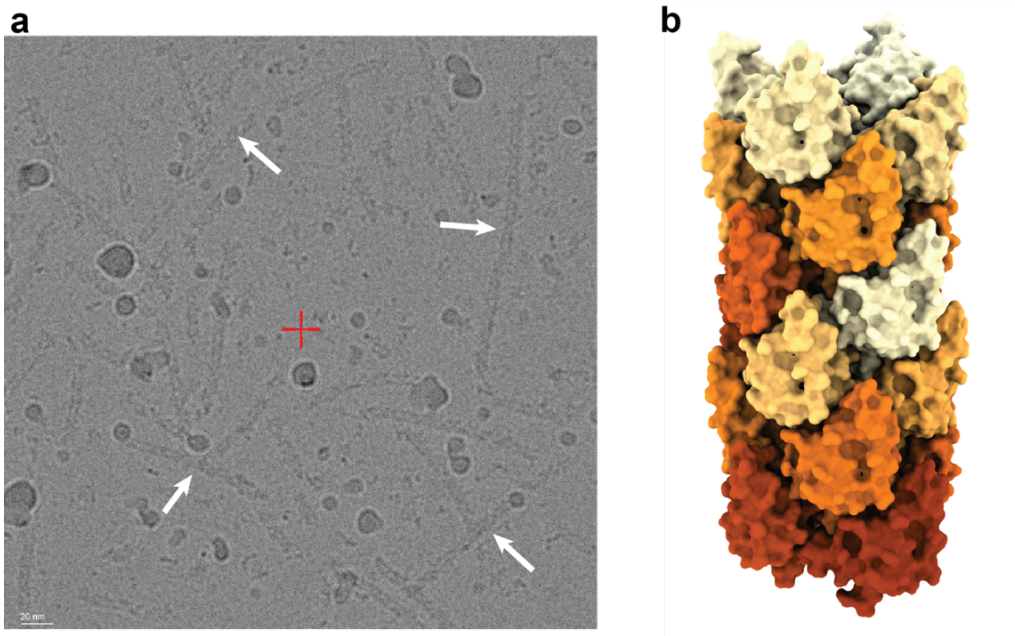


Figure 5.7 – Cryo-EM micrograph of the MyD88^{DD-ID} oligomer
a Cryo-EM micrograph of MyD88^{DD-ID} oligomers. White arrows indicate MyD88^{DD-ID} oligomers.
b Cryo-EM structure of the MyD88 Death domain oligomer (PDB ID 6I3N) [205]. Surface representation of structure is coloured by MyD88 protomer.

Overall, the cryo-EM micrograph confirmed that MyD88^{DD-ID} was forming ordered oligomers (**Fig. 5.7 a**). Since the structure of the MyD88 Death domain oligomer has already been elucidated, efforts were then focussed on gaining the structure of the MyD88^{DD-ID} oligomer in complex with oligomeric SPOP$^{18-359}$.

5.5 - Purification of the SPOP$^{18-359}$ and MyD88^{DD-ID} oligomer complex

In order to reconstitute a SPOP-MyD88 complex, the SPOP oligomer (SPOP$^{18-359}$) and truncated oligomeric MyD88 (MyD88^{DD-ID}) were individually recombinantly expressed and isolated from *E. coli* cells. To form the SPOP-MyD88 complex, the two SEC purified proteins were combined, followed by a final SEC purification (**Fig. 5.8 a**). A peak beginning at 8 mL and tailing off at later elution volumes corresponded to the SPOP-MyD88 complex, as shown by SDS-PAGE (**Fig. 5.8 a, b**).

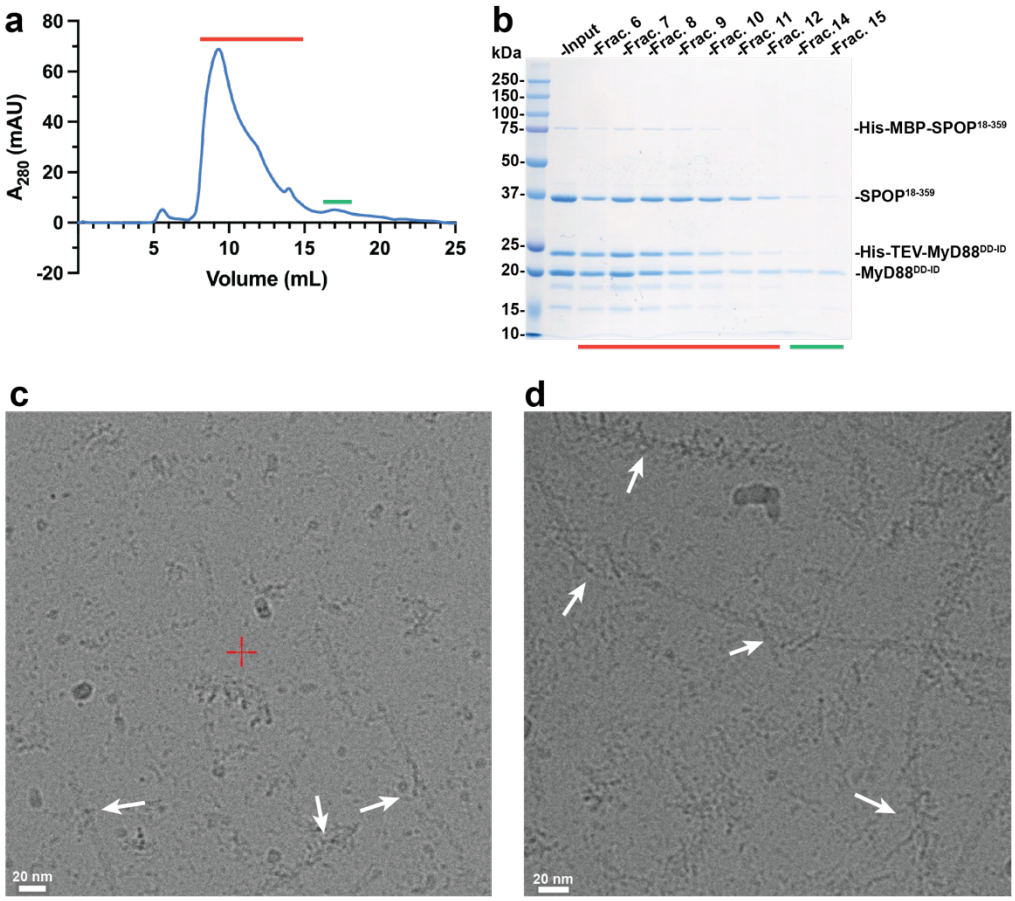


Figure 5.8 – Co-purification of the SPOP$^{18-359}$-MyD88^{DD-ID} oligomer complex
a Size-exclusion chromatogram of SPOP$^{18-359}$ and MyD88^{DD-ID}, using a Superose6 Increase 10/300 gl column.
b SDS-PAGE gel of protein fractions, as in **a**.
c Cryo-EM micrograph of the SPOP$^{18-359}$-MyD88^{DD-ID} complex, purified as in **a** and **b**. White arrows indicate potential complex.
d Cryo-EM micrograph of the SPOP$^{18-359}$-MyD88^{DD-ID} complex formed by mixing of the two individually purified protein. White arrows indicate potential complex.

Fraction 7 of this SPOP-MyD88 complex was then used to make cryo-EM grids (**Fig. 5.8 b**). The cryo-EM micrographs show straight MyD88^{DD-ID} oligomers and curved SPOP$^{18-359}$ oligomers of varying lengths (**Fig. 5.8 c**). Regions which appear to resemble a SPOP-MyD88 complex show a curved SPOP oligomer seemingly wrapping around a straight MyD88 filament (**Fig. 5.8 c**). Overall, the oligomers seem fairly short and there appears to be significant dissociation of the oligomers, as seen by the background noise in the cryo-EM micrograph (**Fig. 5.8 c**). This may be due to the fact that the SEC peak fraction used to make these cryo-EM grids eluted from the column at 0.33 mg/mL (19 μM), causing oligomer dissociation of both proteins.

Again, efforts were focussed in increasing the respective protein concentrations in order to obtain longer oligomers. During previous protein preparations, the purity of the MyD88^{DD-ID} oligomer did not seem to improve after the SEC step, compared to the input sample (**Fig. 5.6 b, c**). Therefore, the more concentrated "input" stock of MyD88^{DD-ID} (2 mg/mL; 114 μM) was mixed with purified SPOP$^{18-359}$ oligomer (0.72 mg/mL; 19 μM) at 3:1, 1:1 or 1:3 volumetric ratios. Rather than following this with an additional SEC step, these mixed samples were directly used to make cryo-EM grids. The resulting cryo-EM micrographs appeared to have longer oligomers with more clear complex formation, with the best ratio being a 1:3 ratio of MyD88 to SPOP (**Fig. 5.8 d**).

5.6 - Cryo-EM dataset of the SPOP$^{18-359}$-MyD88^{DD-ID} oligomer complex

A cryo-EM dataset of 5,624 movies was collected for the SPOP$^{18-359}$-MyD88^{DD-ID} complex (full data collection and processing parameters in **Table 2.4**). Single particles were picked centered on the straight MyD88 filaments and the particles were extracted with a box size large enough to encompass MyD88 and the surrounding SPOP oligomer region (**Fig. 5.9 a**).

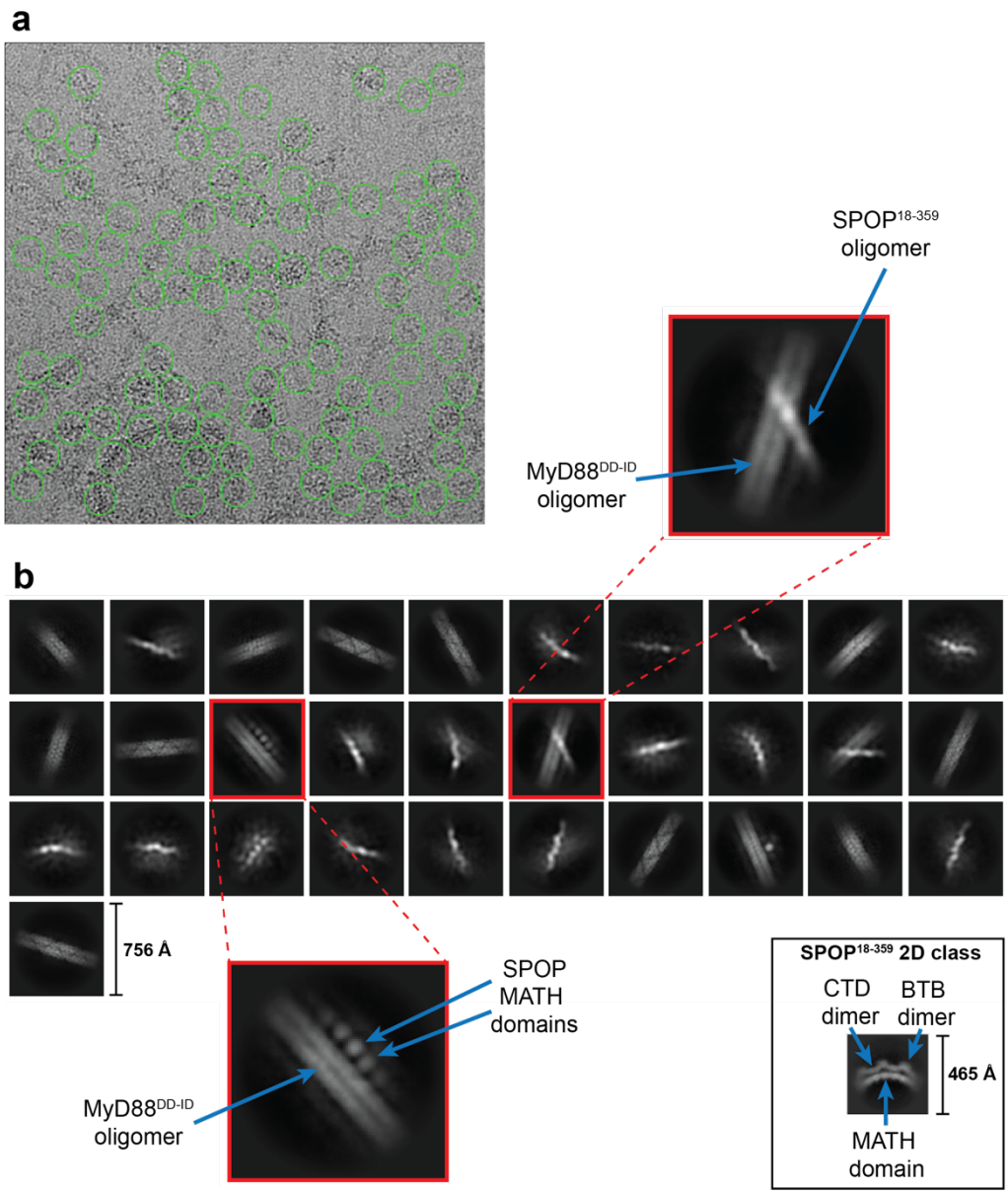


Figure 5.9 – Cryo-EM dataset of the high-order SPOP$^{18-359}$-MyD88^{DD-ID} oligomer complex
a Representative micrograph showing autopicked particles as green circles.
b Selected 2D classes. Red boxes show an enlarged view of 2D classes that potentially show signal for both MyD88^{DD-ID} and SPOP$^{18-359}$. A 2D class from the SPOP$^{18-359}$ cryo-EM dataset is shown at the bottom right for comparison, highlighting the different SPOP domains. The box sizes used for the SPOP$^{18-359}$-MyD88^{DD-ID} dataset (756 Å) and the SPOP$^{18-359}$ (465 Å) dataset are shown.

2D classes were then generated from the picked particles. The resulting 2D classes showed good secondary structure features for the MyD88^{DD-ID} oligomer in some classes (**Fig. 5.9 b**). Signal for SPOP was less clear

in these classes, with some classes showing uniform circular-shaped signal along the MyD88 filament, which may represent the MATH domains of SPOP binding to $MyD88^{DD-ID}$ (**Fig. 5.9 b**). There are also some classes which show the $SPOP^{18-359}$ oligomer diagonally overlapping with the $MyD88^{DD-ID}$ oligomer, which likely represents a SPOP-MyD88 complex (**Fig. 5.9 b**). Importantly, the dimensions of SPOP observed in these 2D classes is similar to what was observed for the $SPOP^{18-359}$ cryo-EM dataset 2D classes (**Fig. 5.9 b**).

Again, no helical symmetry was imposed during 3D map reconstruction as MyD88 and SPOP have different helical parameters. Reconstruction of a 3D density map followed by refinement resulted in a map with a reported resolution of 7.6 Å (**Fig. 5.10**). The reported resolution is likely overestimated as the quality of the map does not reflect this. This overestimated resolution may be the result of using a mask during refinement which is too tight or may indicate the presence of duplicate particles from overlapping particle picking and extraction, which would overinflate the resolution.

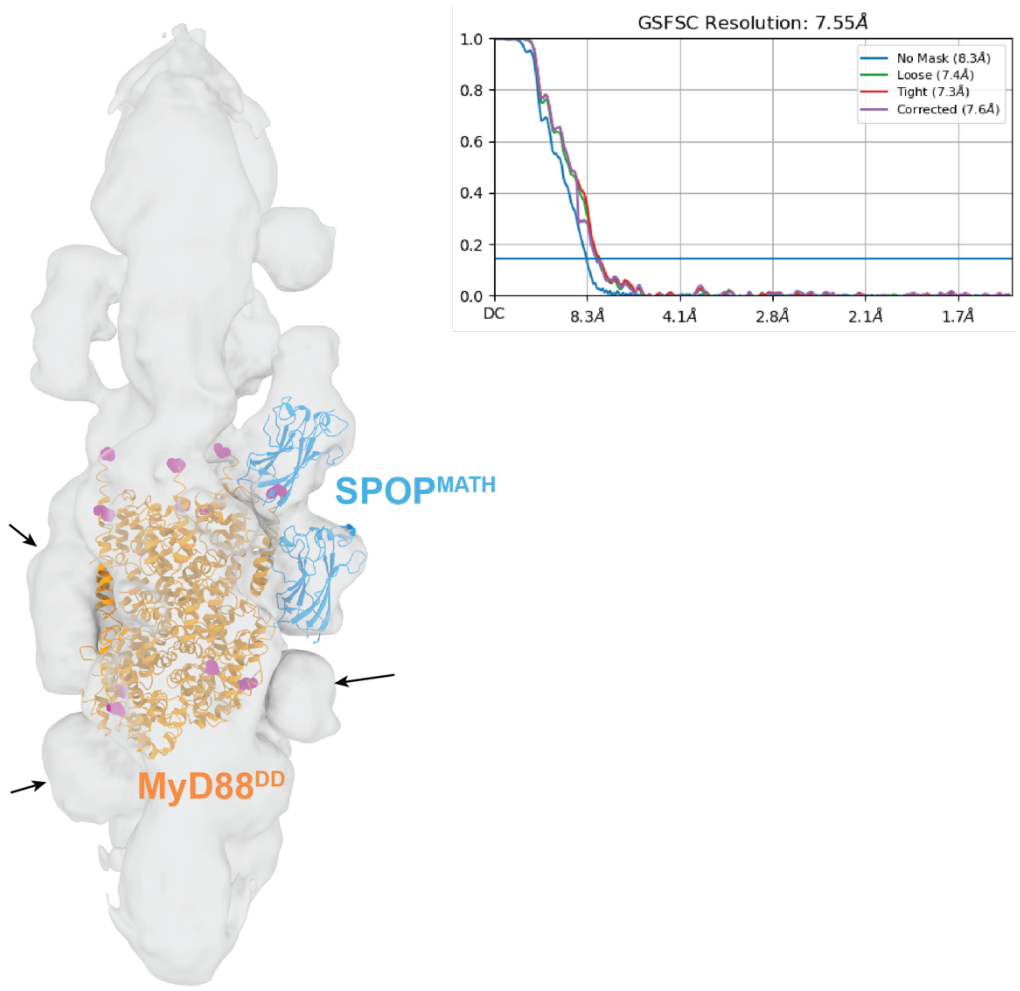


Figure 5.10 – High-order SPOP$^{18-359}$-MyD88^{DD-ID} oligomer complex cryo-EM map
3D refined cryo-EM map of the SPOP$^{18-359}$-MyD88^{DD-ID} oligomer shown in transparent grey. Cryo-EM structure of the MyD88 Death domain oligomer (PDB ID 3MOP) [205] is rigid body fitted into the density. Magenta spheres are the most C-terminal residue present in the MyD88 Death domain oligomer structure, which terminates just prior to the SPOP-binding consensus motif-containing intermediate domain. SPOP^{MATH} domains are rigid body fitted into the density. Black arrows indicate additional density that may also represent SPOP$^{18-359}$. FSC curve is shown, calculated using the gold standard FSC cutoff of 0.143.

Overall, the cryo-EM map has a long and thin shape at its centre with additional blobs of density along the outside (**Fig. 5.10**). The strongest density is located towards the middle section along the length of the oligomer, with the density becoming less discernible at the extreme ends (**Fig. 5.10**). Rigid body fitting of the MyD88 Death domain oligomer structure [205] into this middle section of the density shows a reasonable

fit, with additional circular blobs of density observed just adjacent, similar to what was observed in the 2D classes (**Fig. 5.9 b**, **Fig. 5.10**). Rigid body fitting of $SPOP^{MATH}$ domains into this additional density shows a fairly good fit in terms of the size (**Fig. 5.10**). Additionally, this placement of the $SPOP^{MATH}$ domains is consistent with their expected placement to enable binding of the MyD88 SBCs.

It is clear that further 3D classification steps and the collection of more cryo-EM data is needed to improve the resolution and interpretability of this structure. Overall, this low-resolution map provides a first glimpse at how the $SPOP^{18-359}$ oligomer engages the $MyD88^{DD-ID}$ oligomer.

5.7 - Chapter summary

Here it was shown that a truncation of MyD88 including residues 1-158 ($MyD88^{DD-ID}$) is able to form high-order oligomers. Additionally, it was shown that $MyD88^{DD-ID}$ is able to form a complex with oligomeric $SPOP^{18-359}$. This high-order SPOP-MyD88 complex was imaged using cryo-EM, which revealed a low-resolution density map showing curved SPOP oligomers wrapping around a straight MyD88 oligomer, with the MATH domains of SPOP engaging MyD88.

Chapter 6 – High-resolution cryo-EM structure of the UFL1 E3 ligase complex bound to the 60S ribosome

6.1 - Chapter introduction

UFMylation is the covalent attachment of the ubiquitin-like molecule, ubiquitin fold modifier 1 (UFM1), onto lysine residues of target proteins [121]. This unusual post-translational modification is catalysed by a single known E3 ligase, the UFL1 E3 ligase complex. The UFMylation target of this ligase was recently shown to be endoplasmic reticulum (ER) bound ribosomes, in a mechanism implicated in the ribosome quality control pathway after protein translation stalling [137, 138]. The structure of the UFL1 ligase complex and the mechanism of ribosome UFMylation remain unknown. To explore this, cryo-EM was used to elucidate the high-resolution structure of the UFL1 ligase complex bound to the 60S ribosome subunit.

6.2 - The UFL1 E3 ligase complex

The UFL1 E3 ligase complex used for cryo-EM studies includes five proteins, all of which are required for ribosome UFMylation. First, is UFL1 which is the putative E3 ligase. This E3 ligase is unique as it does not share any domain homology with the RING and HECT classes of ubiquitin E3 ligases (**Fig. 6.1**). Next is UFBP1, a UFL1 adaptor protein which localises the entire ligase complex to the ER membrane via an N-terminal transmembrane helix (TMH) (**Fig. 6.1**) [132]. CDK5RAP3 is an adaptor protein that directs ligase activity towards the ribosome [131]. Lastly, the complex includes a UFC1-UFM1 mimic where the E2 enzyme, UFC1, is linked to the C-terminal G-83 of UFM1 through an engineered isopeptide bond to prevent discharge of UFM1 onto the ribosome (**Fig. 6.1**). UFC1 contains a UBC domain that is similar to ubiquitin E2

enzymes (**Fig. 6.1**). However, UFC1 lacks a recognisable HPN motif that is present in ubiquitin E2s, but does contain a catalytic cysteine (C-116).

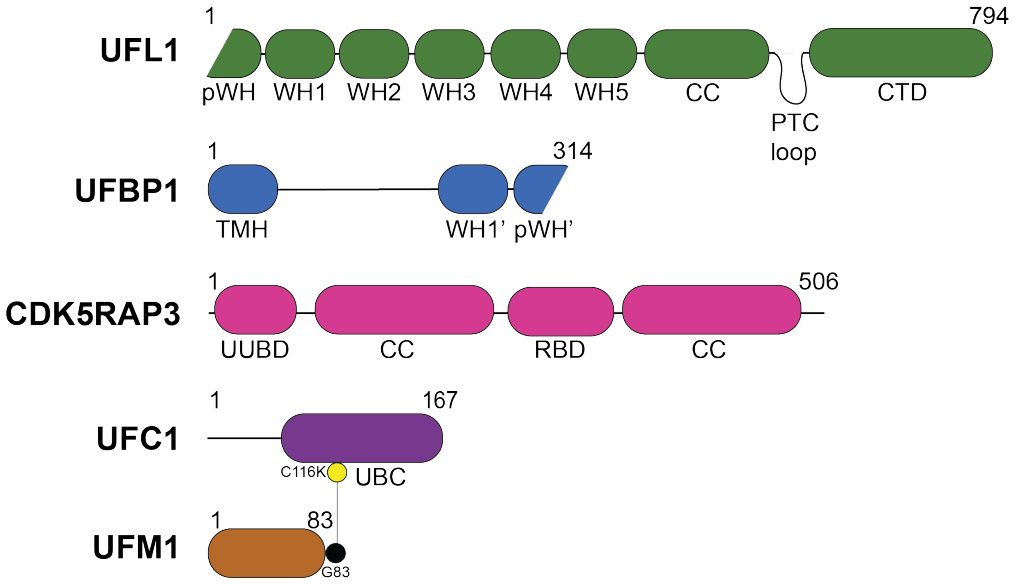


Figure 6.1 – Schematic of protein domain architecture of the UFL1 E3 ligase complex
Amino acid numbers are indicated. WH is winged helix domain. pWH is partial WH. CC is coiled-coil. CTD is C-terminal domain. TMH is trans-membrane helix. UUBD is UFM1-UFBP1 binding domain. RBD is ribosome binding domain. UBC is ubiquitin-conjugation domain. C116K is UFC1 catalytic cysteine-116 to lysine mutation, represented by yellow circle. Black circle represents UFM1 C-terminal glycine-83. Line connecting yellow and black circles represents isopeptide link.

6.3 - Cryo-EM structure determination of the UFL1 ligase – ribosome complex

6.3.1 - UFL1 ligase complex-60S ribosome sample preparation

All protein expression and purification steps were performed by Dr Joshua Peter (Prof Yogesh Kulathu lab). The recombinant *E. coli* expressed UFL1 E3 ligase complex consisting UFL1, UFBP1, CDK5RAP3 and UFC1-UFM1 was *in vitro* reconstituted. All proteins were FL except UFBP1 which began at amino acid 29, removing the N-terminal trans-membrane helix. The UFL1 ligase complex was combined with 60S ribosome subunits, isolated from HEK293 cells. This UFL1 ligase-60S ribosome complex was stabilised by glutaraldehyde

crosslinking and purified by sucrose gradient centrifugation before being imaged using cryo-EM.

6.3.2 - Cryo-EM data collection and processing

A dataset of ~59,000 cryo-EM micrographs was collected (full data collection and processing parameters in **Table 2.4**). Around 2.2 million particles were picked using crYOLO's automated particle picking software [190] (**Fig. 6.2**). Seven rounds of reference-free 2D classification resulted in ~1.6 million selected particles. These particles were then used to generate an initial 3D model, followed by a refined 3D model. During 3D refinement, the dynamic mask start resolution parameter was set to a value below the resolution of the data (i.e., 1 Å) to generate an output mask for subsequent 3D variability analysis (3DVA). This mask would allow for the 3DVA algorithm to search for variability within the entire 3D volume box. The first round of 3DVA separated three major classes: 80S ribosome, 60S ribosome subunit and UFL1 ligase-bound 60S (**Fig. 6.2**). These three classes then underwent further rounds of 3D classification using 3DVA, as well as a final classification step using cryoDRGN, to obtain a homogenous particle population consisting of 299,008 particles. This was then followed by a final round of 3D refinement with per-particle defocus refinement, Ewald sphere correction and CTF refinement to generate a ligase-bound 60S ribosome map at a nominal resolution of 2.2 Å (**Fig. 6.2**).

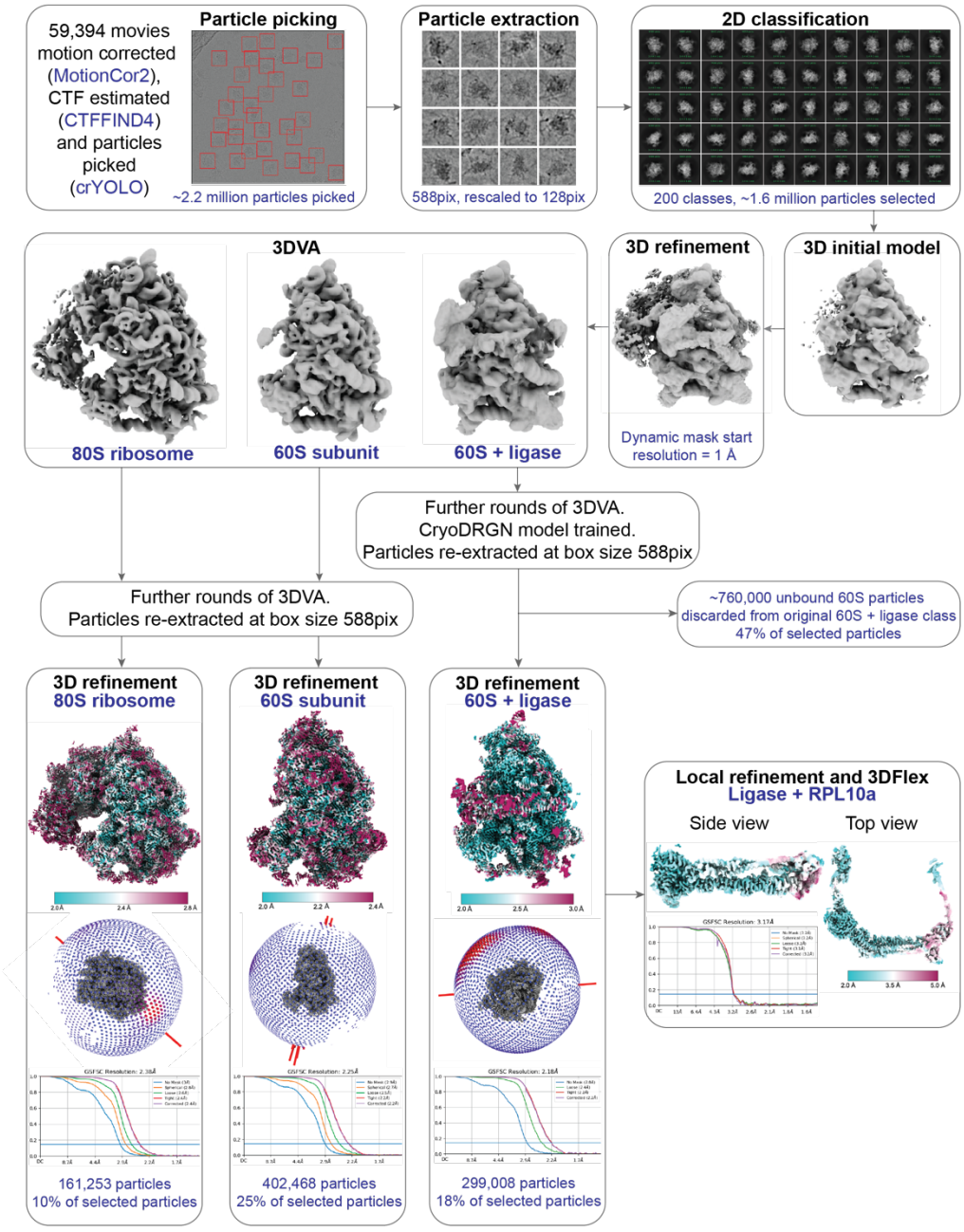


Figure 6.2 – UFL1 ligase complex-bound 60S ribosome cryo-EM data processing pipeline

Cryo-EM data processing steps to obtain cryo-EM maps for the 80S ribosome, 60S ribosome subunit, UFL1 ligase-bound 60S ribosome and the UFL1 ligase complex bound to RPL10a. ~2.2 million picked particles were extracted using a box size of 588 pixels (pix), rescaled to 128 pix. After several rounds of 2D classification, ~1.6 million ribosome like particles were selected. All ribosome-like particles were pooled to generate an initial 3D map, followed by 3D refinement. During 3D refinement, the dynamic mask start resolution parameter was set to 1 Å to generate a mask for subsequent 3D variability analysis (3DVA), which separated three major classes: 80S ribosome, 60S subunit and UFL1 ligase-bound 60S (60S+ligase). These underwent further rounds of 3DVA, with the 60S+ligase particles undergoing an additional classification step using cryoDRGN. The final particle stacks were then re-extracted using the original box size, followed by final 3D refinement. To generate the ligase+RPL10a map, signal corresponding to the 60S ribosome was subtracted

from the 60S+ligase particles and the region corresponding to the UFL1 ligase and RPL10a was locally refined. This was followed by 3D flexible refinement to generate the final map. Final maps were sharpened and are coloured by local resolution. 3D representation of the angular distributions and FSC curves are shown for final maps, calculated using the gold standard FSC cutoff of 0.143.

The resulting cryo-EM map showed high-resolution detail for the majority of the ribosome density, with additional density corresponding to the ligase complex (**Fig. 6.3**). However, the density for the ligase complex was at a lower resolution when compared to the ribosome and was fragmented in places, with large regions of the ligase complex unaccounted for (**Fig. 6.3**).

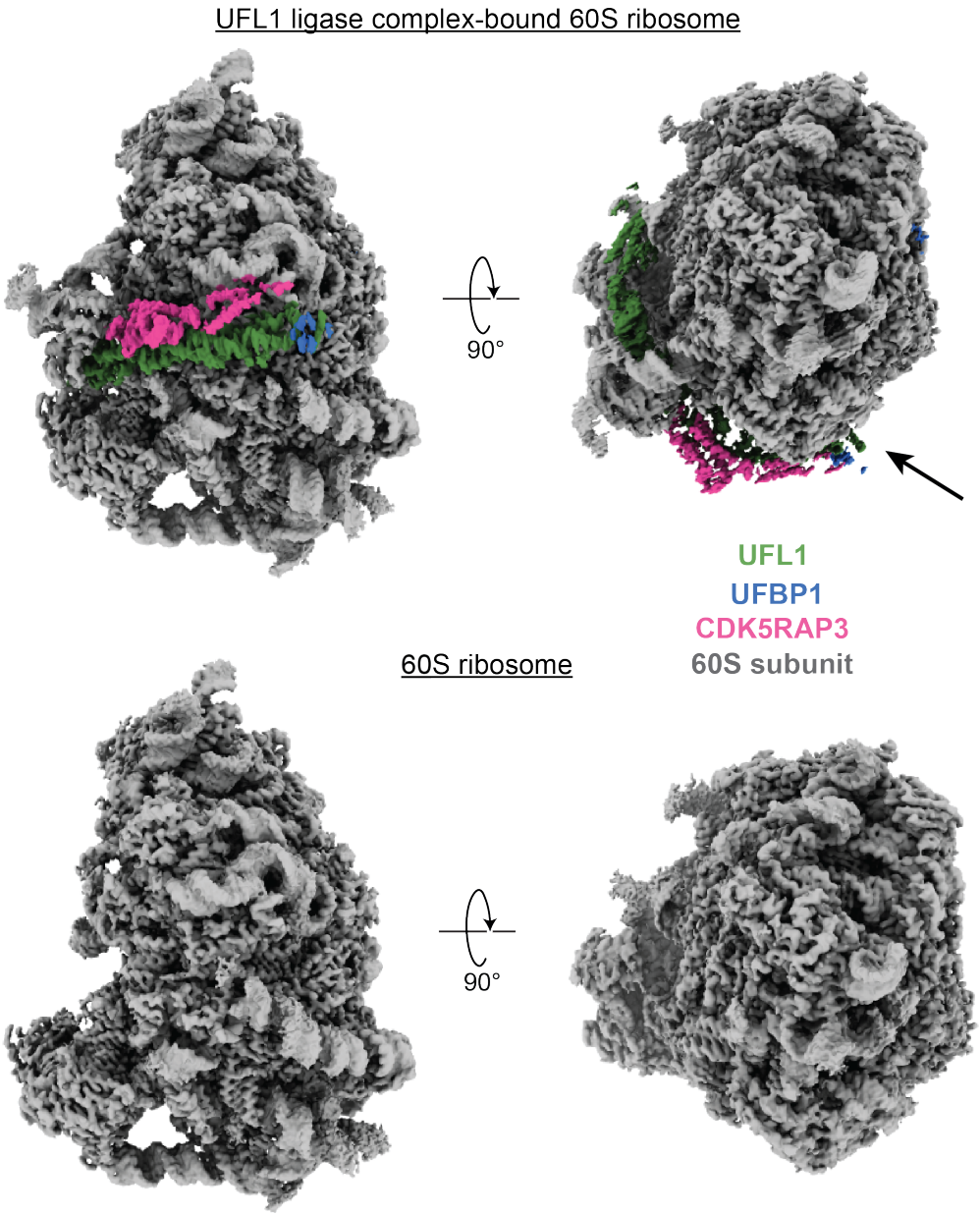


Figure 6.3 – Comparison of UFL1 ligase complex-bound 60S ribosome and 60S ribosome only cryo-EM maps
The UFL1 ligase complex-bound 60S cryo-EM map shows additional density in comparison to the 60S only map. Black arrow indicates fragmented and missing ligase density.

This missing density was either due to the relative flexibility of the ligase complex or a lower occupancy of the ligase bound to the ribosome. For the latter instance, further classification steps did not separate out free 60S ribosome particles from ligase-bound 60S particles. In an attempt to improve the ligase density, signal corresponding to the ribosome was subtracted and local refinement of just the ligase region was performed.

This resulted in a 3.2 Å reconstruction of the ligase complex, which showed more complete density and high-resolution features in most regions (**Fig. 6.4 a**). However, some regions of the ligase still displayed poor density, as shown by the lower local resolution, hindering atomic model building in these areas (**Fig. 6.4 a**).

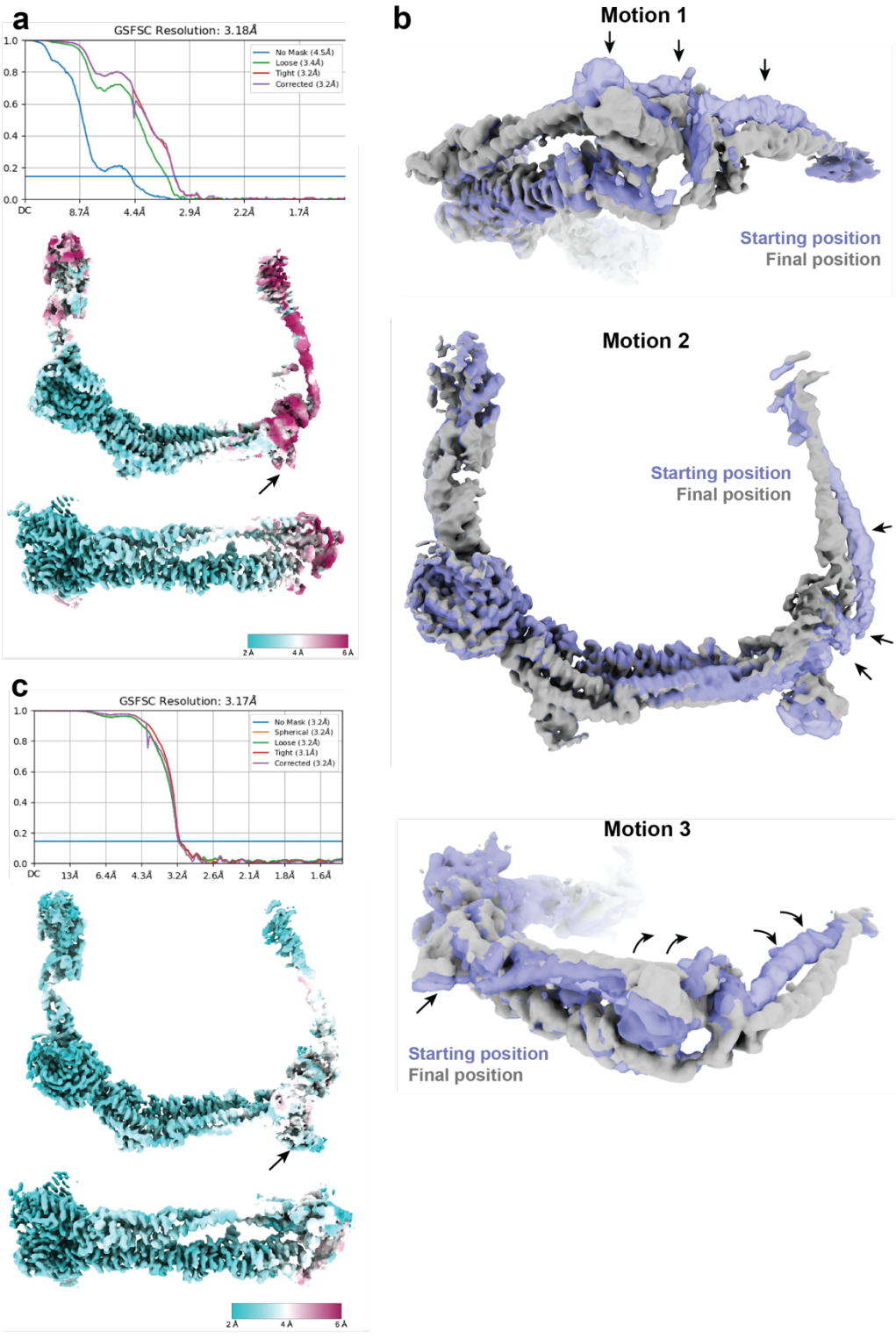


Figure 6.4 – Local refinement of the UFL1 ligase complex and modelling motion of the UFL1 ligase complex using 3DFlex
a Cryo-EM map of the UFL1 ligase complex after local refinement, coloured by local resolution. FSC curve is shown, calculated using the gold-standard FSC cut-off at 0.143. Black arrow points to missing CDK5RAP3 density.
b 3DFlex generated motion models of the ligase complex. Motion of the ligase complex is illustrated by showing the start (transparent purple) and end (solid grey) positions of the ligase complex. Black arrows indicate direction of motion.
c Cryo-EM map of the ligase complex after 3DFlex reconstruction, coloured by local resolution. FSC curve is shown, calculated using the gold-standard FSC cut-off at 0.143. Black arrow points to resolved CDK5RAP3 density.

To try to account for the flexibility of the ligase complex, CryoSPARC's 3DFlex [191] algorithm was used to model the flexible motion of the ligase and to generate a consensus 3D reconstruction. As expected, the flexibility modelling showed large motions in the regions of the ligase that were lower resolution (**Fig. 6.4 a, b**). The resulting 3DFlex reconstruction showed a much-improved local resolution in the flexible regions (**Fig. 6.4 c**). There is also additional density that is reconstructed, which corresponds to part of CDK5RAP3 (**Fig. 6.4 a, c**)**.** This improved ligase-only cryo-EM map allowed for atomic model building of the majority of the ligase complex (**Fig. 6.4 a**).

6.4 - Overview of UFL1 ligase complex-bound 60S ribosome structure

The UFL1 ligase complex wraps around the ribosome, contacting the tRNA binding sites at one end and the peptide exit tunnel/SEC61 translocon binding region at the other (**Fig. 6.5 a, b**). UFL1 and UFBP1 form the main ligase scaffold, with CDK5RAP3 bridging across these two proteins (**Fig. 6.5 a**). UFM1 is situated between CDK5RAP3 and UFBP1 (**Fig. 6.5 a).** However, there was no clear density for UFC1 and so this could not be built into the atomic model (**Fig. 6.5 a**).

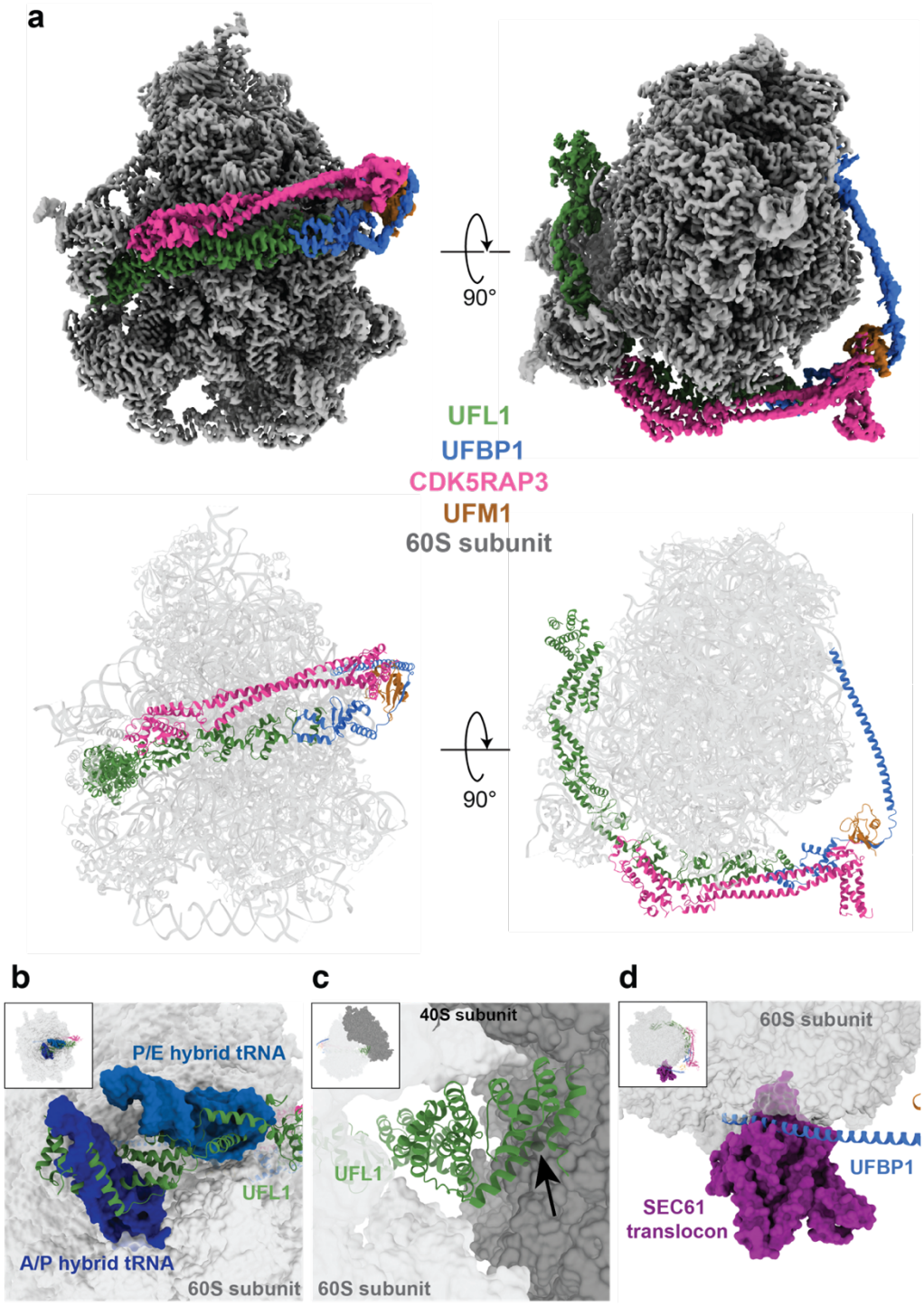


Figure 6.5 – Cryo-EM map and model of the UFL1 E3 ligase complex bound to the 60S ribosome

a Top - Composite cryo-EM density maps of the UFL1 ligase complex bound to the 60S ribosome. The 60S ribosome density map is coloured in grey; the ligase density map is coloured by protein. Bottom – Atomic model of the UFL1 ligase complex bound to the 60S ribosome.

b Superpositions of this study's structure with published ribosome structures, highlighting clashes with 60S ribosome binding components. UFL1 C-terminal domain (CTD) occupies the aminoacyl (A)-, peptidyl (P)- and exit (E)-tRNA binding sites, as illustrated by clashes with superimposed A/P and P/E tRNA (PDB ID 6W6L) [207].

c UFL1 CTD clashes with the 40S subunit (PDB ID 6IP8) [208]. Arrow points to clash.
d UFBP1 α-helix clashes with the SEC61 translocon (PDB ID 6R7Q) [209], which binds at the peptide exit tunnel.

The CTD of UFL1 sits in the tRNA binding sites and superposition of tRNA molecules show that UFL1 would clash with tRNA bound to the ribosome, suggesting that tRNA would have to dissociate prior to ligase binding (**Fig. 6.5 b**). In addition, when superimposing the 40S subunit of the ribosome, this same region of the ligase shows clashes with the 40S subunit (**Fig. 6.5 c**). Within the cryo-EM dataset, a population of 80S ribosome particles were reconstructed to a resolution of 2.4 Å (**Fig. 6.2**). This final 3D reconstruction of the 80S ribosome, as well as earlier iterations during the data processing pipeline, did not show any ligase density present. This suggests that the UFL1 ligase complex preferentially binds to the 60S ribosome subunit rather than 80S ribosomes and that re-association of the 40S subunit would be blocked by the ligase. Although, it cannot be discounted that other conformational states of the 40S subunit bound to the 60S, such as the various rotated states of the 40S, may accommodate UFL1 binding.

At the other end of the ligase complex, a long UFBP1 α-helix binds proximal to the ribosome peptide exit tunnel, where newly formed polypeptide chains egress from the ribosome (**Fig. 6.5 d**). This region of the ribosome is also where the SEC61 translocon is known to interact. The SEC61 translocon localises ribosomes to the ER membrane and allows translocation of nascent polypeptides from the peptide exit tunnel into the ER lumen. Therefore, the structure suggests that the ribosome would have to be released from SEC61 for the ligase complex to associate. Collectively, the structural analysis suggests that the UFL1 ligase complex preferentially recognises 60S ribosomes rather than 80S ribosomes, after removal of the 40S subunit, tRNAs and the SEC61 translocon (**Fig. 6.6**).

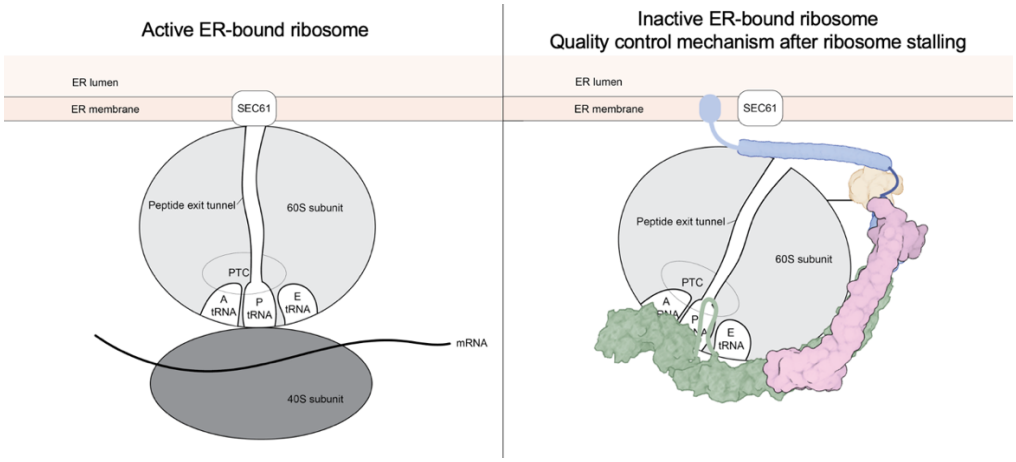


Figure 6.6 – Schematic of an ER-bound 80S ribosome and a UFL1 ligase complex-bound 60S ribosome
Left is an ER-bound 80S ribosome that is translationally active. Right is a UFL1 ligase complex-bound 60S subunit that is incapable of initiating translation. UFL1 is in green. CDK5RAP3 is in magenta. UFBP1 is in blue. UFM1 is in beige. The UFL1 ligase complex keeps the 60S subunit localised to the ER-membrane via a transmembrane helix of UFBP1. PTC is peptidyl transferase centre.

6.4.1 - UFL1-UFBP1 interactions

The structure reveals the presence of a total of seven WH domains, five from UFL1 (WH1-5), one from UFBP1 (WH1'), and a composite WH domain (pWH/pWH') (**Fig. 6.7 a**). All WH domains except UFBP1 WH1' contact the ribosome. The interaction interface between UFL1 and UFBP1 forms across the partial WH domains from the N-terminus of UFL1 (pWH) and the C-terminus of UFBP1 (pWH'), which come together to form a composite WH domain (pWH/pWH') (**Fig. 6.7 b**). This interface is mainly mediated by hydrophobic residues (**Fig. 6.7 b**).

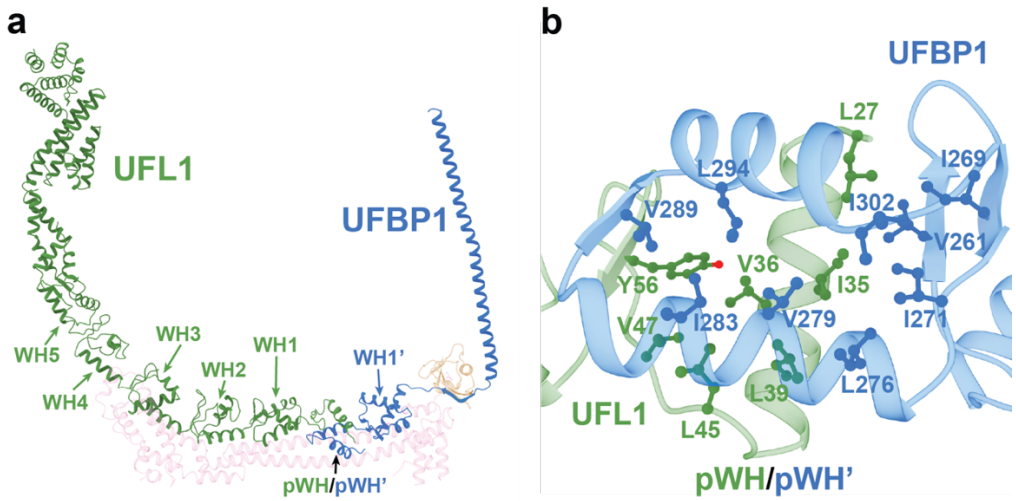


Figure 6.7 – UFL1 and UFBP1 winged-helix domains and binding interface
a Structure of the UFL1 ligase complex with winged helix (WH) domains annotated. pWH is partial WH.
b Close up view of the composite WH formed of pWHs from UFL1 and UFBP1. Main interacting residues at this interface are shown as ball and stick.

6.4.2 - CDK5RAP3 is a ribosome substrate adaptor

CDK5RAP3 forms the central part of the ligase complex, acting as a bridge that contacts UFL1, UFBP1 and UFM1 (**Fig. 6.8 a**). Furthermore, CDK5RAP3 also helps to anchor the ligase complex onto the ribosome through a CDK5RAP3 module that interacts with RPL10a, termed the RPL10a binding domain (RBD) (**Fig. 6.8 a**). Additionally, CDK5RAP3 also has a coiled coil domain (CCD) and a UFM1-UFBP1 binding domain (UUBD) (**Fig. 6.8 a**).

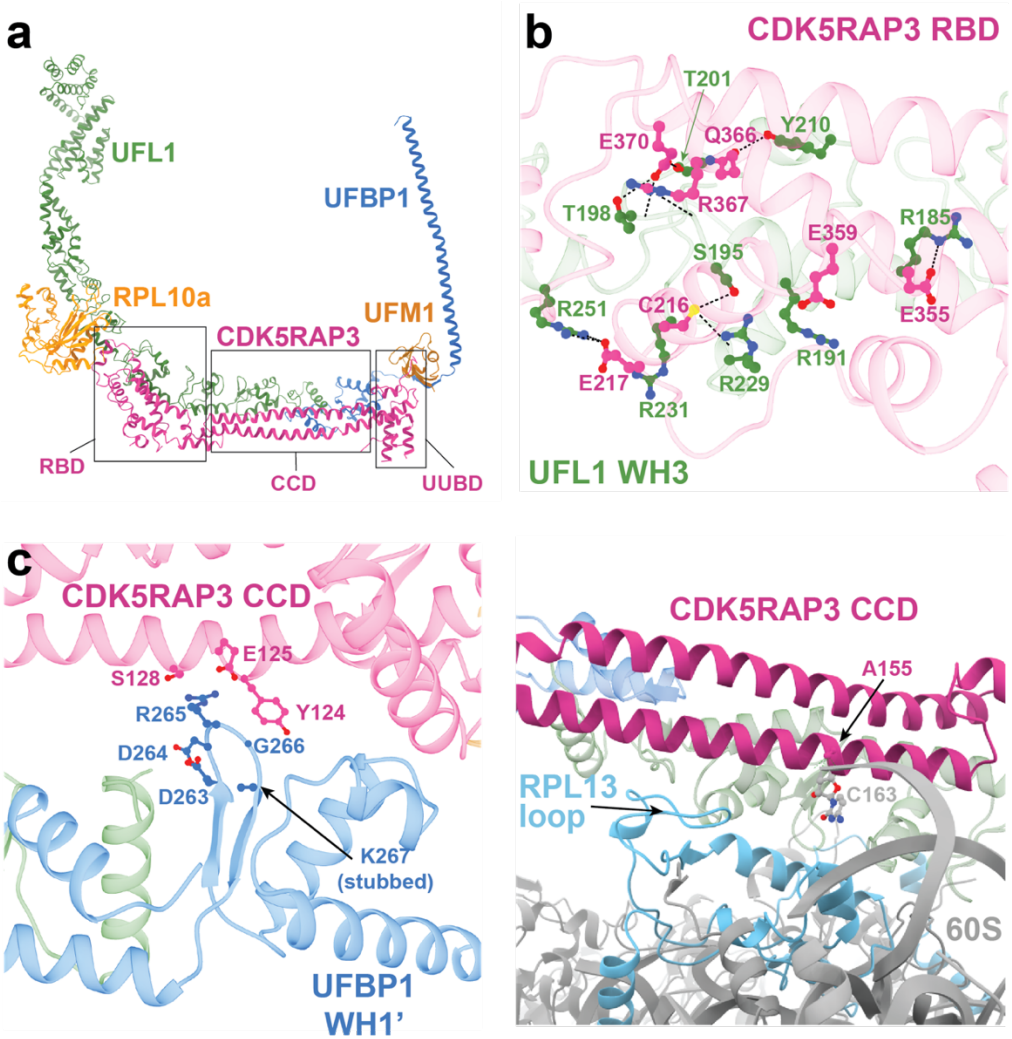


Figure 6.8 – CDK5RAP3 domain interactions with the UFL1 ligase complex and 60S ribosome
a Structural model of the UFL1 ligase complex and ribosome protein RPL10a, highlighting three protein domains of CDK5RAP3. RBD is ribosome binding domain. CCD is coiled-coil domain. UUBD is UFBP1 and UFM1 binding domain.
b Interactions between the RBD of CDK5RAP3 and UFL1 winged helix (WH) domain 3. Interacting residues shown as ball and stick. Hydrogen bonds are shown as black dashed lines.
c Contacts between arginine-265 of the UFBP1 DDRGK motif and the CCD of CDK5RAP3. Side chains of the DDRGK motif are also displayed.
d Contacts between the CCD of CDK5RAP3 and ribosome 28S rRNA helix 16 and RPL13. Contacting residues are shown as ball and stick. No contacts were measured between CDK5RAP3 CCD and RPL13 due to stubbing of side chains. RPL13 loop highlighted as likely region of contact with CDK5RAP3.

All three domains of CDK5RAP3 make continuous contacts with one or more ligase subunits, which is consistent with the idea that CDK5RAP3 is an integral component of the E3 ligase complex (**Fig. 6.8 a**). The RBD

of CDK5RAP3 makes contacts with WH2, 3 and 4 of UFL1, primarily through electrostatic interactions (**Fig. 6.8 a, b**). The CCD contacts WH2 of UFL1 and a loop in UFBP1 WH1' that contains the conserved 'DDRGK' motif (**Fig. 6.8 a, c**). The K-267 of the DDRGK motif has been suggested to be a target of UFMylation [130, 210]. With the addition of CDK5RAP3 to the UFL1 ligase complex, this DDRGK motif is sequestered by the CCD, likely preventing UFMylation of UFBP1 at K-267 and instead directing UFMylation towards the ribosome (**Fig. 6.8 c**). Further, UUBD of CDK5RAP3 interacts with UFBP1 WH1' and UFM1, although the resolution in this region is too low to see clear side chain contacts (**Fig. 6.8 a)**. In addition to contacting RPL10a, CDK5RAP3 also makes contacts with RPL13 and rRNA helix 16 via its CCD (**Fig. 6.8 d**). These extensive interactions of CDK5RAP3 with the ribosome and other ligase components highlight a functionally critical role for CDK5RAP3 in ribosome UFMylation.

6.4.3 - UFL1 ligase complex interactions with the ribosome

One of the major binding interfaces between the ligase complex and the ribosome is with the ribosome protein RPL10a (also known as uL1). This ribosome protein is part of the L1 stalk, a highly dynamic region composed of RPL10a and helices 76-78 of 28S rRNA (**Fig. 6.9 a**). The L1 stalk has a crucial role during protein translation and is involved in the rotation of the 40S subunit, translocation of tRNA across the P- and E-tRNA binding sites and removal of E-tRNA from the ribosome [211-213]. Upon UFL1 ligase complex binding, the L1 stalk is locked in a more closed conformation, contacting both UFL1 and CDK5RAP3 across a buried surface area of ~1,075 $Å^2$ (**Fig. 6.9 a, b**). This interaction is predominantly mediated by WH4 and WH5 of UFL1 and the RBD of CDK5RAP3, with extensive hydrophobic and electrostatic interactions across the interface (**Fig. 6.9 b, c, d**).

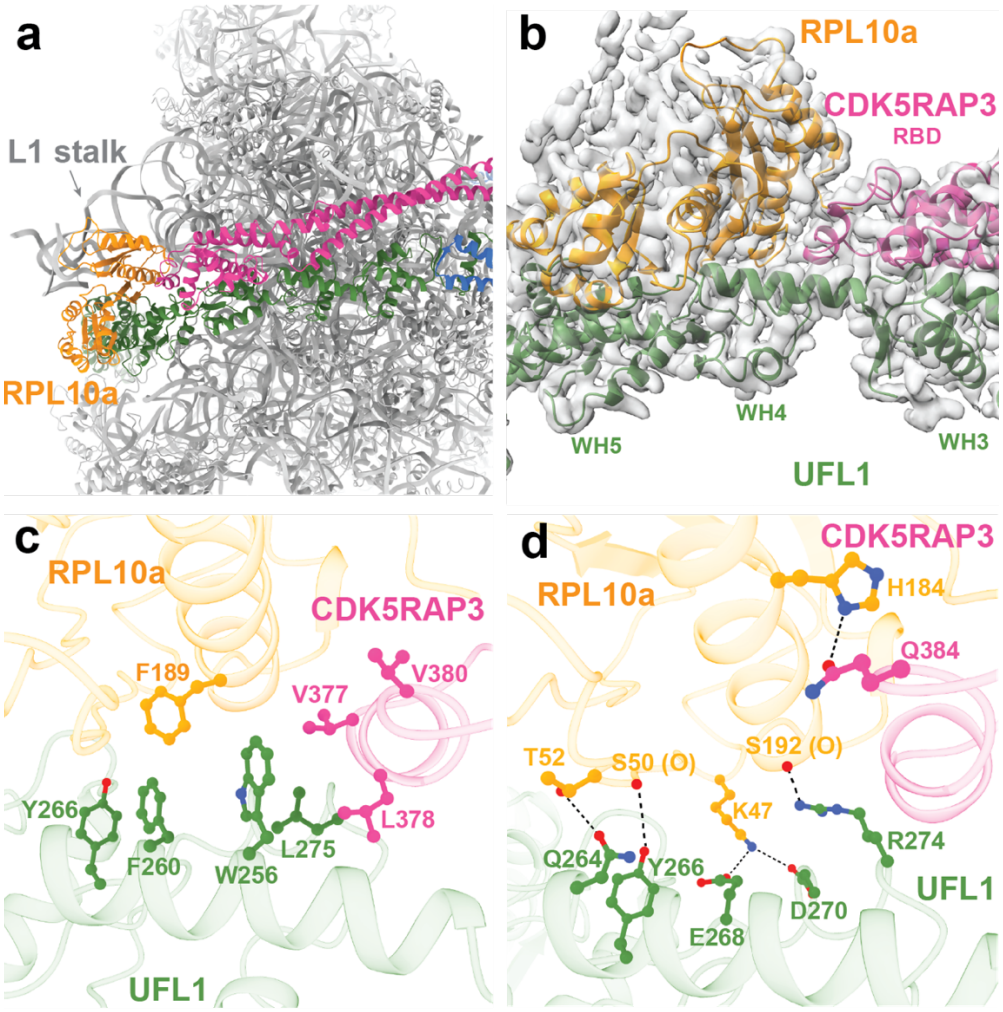


Figure 6.9 – UFL1 ligase complex interactions with ribosome protein RPL10a
a Cartoon model of the UFL1 ligase complex bound 60S ribosome cryo-EM structure, highlighting ribosome L1 stalk.
b UFL1 and CDK5RAP3 bind to RPL10a of the L1 stalk. Cartoon model is coloured by protein and cryo-EM density is shown in transparent grey. RBD is CDK5RAP3 ribosome binding domain. WH is winged helix domain.
c Main hydrophobic residues involved in RPL10a-ligase interface.
d Hydrogen bonding residues involved in RPL10a-ligase interface. Hydrogen bonds shown as black dashed lines.

UFL1 also interacts with other ribosome proteins, burying a total surface area of ~2,356 Å^2 (**Fig. 6.10 a**). The UFL1 WH domains almost exclusively interact with ribosomal proteins rather than rRNA, contacting RPL36a, RPL36, and RPL13 (**Fig 6.10 a, b**). Besides RPL10a, RPL13 seems to be important for ligase binding and has several hydrogen bonds with the N-terminal region of UFL1 (**Fig. 6.10 b**). Additionally, the UFL1 CTD contacts RPL11 and 28S rRNA helices 69 and 38a engage the

UFL1 CTD (**Fig. 6.10 c**). These extensive interactions between UFL1 and the 60S ribosome highlight a role for UFL1 as a substrate-recruiting ligase scaffold.

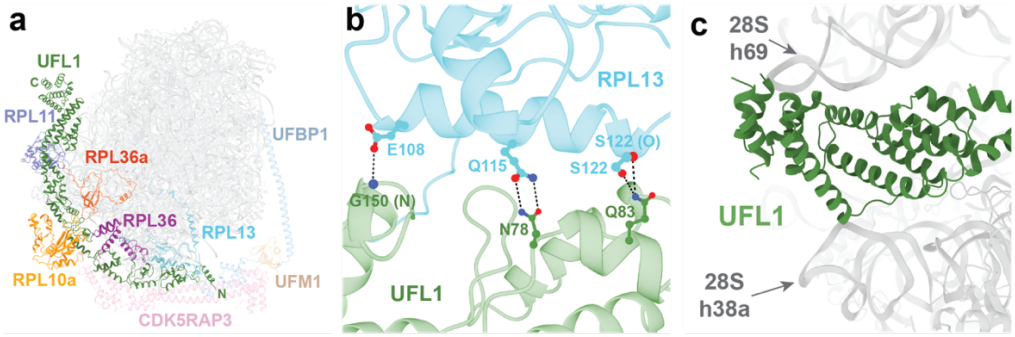


Figure 6.10 – UFL1 interactions with the 60S ribosome
a Atomic model of the UFL1 ligase complex bound to 60S ribosome with UFL1 interacting ribosome proteins highlighted. N-terminus (N) and C-terminus (C) of UFL1 annotated.
b View of UFL1 and RPL13 interactions. Interacting residues shown as ball and stick. Hydrogen bonds shown as dashed lines.
c C-terminal domain of UFL1 binds between 28S rRNA helices 38a and 69, occluding the tRNA binding sites.

6.4.4 - Catalytic region of the UFL1 ligase complex

The ribosomal protein RPL26 is UFMylated at K-134, located at its surface exposed C-terminus [137, 138]. The cryo-EM structure shows the ligase complex wraps around the ribosome in such a way that UFM1, located at the intersection of UFL1, UFBP1 and CDK5RAP3, orients its C-terminal G-83 towards the C-terminus of RPL26 (**Fig. 6.11**). Density for the RPL26 and UFM1 C-termini are not visible in the cryo-EM maps, likely due to their flexibility and poor resolution of this area. However, the missing residues would be long enough to bridge the ~25 Å gap from RPL26 to the ligase-bound UFM1, to allow for UFMylation of RPL26 K-134 (**Fig. 6.11**).

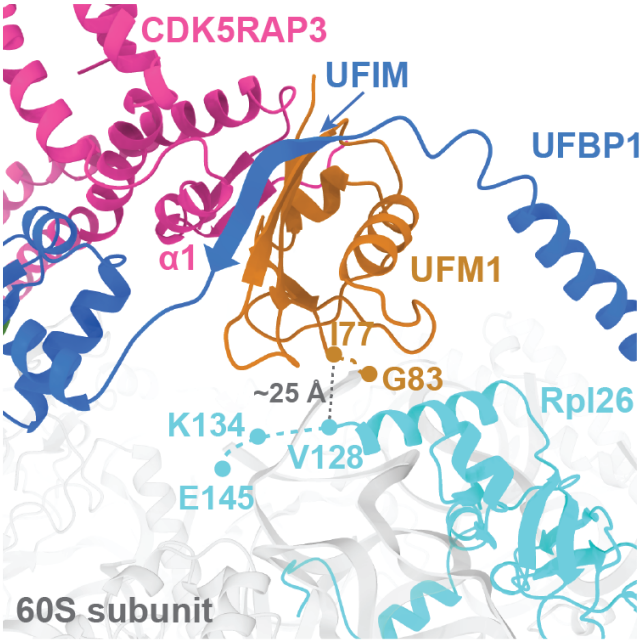


Figure 6.11 – Catalytic region of the UFL1 ligase complex
View of the catalytic region of the UFL1 ligase complex bound to the 60S ribosome. The UFBP1 UFM1-interacting motif (UFIM) and CDK5RAP3 N-terminus and alpha helix 1 (α1) are bound to UFM1. Distance between the C-termini of UFM1 and RPL26 in the cryo-EM model is shown as a black dashed line. Missing residues of UFM1 and RPL26 are shown as dashed lines in their respective colours. K134 is the RPL26 lysine-134 that is UFMylated.

A 10-residue segment (UFBP1$^{196-205}$) connecting WH1' and the long α-helix arm of UFBP1 binds to UFM1 (**Fig. 6.11**). The density in this region was not sufficient to build side chains into the model; therefore, ColabFold [204] multimer prediction was used to aid model building. This showed this UFBP1 linker forming a β-strand when bound to β2 of UFM1, called the UFM1-interacting motif (UFIM) (**Fig. 6.11**). Additional interactions with UFM1 are mediated by helix α1 and the N-terminus of CDK5RAP3 (**Fig. 6.11**). However, the local resolution of this region was ~4-5 Å, which lacked discernible side chain density to unambiguously locate interacting residues.

Surprisingly, there is no density for the E2 enzyme, UFC1, despite UFC1 being linked to UFM1 through an isopeptide linkage. This lack of UFC1 density could be due to flexibility of UFC1, or could indicate that the cryo-EM structure represents free UFM1, either pre- or post-ribosome

UFMylation. Superposition of a UFC1-UFM1 crystal structure (solved by Dr Helge Magnussen, Prof Yogesh Kulathu lab) onto the cryo-EM structure shows that the ligase is positioned in such a way to allow room for UFC1 binding proximal to RPL26 (**Fig. 6.12 a**).

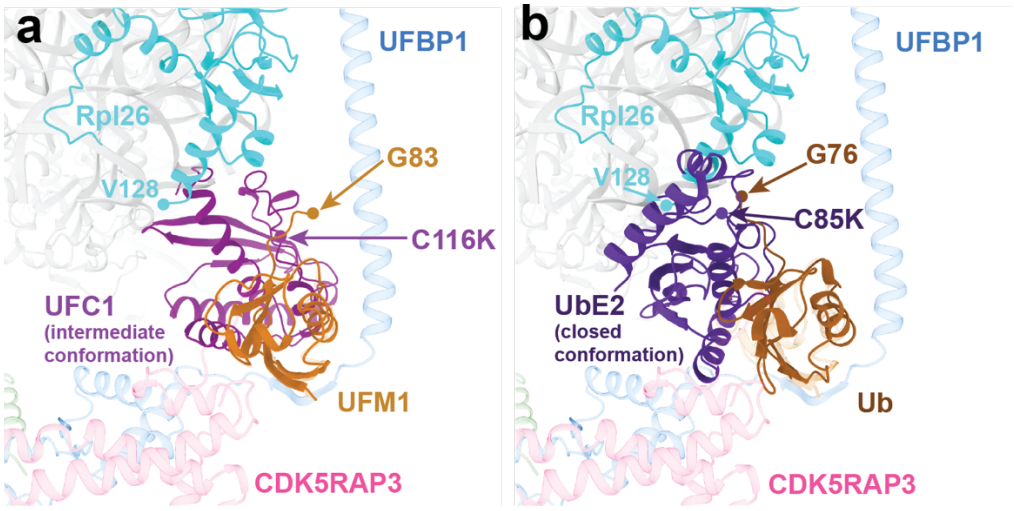


Figure 6.12 – Modelling in the UFC1 E2 enzyme onto the cryo-EM structure of UFL1 ligase complex-bound 60S ribosome
a Crystal structure of a UFC1-UFM1 conjugate (intermediate conformation) (solved by Dr Helge Magnussen, Prof Yogesh Kulathu lab) superposed onto the cryo-EM structure of the UFL1 ligase complex-bound 60S ribosome. C-terminal glycine of UFM1 (G83), catalytic cysteine to lysine mutation of UFC1 (C116K) and most C-terminal residue of RPL26 for which density is present in the cryo-EM map (V128) are depicted as circles.
b Crystal structure of a ubiquitin-E2 conjugate (PDB ID 7R71) [194] in a closed conformation superposed onto the cryo-EM structure of the UFL1 ligase complex-bound 60S ribosome, in the same view as **a**. C-terminal glycine of ubiquitin (G76), catalytic cysteine to lysine mutation of ubiquitin E2 (UbE2) (C85K) and most C-terminal residue of RPL26 for which density is present in the cryo-EM map (V128) are depicted as circles.

Compared to ubiquitin-E2 structures, the UFC1-UFM1 crystal structure shows UFC1 to be positioned in an intermediate conformation, between an open and closed conformation (**Fig. 6.12 a**). Comparison with a closed conformation ubiquitin-E2 structure [194] superposed onto the cryo-EM structure shows that a more closed conformation would bring the E2 catalytic cysteine and ubiquitin C-terminal glycine in closer proximity to the C-terminus of RPL26, which contains K-134 (**Fig. 6.12 b**).

6.5 – Cryo-EM of the UFMylated ribosome

To investigate what the structure of the post-UFMylated ribosome looks like, a UFMylation reaction was allowed proceed by incubating 60S ribosomes with the UFL1 E3 ligase complex, UFM1, ATP, E1 enzyme (UBA5) and E2 enzyme (UFC1) (performed by Dr Joshua Peter, Prof Yogesh Kulathu lab). This reaction was then imaged using cryo-EM. The resulting cryo-EM reconstruction shows density for the ligase complex bound to the ribosome, with UFM1 positioned similarly in comparison to the high-resolution dataset where UFM1 was isopeptide linked to UFC1 (**Fig. 6.13**). It is therefore possible that the high-resolution pre-UFMylation cryo-EM structure may actually mimic the post-UFMylated state, where UFM1 is then stably bound by UFBP1 and CDK5RAP3.

Although, it is possible that in the case UFMylated ribosomes where the UFL1 ligase complex is not bound, it would be difficult to see the relatively small UFM1 covalently attached to the ribosome, particularly as the linkage is quite flexible. In addition, the dataset is relatively small and so rarer states, such as a potential non-ligase bound UFMylated ribosome, may be difficult to detect. Regardless, the UFMylated ribosome dataset highlights a state where the ligase complex remains bound to the 60S ribosome, even after UFMylation of the ribosome has occurred (**Fig. 6.13**). This raises the possibility that the ligase is a “reader” as well as a “writer” of UFM1 modification, and that the longer E3 ligase complex residual time on the ribosome allows for UFBP1-mediated dissociation of the SEC61 translocon (**Fig. 6.5 d**).

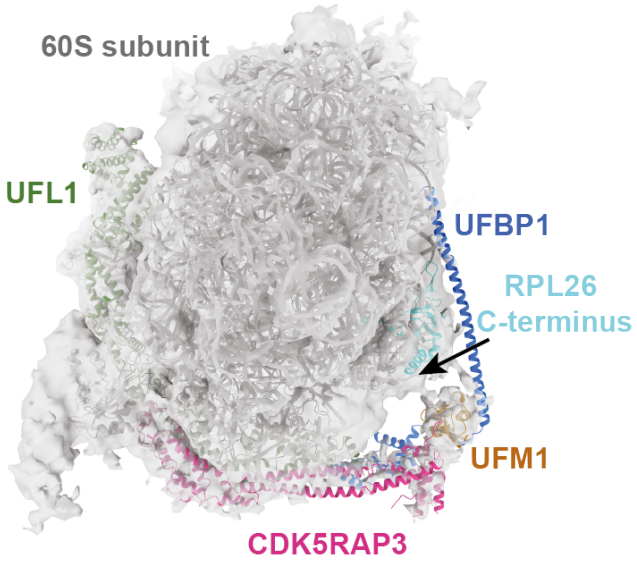


Figure 6.13 – Cryo-EM density map of the UFMylated ribosome
Cryo-EM density map of UFMylated 60S ribosome (transparent grey). Structure of UFL1 ligase complex-bound 60S ribosome is rigid body fitted into the cryo-EM map. Black arrow points to C-terminal region of RPL26, which contains lysine-134, the site for UFMylation.

6.5.1 – The UFL1 ligase complex dissociates the 60S ribosome from the SEC61 translocon

Binding of the UFL1 ligase complex to the 60S would require dissociation of the 60S from the SEC61 translocon at the ER membrane, leaving the ligase-60S complex localised to the ER membrane via the TMH of UFBP1 (**Fig. 6.14**). In addition to clashes with the SEC61 translocon, the catalytic region of the ligase complex where UFM1 binds would likely deform the ER membrane (**Fig. 6.14**). With the revelation that the ligase complex remains bound to the UFMylated ribosome, it was speculated that ribosome UFMylation may trigger the dissociation of 60S from SEC61. To explore this, the ribosome UFMylation dependency on SEC61 binding was explored (performed by Mr Rohan Thakur, Prof Yogesh Kulathu lab).

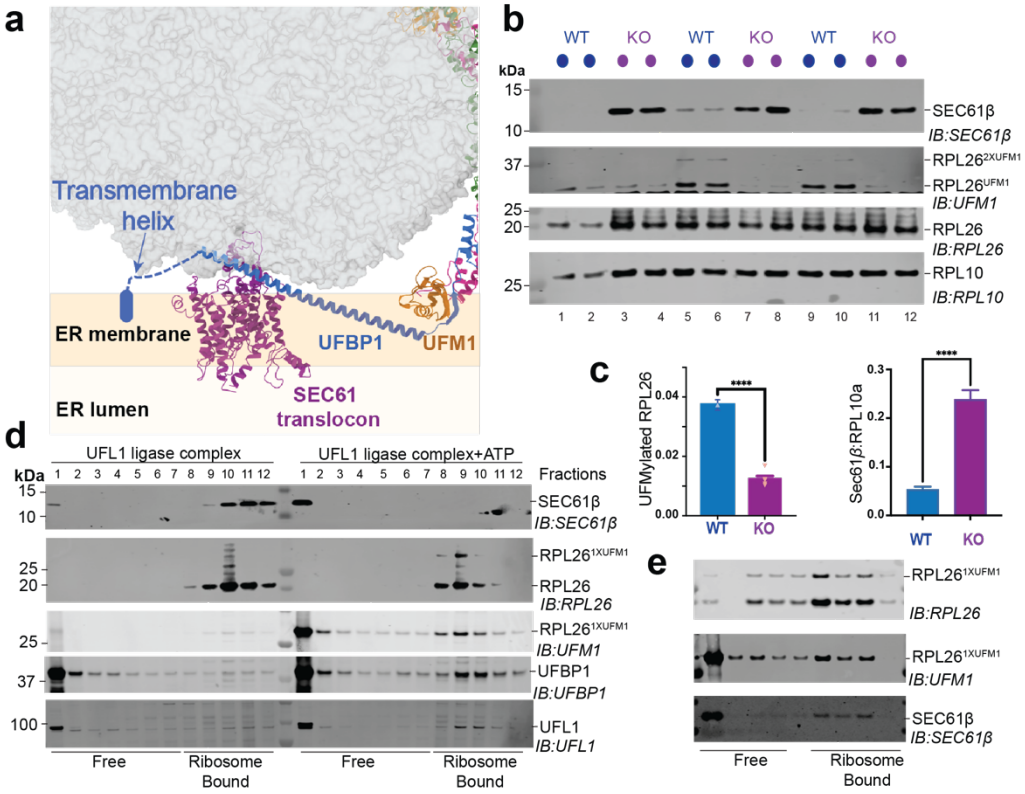


Figure 6.14 – UFMylation of the 60S helps to dissociate the 60S ribosome from the SEC61 translocon

Data in **b**-**e** were performed by Mr Rohan Thakur, Prof Yogesh Kulathu lab.

a Ligase binding to the SEC61 translocon-bound ribosome may destabilize the 60S-SEC61 interaction. UFL1 ligase complex structure superimposed onto the structure of the SEC61-80S ribosome (PBD ID 6r7q) [209]. Approximate position of the ER membrane and ER lumen relative to the SEC61 translocon are shown.

b Loss of UFMylation leads to accumulation of 60S ribosomes with the SEC61 translocon. SEC61β association with 60S ribosomes was analysed in membrane fractions from parental WT and CDK5RAP3 KO cells. Lysates were normalised to uniform RNA concentration (using absorbance at 254nm) and fractionated on 10-30% sucrose density gradients. Fractions corresponding to 60S from 3 independent replicates were analysed by immunoblotting with indicated antibodies.

c Quantification of immunoblots from **b**. UFMylated RPL26 band intensity normalized to intensity of total RPL26 (left) and SEC61β band intensities normalized to RPL10a. Data show mean ± SD. ***p< 0.0001 (Student t-test). Data are representative of n=2 independent experiments.

d UFMylation mediates dissociation of the 60S ribosome from the SEC61 translocon. *In vitro* UFMylation reactions were performed on membrane associated 60S-SEC61 translocon complexes isolated from CDK5RAP3 KO cells. 60S-SEC61 complexes were incubated with UBA5, UFC1, UFL1 ligase complex and UFM1, either in the presence or absence of ATP. The reaction products were separated on a sucrose gradient and analysed by immunoblotting with the indicated antibodies. Blot is representative of n=2 independent experiments

e Dissociation of 60S from the SEC61 translocon requires the UFIM motif. *In vitro* 60S-SEC61 UFMylation and translocon dissociation assay performed as in **d,** using UFL1 ligase complex with the UFBP1 UFIM mutant F196A.

Indeed, in CDK5RAP3 knockout cells, where ribosome UFMylation is inhibited, there is an increase in ribosome bound to SEC61 compared to wild-type cells (**Fig. 6.14 b, c**). In addition, *in vitro* UFMylation reactions performed on isolated ER membrane fractions containing SEC61-bound ribosomes also displayed an increased dissociation from SEC61 when UFMylation was allowed to proceed (**Fig. 6.14 d**). Lastly, a F196A mutation of the UFBP1 UFIM motif, which prevents UFM1 binding, showed a decrease in SEC61 dissociation (**Fig. 6.14 e**). These data combined suggest a key role for ribosome UFMylation in the dissociation from the SEC61 translocon. It is possible that RPL26 UFMylation and stable binding of UFM1 post-UFMylation causes the UFBP1 helical arm to act as a lever, pushing the ribosome from SEC61, allowing the ligase complex to bind in the conformation seen in the cryo-EM structure.

6.6 - UFL1 loop projects into the ribosome peptidyl transferase centre

Intriguingly, the cryo-EM map showed an additional density present in the PTC which could not be attributed to the ribosome. Importantly, this additional density was not present in the non-ligase bound 60S map (**Fig. 6.15 a**). Aside from the additional density in the PTC, there is no additional density observed for nascent polypeptide within the peptide exit tunnel or any major conformational changes to the peptide exit tunnel, including tunnel proteins RPL4 (uL4), RPL17 (uL22), and RPL39 (eL39). The well resolved density allowed the building of a 13-residue segment of UFL1 ($UFL1^{436-448}$) that is part of an 86-residue UFL1 protrusion from the UFL1 CCD, hereby referred to as the PTC loop (**Fig. 6.15 b, c**).

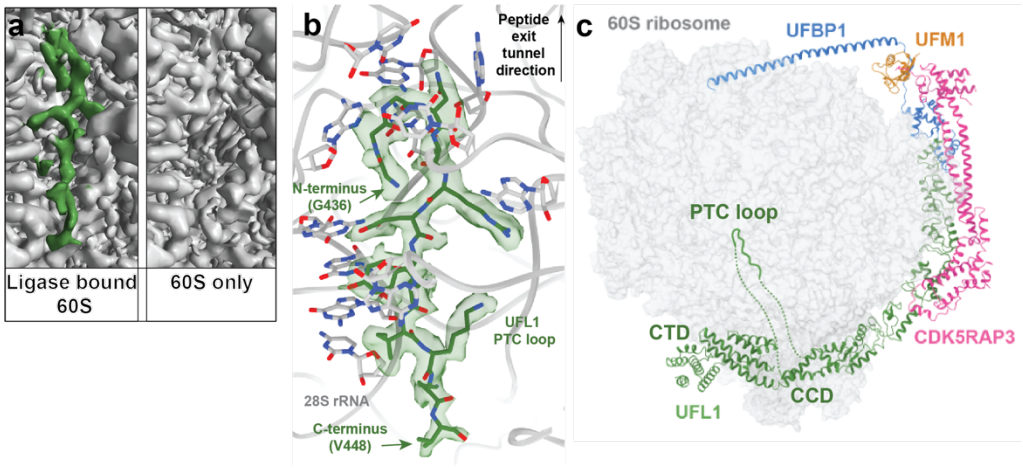


Figure 6.15 – UFL1 PTC loop density and atomic model
a Comparison of cryo-EM density at the ribosome PTC between the UFL1 ligase complex-bound 60S and non-ligase bound 60S maps. Density corresponding to the UFL1 PTC loop is coloured in green.
b Atomic model of the UFL1 PTC loop bound to the 60S ribosome PTC. Main chain and side chains of UFL1 shown as sticks. Density for UFL1 loop shown in transparent green. UFL1 PTC loop contains residues ^{436}GGGNAREYKIKKV^{448}. Contacting rRNA bases are shown as sticks.
c UFL1 ligase complex bound 60S ribosome with ligase complex shown as a cartoon representation and 60S ribosome shown as transparent grey surface representation. CCD is coiled-coil domain. CTD is C-terminal domain. Dashed line represents part of the UFL1 PTC loop for which density is missing.

This PTC loop is positioned proximal to the P-tRNA binding site and remarkably, appears to remodel key translation elongation and termination bases. 28S rRNA A4548 rotates towards the P-site to stack with the aromatic ring of UFL1 Y-443 (**Fig. 6.16 a**). Additionally, 28S rRNA U4452 flips ~90° to be in proximity to UFL1 G-437, which partially occludes the aminoacyl (A)-site (**Fig. 6.16 b**). Importantly, A4548 is a highly mobile base within the PTC which aids in tRNA positioning for peptide bond formation and peptide release, while U4452 is involved with A-tRNA binding [153-156]. The peptide bond formation catalytic base A3908, as well as A4385, stacks with UFL1 D-439 and R-441 respectively (**Fig. 6.16 c, d**). In addition, the positively charged R-441 of the UFL1 PTC loop sits in a pocket surrounded by the phosphate backbone of several 28S rRNA bases, with several hydrogen bonds formed (**Fig. 6.16 d**).

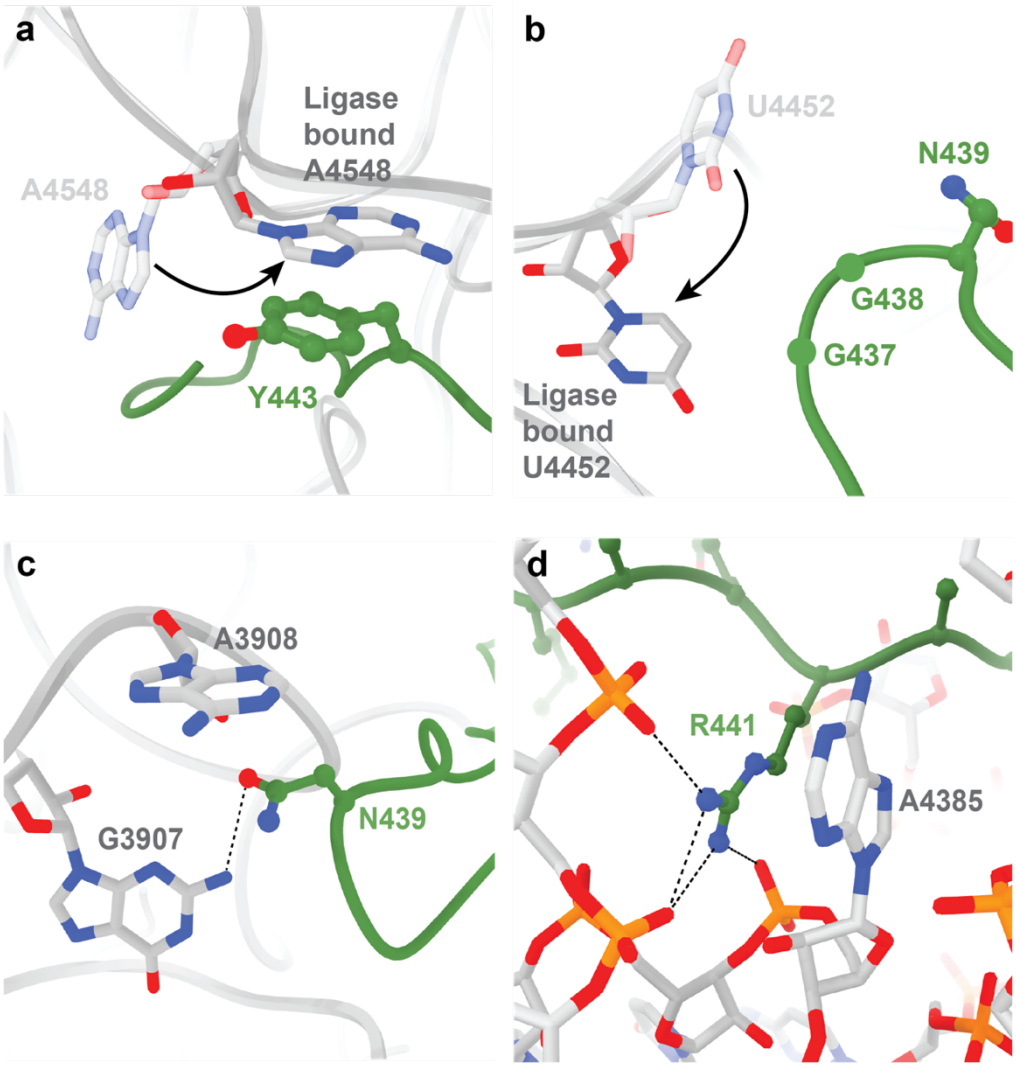


Figure 6.16 – Interactions between the UFL1 PTC loop and surrounding 28S rRNA bases
UFL1 PTC loop shown in green. UFL1 residues shown as ball and stick. 28S rRNA of the 60S ribosome is shown in grey. 60S bases shown as sticks. Hydrogen bonds shown as black dashed line.
a 28S RNA adenine-4548 moves towards the peptidyl (P)-site to stack with UFL1 tyrosine-443. Transparent grey 60S model represents non-ligase bound 60S (PDB ID 6r7q) [209]. Opaque grey 60S model is UFL1 ligase complex-bound 60S. Black arrow shows the direction of movement.
b 28S RNA uracyl-4452 moves towards the aminoacyl (A)-site to sit proximal to glycine-437. Transparent grey 60S model represents non-ligase bound 60S (PDB ID 6r7q) [209]. Opaque grey 60S model is UFL1 ligase complex-bound 60S. Black arrow shows the direction of movement.
c UFL1 asparagine-439 stacks with 28S rRNA adenine-3908 and hydrogen bonds with guanine-3807.
d UFL1 arginine-441 stacks with 28S rRNA adenine-4385 and hydrogen bonds with surrounding phosphates of the 28S rRNA backbone.

6.6.1 - Potential function of the UFL1 PTC loop

This binding to and re-arrangement of the PTC bases resembles the action of protein translation stalling agents, such as the translation of arrest peptides or binding of antibiotics such as anisomycin (**Fig. 6.17**) [214]. However, binding of the PTC loop in order to induce ribosome stalling does not make much sense in the context of biological function, as the UFL1 ligase complex is implicated in rescuing stalled ribosomes rather than causing the ribosomes to stall. Alternatively, the UFL1 PTC loop may bind to the PTC after a ribosome stalling event in order the keep the ribosome in an inactive state while the downstream effects of ribosome UFMylation, ribosome quality control and/or SEC61 release, takes place. This however contrasts with binding of the UFL1 CTD, which sterically blocks the tRNA binding sites and 40S subunit, effectively preventing protein translation from re-initiating (**Fig. 6.5 b, c**). Instead, the UFL1 PTC loop may be a detection mechanism for translationally terminated or stalled ribosomes, acting as a signal to recruit the UFL1 ligase complex. Regardless, as these remodelled PTC bases are critical for translational elongation and termination, the intricate interactions and remodelling of the PTC suggest an important role for the PTC loop and may even suggest UFMylation independent functions for UFL1.

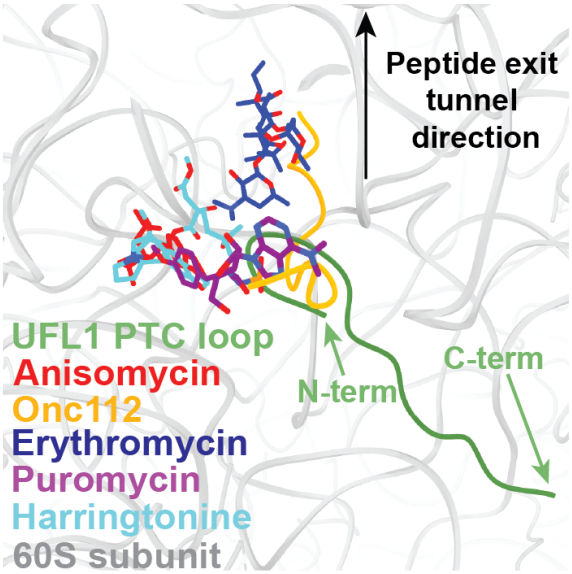


Figure 6.17 – Comparison of the UFL1 PTC loop bound to the ribosome with ribosome stalling agents

View of the ribosome PTC, highlighting that the UFL1 PTC loop and ribosome stalling agents bind to similar regions. UFL1 and Onc112 are polypeptides. Anisomycin, Erythromycin, Puromycin and Harringtonine are small molecules. Anisomycin PDB ID 4u3m [215], Onc112 PDB ID 4zer [216], Erythromycin PDB ID 4wfn [217], Puromycin PDB ID 1q7y [218] and Harringtonine PDB ID 7ucj [219].

The binding of the PTC loop also resembles the binding mode of the ribosome release factor, eukaryotic release factor 1 (eRF1). During translation termination, eRF1 catalyses the release of nascent polypeptide from ribosome bound P-tRNA in response to sensing a stop codon in the mRNA decoding centre [220]. Both the eRF1 catalytic domain and the UFL1 PTC loop bind to the PTC and share similar residues (**Fig. 6.18 a**). The apex of the eRF1 catalytic domain contains the catalytic motif, GGQ, whereas the PTC loop apex contains residues ^{437}GGN^{439} (**Fig. 6.18 a**). During eRF1-mediated tRNA-peptide hydrolysis, the glutamine in the GGQ motif is thought to coordinate a water molecule to mediate nucleophilic attack of the peptide-tRNA ester bond [165]. Since asparagine is structurally similar to glutamine, where both amino acids have an amide group in their side chains, it is possible that the asparagine in the UFL1 GGN motif may also perform a similar catalytic role. Superimposition of the PTC loop with structures of P-tRNA-peptide ribosomes shows that the UFL1 GGN motif coincides approximately with the tRNA-peptide junction (**Fig. 6.18 b**). However, the UFL1 PTC loop differs from the eRF1 central domain in its positioning within the PTC. eRF1 occupies the A-site of the PTC to catalyse hydrolysis of the adjacent P-site tRNA-peptide bond, whereas the UFL1 PTC loop occupies the P-site (**Fig. 6.18 a**). Therefore, if the UFL1 PTC loop were performing a similar catalytic role to eRF1 then it would likely act upon tRNA-peptide occupying the A-site, perhaps after a failure in translocation of tRNA-peptide from the A-site to the P-site during a ribosome stalling event. This role for the PTC loop acting as a release factor is consistent with the UFMylation role in releasing stalled ribosomes from SEC61, whereby the nascent polypeptide can be released from the ribosome and then be accessed by cytosolic factors to result in the degradation of the polypeptide.

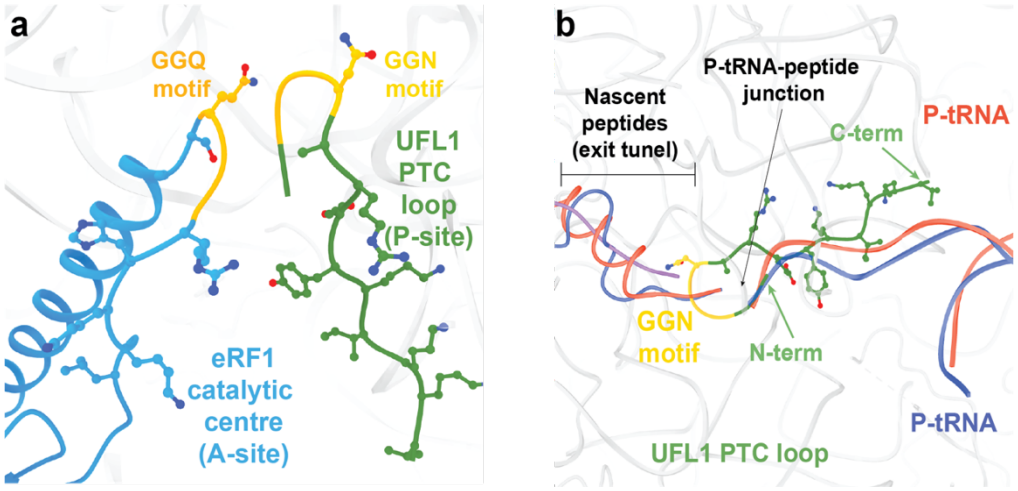


Figure 6.18 – Comparison of UFL1 PTC loop with the eRF1 central domain
a Comparison of the UFL1 PTC loop (P-site) with eRF1 catalytic centre (A-site; PDB ID 6ip8 [208]). eRF1 catalytic residues GGQ and UFL1 loop residues GGN coloured in yellow.
b Comparison of UFL1 PTC loop positioning to structures of stalled ribosomes with nascent peptides in exit tunnel and tRNA in P-site. UFL1 loop shown in green, PDB ID 3j92 [221] in blue, PDB ID 7qwr [193] in purple, PDB ID 6xa1 [222] in red. N-terminus and C-terminus of loop indicated with green arrows. UFL1 residues GGN coloured in yellow, sits approximately at the P-tRNA-peptide junction.

eRF1 senses all three stop codons in the mRNA decoding centre by recognizing the geometry of the UAA, UAG and UGA stop codons during translation termination, leading to polypeptide release. It is less well known how polypeptides are released in the case of a ribosome stall, such as a readthrough of the mRNA poly-A tail (AAA codon). Additionally, it has been reported that the translation of a hCMV stall peptide re-arranges the PTC so as to inhibit eRF1 activity [223]. This perhaps explains the need for an additional release factor other than eRF1, as different geometry in the mRNA decoding centre (e.g., poly-A tail) and a re-arranged PTC (induced translation stalling) may prevent the release factor activity of eRF1.

A key question is whether UFL1 PTC loop binding is coupled to ribosome UFMylation, which occurs towards the other end of the ligase complex. Further experimentation is needed to clarify exactly what the role of the UFL1 PTC loop is and how exactly the UFMylation of the 60S ribosome

by the UFL1 ligase complex promotes the ribosome quality control pathway in cells.

6.7 - Chapter summary

For the first time, the high-resolution cryo-EM structure of the UFL1 E3 ligase complex-bound 60S ribosome has been revealed. The structure shows the ligase complex wrapped around the ribosome in a unique substrate binding mode not seen in other E3 ligases. The structure not only reveals the ligase complex assembly, but also shows how the ligase complex binds to the ribosome, with UFM1 positioned close to its UFMylation target, RPL26. The structure solidifies CDK5RAP3 as an integral part of the UFL1 ligase complex and highlights its role as a ribosome substrate adaptor. Intriguingly, a UFL1 loop that binds to and re-arranges the PTC is identified, in a function that remains to be investigated. Collectively, these results show a fundamental role for the UFL1 ligase complex in releasing 60S ribosomes from the SEC61 translocon and elucidate the molecular basis for UFMylation in regulating protein quality control at the ER.

Chapter 7 – Discussion

The post-translational modification of substrates with ubiquitin or UFM1 controls many cellular processes, which are tightly regulated by the action of specific E3 ligase complexes. E3 ligases can assemble into a variety of structures, with some ligases functioning as multi-subunit protein complexes, as well as forming various oligomeric states, for their enzymatic action. This diversity in E3 ligase assembly and structure allows the targeting of a diverse substrate pool. While many structural studies of E3 ligases have revealed the molecular mechanisms of substrate specificity and ubiquitylation, there is still a lack of structural information on how multi-subunit high-order oligomer E3 ligases, such as the SPOP-Cullin-3 ligase complex, assemble. Additionally, there is no structural information for the single known UFM1 E3 ligase, the UFL1 E3 ligase complex. Gaining the structures of these E3 ligase complexes will help to reveal how they target substrates and the molecular mechanisms of substrate modification. In this chapter the results obtained for this thesis are summarised and discussed, highlighting why this work was necessary, how this work benefits the field and what should be studied next.

7.1 – Characterising SPOP binding to MyD88

At the beginning of this project, an interaction between the ubiquitin E3 ligase substrate binding adaptor, SPOP, and the integral Myddosome complex protein, MyD88, had been identified [62, 118, 119]. SPOP was shown to negatively regulate immune signalling by ubiquitylating MyD88, although various reports had come to different conclusions as to how this was achieved. Using a combination of biochemical assays and structural biology, the interaction between SPOP and MyD88 was studied in **Chapter 3**.

7.1.1 - MyD88 binds to SPOP using an extended SPOP-binding consensus motif

In **Chapter 3**, it was shown that a MyD88 peptide containing the minimal SBC (A**VDSSV**P) bound to SPOP^{MATH} with a K_d of 32.25 ± 2.08 μM (**Fig. 3.2 e**, **f**). This binding is of moderate strength when compared to other substrate SBCs, which typically have binding affinities in the mid to high μM range. Analysis of the MyD88 protein sequence showed a region of high conservation just N-terminal to the SBC, which was hypothesised could be involved in SPOP binding (**Fig. 3.2 c**). Indeed, the binding affinity of a longer MyD88 peptide (AEKPLQVAA**VDSSV**PRT) was measured with a K_d of ~2-4 μM, making it the highest known binding affinity amongst SPOP substrates (**Fig. 3.2 e**, **f, Fig. 3.6 a, e**). The SPOP^{MATH}-MyD88 peptide X-ray crystal structure helped to corroborate the binding data (**Fig. 3.8**). This revealed an unusual extended binding mode of MyD88 which has only been shown in one other SPOP substrate, Pdx1 [63, 99]. From a biological context, the need for MyD88 to possess an extended SPOP-binding motif remains unknown. One possible reason is that these additional binding residues compensate for a non-standard SBC, such as MyD88 having a non-standard residue at SBC position five. Directly N-terminal and C-terminal to the MyD88 peptide used in the crystal structure are the Death and TIR domains, which both oligomerise to form part of the large multiprotein complex, the Myddosome (**Fig. 3.2 d**). Therefore, another possibility is that the binding of this extended SBC to SPOP orients MyD88 so as to avoid clashes of the adjacent domains. In addition, this positioning may orient MyD88 in order to allow optimal access to lysine residues for ubiquitylation by the Cullin-3 E3 ligase.

The crystal structure of the SPOP^{MATH}-MyD88 peptide complex revealed the molecular interactions of this extended SPOP-MyD88 binding. Interestingly, this revealed MyD88 residues ^{129}LQ^{130} sitting within a hydrophobic pocket of SPOP, which were further validated to be important for binding through FP peptide competition assays and SPOP

mutant co-IPs with MyD88 (data of Dr Mukul Mishra, Dr Luca Busino lab) (**Fig. 3.9, Fig. 3.10**). This region is also important for the Pdx1 extended binding motif and likely represents a new secondary binding interface to SPOP.

This characterisation of the MyD88 extended binding motif builds upon the original work by Zhuang *et al.*, which first identified the five-residue SPOP-binding consensus motif in SPOP substrates [55]. Upon the discovery of Pdx1 as an extended SPOP binder, which also has a non-standard SBC, a more relaxed SBC was proposed of φ-π-S-π-π (φ is non-polar amino acid, π is polar amino acid) [99]. MyD88 diverges slightly from this new SBC as it possesses a valine at position 5. With the identification of Pdx1 and MyD88 possessing extended SBCs, which also have unusual SBC residues at position 4 and/or 5, it is possible that additional substrates of SPOP can be identified using a new extended binding motif. Taking this into account, an extended SBC is proposed of λ-Q-X-X-X-φ-π-S-X-X (X is any amino acid, λ is large amino acid).

Analysis of the protein sequences of known SPOP binders identified six potential substrates which may bind to SPOP using an extended SBC (**Fig. 3.13 c**). Preliminary data has identified Caprin1 as a potential extended binder, with potentially more substrates yet to be discovered (**Fig. 3.14**). Overall, the identification of a MyD88 extended binding motif, the second known substrate to possess this, broadens our understanding on how SPOP recognises substrates.

There is potentially a second SBC on MyD88 at its N-terminus, with binding experiments showing $SPOP^{MATH}$ is able to bind to this sequence with a similar affinity as the intermediate domain SBC (**Fig. 3.4**). This is in contrast with a previous report which states that this SBC was dispensable for SPOP binding [119]. Whether this is a true SPOP binding site *in vivo* still needs to be explored. In the context of the Myddosome complex, several MyD88 molecules would present SBCs for SPOP to

bind to, potentially circumventing the need for multiple degrons in a protein sequence, as is seen for other substrates.

7.2 - Cryo-EM structure determination of the SPOP oligomer

SPOP undergoes homo-oligomerisation to form high-order helical oligomers. This oligomerisation of SPOP was shown to be important for the efficient ubiquitylation of target substrates, likely by recruiting multiple Cullin-3 ligase complexes to the substrate. The structural and biochemical analysis of the interaction between $SPOP^{MATH}$ and MyD88 peptides in **Chapter 3** provided insights into how these two proteins interact. However, a key question is how do these two proteins interact in the context of their full-length oligomeric structure, and how does this oligomeric structure help to enhance ubiquitylation? To try to answer this, SPOP constructs capable of forming high-order oligomers were produced for structural characterisation using cryo-EM.

7.2.1 - SPOP residues 18-28 are required for stable SPOP oligomer formation

At the beginning of this project, SPOP residues 28-359 had been used to study oligomeric SPOP, as the full-length protein was less soluble and tended to form nuclear speckles that were difficult to purify from cells for *in vitro* study. Analysis of the AlphaFold predicted SPOP monomer and dimer structures showed an additional β-sheet forming between two MATH domains, formed by residues that were not present in the current construct residue boundaries. Concurrent with this, in **chapter 4** it was shown that that $SPOP^{18-359}$ produced more stable oligomers than $SPOP^{28-359}$ (**Fig. 4.9**). Cryo-EM was used to try to determine the structure of the $SPOP^{18-359}$ oligomer, which produced a low-resolution cryo-EM density map that fit the predicted SPOP oligomer model well (**Fig. 4.10 e**). Some E3 ligase substrate adaptors have the ability to undergo

dimerisation or tetramerization, which can either activate or inhibit its E3 ligase activity [224-226]. SPOP is an unusual E3 ligase adaptor which has the ability to form seemingly indefinite high-order oligomers. This study provides structural confirmation that SPOP achieves this high-order oligomerisation through alternating dimerization through its BTB domain dimers and CTD domain dimers.

7.2.2 - Comparison of this study's SPOP oligomer structure with the published SPOP oligomer structure

When optimising the SPOP$^{18-359}$ sample for obtaining better cryo-EM data, a study published the structure of the full-length SPOP oligomer [188]. Comparison of the two cryo-EM maps showed large differences in the helical parameters, with full-length SPOP having twelve dimers per turn and this study's SPOP$^{18-359}$ map having eight dimers per turn. 3DFlex motion modelling was used on this study's cryo-EM dataset of SPOP$^{18-359}$, which revealed significant motion of the SPOP oligomer (**Fig. 4.14**). The SPOP$^{18-359}$ oligomer is seen twisting and relaxing, a conformational change which would likely alter the number of dimers per turn and therefore, the shape of the SPOP oligomer. Currently, the implications of this large plasticity of the SPOP oligomer are unknown. It is possible that SPOP can adopt a range of conformations in order to target a varied pool of substrates, from small monomeric proteins to large oligomeric complexes. Perhaps this motion may also be coupled to the ubiquitylation of substrate where the SPOP oligomer relaxes or contracts to bring the substrate closer to ubiquitin-E2 positioned on a Cullin-3 arm. The exact significance of this flexibility and the molecular mechanism of how this is achieved remains to be studied. High-resolution cryo-EM data would help shed light on the molecular basis of how this flexibility within the SPOP oligomer is achieved. Additionally, the use of new cryo-EM image processing tools such as 3DFlex would likely be useful for studying large flexible oligomeric E3 ligase complexes, such as the SPOP-Cullin-3-Rbx1 ligase complex.

7.3 - Cryo-EM structure determination of the SPOP-MyD88 high-order oligomer complex

There are conflicting reports as to whether SPOP-mediated ubiquitylation of MyD88 even occurs and whether this ubiquitylation occurs at the level of the fully assembled Myddosome, activated MyD88 oligomers at the TLR prior to binding of IRAKs, or at free MyD88 protomers. In **Chapter 3** it was confirmed through binding data and an X-ray crystal structure that this interaction indeed occurs; although, these analyses were performed using minimal domains and peptides, rather than full-length proteins. Since both SPOP and MyD88 form high-order oligomers, a key question was how do these large oligomers come together to form a complex, if at all? To try to answer this, a structural biology approach was used to image the SPOP-MyD88 high-order oligomer complex.

In **Chapter 5**, an attempt was made to produce recombinant full-length MyD88 which proved to be mostly insoluble when expressed in *E. coli* cells (**Fig. 5.2**). The decision was made to instead produce a truncation construct of MyD88, $MyD88^{DD\text{-}ID}$, which was still capable of oligomerisation and importantly contained both the validated and predicted SBCs. It was also confirmed that $MyD88^{DD\text{-}ID}$ is able to form large oligomers in a concentration-dependent manner, as had been reported previously (**Fig. 5.6 b**, **5.8 a**) [205]. $MyD88^{DD\text{-}ID}$ formed a complex with $SPOP^{18\text{-}359}$, which was then imaged using cryo-EM (**Fig. 5.9**). Processing of the cryo-EM data showed the $SPOP^{18\text{-}359}$ oligomer wrapping around a $MyD88^{DD\text{-}ID}$ oligomer, with a low-resolution map suggesting that the MATH domains of SPOP are oriented towards MyD88, as expected (**Fig. 5.10**). This provides a first look at how oligomeric SPOP engages substrate, which has previously only been demonstrated by structures of the minimal SPOP MATH domain and short substrate peptides.

A key question that remains is how oligomeric MyD88 assembles in the Myddosome. As individual domains, both the DD and TIR domains of MyD88 are able to oligomerise into long helical filaments (**Fig. 5.8 b**) [205, 206]. Further structural studies of the full-length MyD88 protein are needed to reveal how the DD oligomerisation domains and TIR oligomerisation domains arrange themselves in the context of the full-length MyD88 oligomer. It is likely that when incorporated into the Myddosome, MyD88 does not form long filamentous oligomers. Instead, the MyD88 DDs could be capped by the IRAKs and the MyD88 TIR domains could be capped by binding to TLRs, limiting the number of MyD88 molecules in the Myddosome. Indeed, a report showed that addition of IRAK4^{DD} caused MyD88^{DD} filaments to instead form 4:6 stoichiometry Myddosomes [205]. A structure of the MyD88-IRAK4-IRAK2 DD complex and live cell-imaging estimation revealed a Myddosome stoichiometry of 6:4:4, with other reports measuring MyD88-IRAK4 stoichiometries of 7:4 and 8:4 [108, 227, 228]. With this stoichiometry, six MyD88 molecules within the Myddosome could potentially present 6-12 SBCs, allowing for multivalent SPOP binding and efficient MyD88 ubiquitylation. Another open question is does the SPOP oligomer ubiquitylate MyD88 in the Myddosome (6-12 SBCs) or at free MyD88 monomers (1-2 SBCs) (**Fig. 7.1**)? SPOP substrates typically present two or more SBCs for SPOP binding. From an evolutionary perspective, the targeting of MyD88 monomers containing a single SBC to high-order SPOP oligomers, which have the capability of multivalent binding to substrates, seems less likely (**Fig. 7.1**). Currently, there is a lack of study on how SPOP interacts with oligomeric substrates. To our knowledge, this study provides the first example of the SPOP oligomer engaging an oligomeric substrate, MyD88.

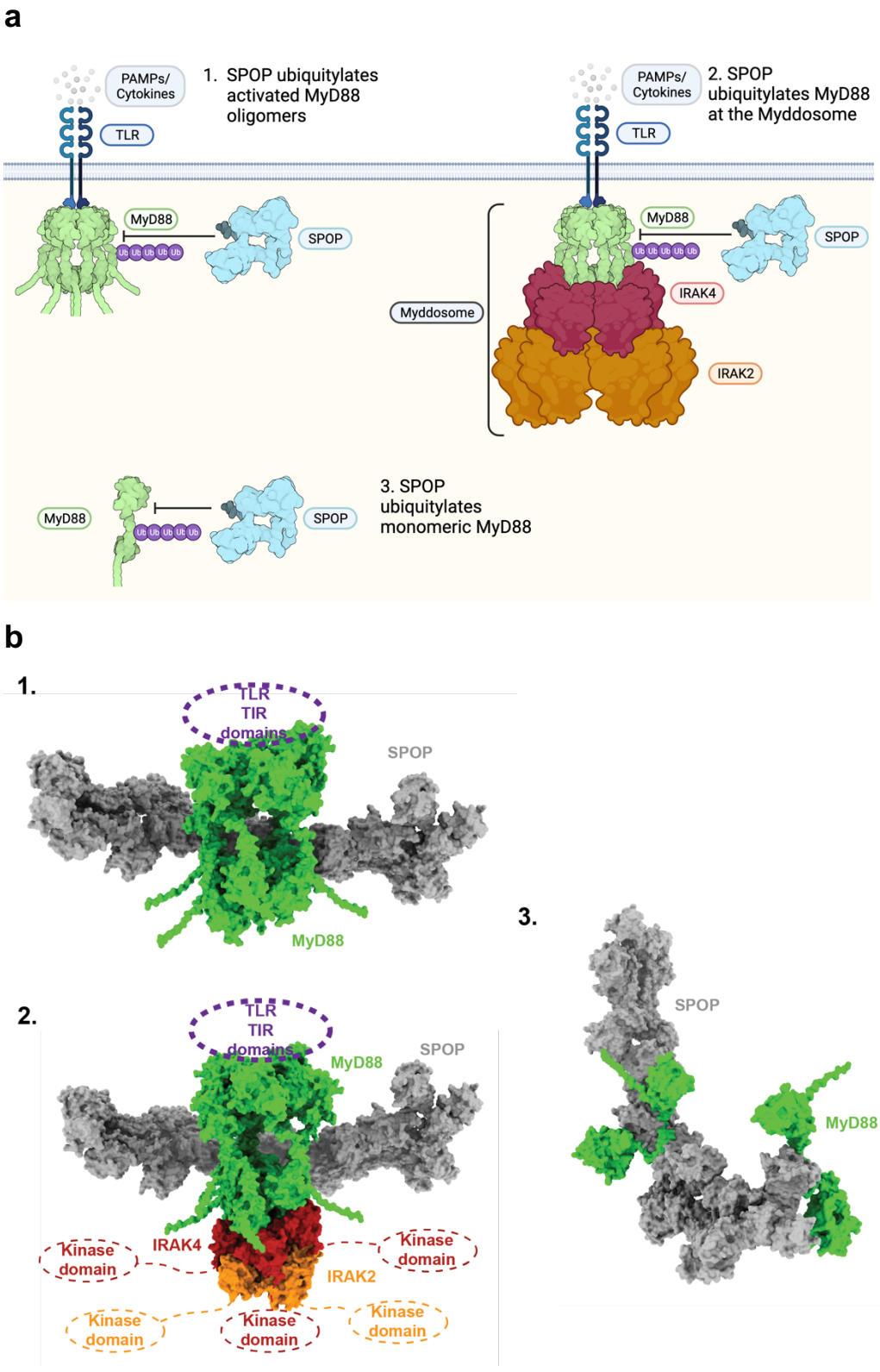


Figure 7.1 – Potential modes of MyD88 regulation by SPOP-mediated ubiquitylation

a Schematic showing the potential binding modes of the SPOP oligomer to MyD88 for its ubiquitylation. 1. represents the SPOP oligomer ubiquitylating MyD88 while it is assembled as a hexa-oligomer at the activated toll-like receptor (TLR), prior to binding of IRAKs, potentially leading to inhibition of IRAK binding and MyD88 degradation. 2. represents the SPOP oligomer

ubiquitylating MyD88 while it is assembled in a 6:4:4 stoichiometry MyD88:IRAK4:IRAK2 Myddosome, potentially leading to disassembly of the Myddosome and MyD88 degradation. 3. represents the SPOP oligomer ubiquitylating free MyD88, preventing it from assembling into the Myddosome. **b** shows the potential SPOP-MyD88 binding modes as in **a**. Surface representation of the structures are shown. The SPOP oligomer is a predicted structural model [100]. The death domain truncation Myddosome structure is shown (PDB ID 3MOP [108]). For 1. and 2., the AlphaFold [96] predicted model of full-length MyD88 is superimposed based on structural similarity onto the MyD88 death domains of the Myddosome model. The approximate binding location of the TLR TIR domains are represented by purple ovals. The SPOP oligomer is approximately placed so the MATH domains of SPOP are in proximity to the MyD88 binding sites. For 2., the missing kinase domains of IRAK4 and IRAK2 are shown as ovals in their respective colours. For 3., the crystal structure of the SPOP MATH domain and the MyD88 peptide from this study is superimposed onto the MATH domains of the predicted SPOP oligomer. The AlphaFold predicted structure of the MyD88 Death and TIR domains are placed relative to the MyD88 peptide.

Despite the difficulty in reconstructing a $SPOP^{18-359}$-$MyD88^{DD-ID}$ cryo-EM 3D map and the limited resolution of the map, the cryo-EM micrographs and 2D classes appear to show a convincing complex being formed between $SPOP^{18-359}$ and $MyD88^{DD-ID}$ (**Fig. 5.10**). Whether or not this type of complex forms in cells remains to be seen. Substrates have been shown to affect the subcellular localisation of SPOP and their binding may nucleate or increase SPOP oligomerisation. The structure of the Myddosome alone or in complex with SPOP could reveal whether the pre-formed SPOP oligomer is able to bind to and ubiquitylate MyD88 at the Myddosome, or if recruitment to MyD88 triggers SPOP oligomerisation (**Fig. 7.1**).

7.3.1 - Future work needed to gain the structure of the SPOP-MyD88 complex

Robust experiments will need to be designed in order to answer the outstanding questions to clarify how exactly SPOP-mediated termination of Myddosome signalling occurs. Optimising the production of the SPOP-Myddosome complex is crucial to fulfil this. Future efforts should focus on producing a stable SPOP-MyD88 or SPOP-Myddosome complex at high enough concentrations for structural characterisation by cryo-EM.

IRAK4 is constitutively active, where it phosphorylates either IRAK2 or IRAK1 for activation of the Myddosome complex. In a cellular context, this may lead to negative regulation of the Myddosome and its ubiquitylation and degradation. This makes the production of recombinant Myddosome complex difficult, as mammalian and insect cell expression systems possess the machinery required to degrade the Myddosome. On the other hand, bacterial expression of the full-length Myddosome has not been achieved due to low expression and solubility.

In order to stabilise IRAK4, there is the potential to use kinase inhibitors such as Staurosporine, as well as mutating the Mg^{2+} co-ordinating D-329 residue of the DFG motif in the kinase domain ($IRAK4^{D329A}$) to prevent the hyper-phosphorylation activity of IRAK4. Additionally, the MyD88 cancer mutation $MyD88^{L252P}$, an oligomer stabilising mutation that promotes Myddosome formation, could be used. This mutation could be used not only to aid in complex stability for experimental studies but is also an important avenue for study of disease biology due to the high prevalence of this mutation in a variety of lymphomas. Since gaining the high-resolution structure of the SPOP-Myddosome complex by cryo-EM is not guaranteed, cross-linking mass spectrometry and hydrogen deuterium exchange mass spectrometry could also be used to visualise protein-protein interactions within the complex as an integrated structural biology approach.

Overall, this could provide a workflow where insect cell expressed Myddosome complex can be reconstituted with various bacterially expressed SPOP truncations (monomer or oligomer) to probe the effect of SPOP oligomerization on Myddosome binding, which may shed light on the mechanism of SPOP termination of Myddosome signalling. It will also be important to validate binding of the full-length proteins using biophysical methods to measure K_ds and association in cells.

7.4 - High resolution cryo-EM structure of the UFL1 E3 ligase complex bound to the 60S ribosome

At the beginning of this project the ribosome had been identified as a UFMylation target, which was linked to ER-associated RQC after translation stalling [137, 175]. How the UFL1 E3 ligase complex assembles and the mechanism of ribosome binding and UFMylation was unknown at the time.

In **chapter 6** the high-resolution cryo-EM structure of the UFL1 ligase complex bound to the 60S ribosome was elucidated. This revealed an unusual ligase architecture which wraps around the 60S ribosome, with three of the complex proteins, UFL1, CDK5RAP3 and UFBP1, binding to the ribosome. Consistent with previous reports stating that the UFL1 ligase complex assists in RQC after translation stalling, the mode of UFL1 ligase complex binding would be incompatible with active ribosome translation, and suggests that the UFL1 ligase complex binds after partial disassembly of stalled or terminated 80S ribosomes. An unexpected aspect of the cryo-EM structure was the identification of a UFL1 loop bound to the PTC (**Fig. 6.14**). This loop resembles the catalytic region of the release factor eRF1, and may even act as an alternate release factor for ER-stalled ribosomes to release stalled polypeptide from tRNA (**Fig. 6.17**).

7.4.1 - The UFL1 ligase complex in RQC

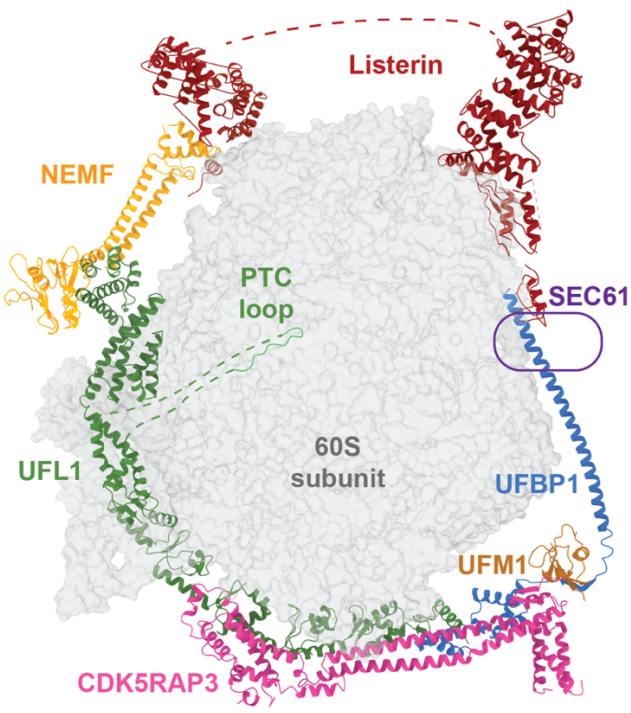
The RQC mechanism for the rescue of stalled cytosolic ribosomes has been well studied. However, in the case of ER-bound ribosomes, the prematurely terminated nascent polypeptide is inaccessible to the cytosolic Listerin E3 ligase complex, potentially leading to build up of unfolded protein in the ER lumen. There is therefore a need for an additional RQC pathway for stalled ER-bound ribosomes, a process that involves the UFL1 E3 ligase complex.

A key question is do the Listerin and UFL1 ligases work together to result in the rescue of stalled ribosomes and degradation of polypeptide? Superimposition of the UFL1 ligase complex and the Listerin ligase complex show that they can bind to the 60S simultaneously, without producing any major clashes (**Fig. 7.2**). This could provide a mechanism for ER-RQC where the UFL1 ligase complex is initially recruited to the stalled peptide-tRNA-60S ribosome complex after dissociation of the 40S subunit. Catalytic action by the UFL1 PTC loop may hydrolyse the peptide-tRNA bond, releasing tRNA from the ribosome and allowing the CTD of UFL1 to fully bind. It is possible that NEMF could then bind to the CTD of UFL1 (**Fig. 7.2 a**). Contrary to the UFL1 ligase complex binding, during cytosolic RQC, NEMF recognises and binds to P-tRNA of the stalled peptide-tRNA-60S complex. Once bound, NEMF then recruits Listerin E3 ligase to the 60S. In the case of ER-RQC, the UFL1 ligase complex-bound 60S after tRNA removal could potentially structurally mimic this peptide-tRNA-60S complex, with the CTD of UFL1 resembling a tRNA molecule, both in localisation and structural shape (**Fig. 7.2 b**). The UFL1 CTD could then allow NEMF to bind and recruit Listerin ligase to the 60S. Currently this is relatively unexplored, but the potential models outlined here may provide a possible mechanism in the role of UFM1 regulation of ER-RQC.

Prior to this work, it was unknown how the ribosome dissociated from SEC61, both during normal translation termination and in the case of ribosome stalling. In CDK5RAP3 knock-out cells, there was an increased accumulation of SEC61 on 60S ribosomes compared to WT, which was shown to be UFMylation dependent (data by Rohan Thakur, Yogesh Kulathu lab) (**Fig. 6.14**). Interestingly, the UFL1 ligase seems to not only act as a writer but also a reader of UFM1, where UFM1 remains stably bound to UFBP1 and CDK5RAP3 after UFMylation of K-134 on RPL26. This stable binding of UFM1 likely clashes with the ER membrane, causing the long alpha helical arm of UFBP1 to displace the 60S from SEC61.

The dissociation of the 60S from SEC61 then reveals the stalled nascent polypeptide to the cytosol, which can now be ubiquitylated by the Listerin ligase complex and degraded. Additionally, this binding mode of the UFL1 ligase complex could help to serve as a mechanism to keep the ribosome inactive while downstream processing takes place, by sterically blocking 80S ribosome reassembly and translation initiation. Whether or not the UFL1 ligase complex directly participates in tRNA removal, is able to recruit NEMF/Listerin and the exact coordination of the ER-RQC steps are unknown. Overall, this role of the UFL1 ligase complex in rescuing stalled ER-bound ribosomes may also explain why ER stress occurs when the UFM1 system is knocked out, as unfolded polypeptides may accumulate in the ER lumen.

a



b

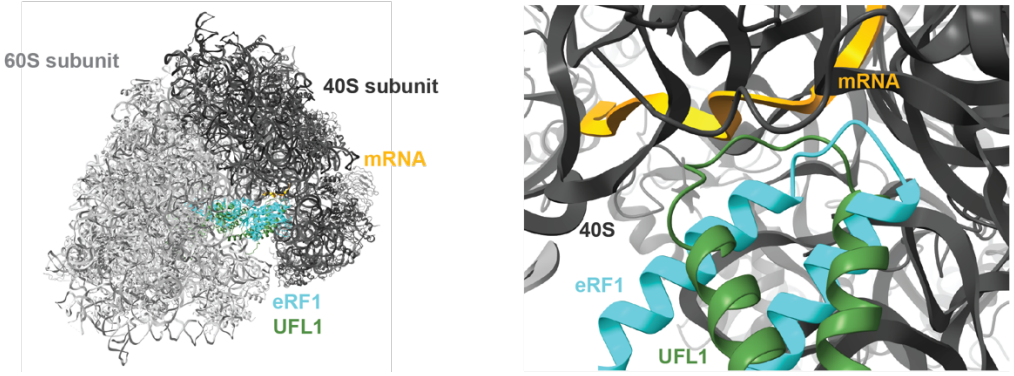


Figure 7.2 – Structures of the UFL1 ligase complex and the Listerin ubiquitin E3 ligase complex bound to the 60S ribosome
a NEMF-Listerin ubiquitin E3 ligase complex (PDB ID 3J92) [221]. Listerin is the E3 ligase. NEMF is a substrate adaptor. Missing Listerin model depicted as dashed line. Approximate location of the SEC61 translocon is depicted as a purple oval.
b C-terminal helices of UFL1 have a similar fold and localisation as the mRNA sensing domain of eRF1 (PDB ID 6ip8) [208], which mimics the structure of an anti-codon stem of a tRNA molecule.

7.4.2 - Future work needed to determine the role of the UFL1 PTC loop

Further study is needed to assess whether the UFL1 ligase complex directly assists in 40S subunit and tRNA removal, either by itself or with other termination factors, as part of the RQC process. Due to similarities with release factors, it will be important to validate whether the UFL1 PTC loop has catalytic activity against the peptide-tRNA bond. Deletion or mutagenesis of the UFL1 PTC loop in cells may reveal whether this loop is critical for stalled ribosome recognition, peptide release from the 60S or other ER-RQC steps. Whether or not the UFMylation of RPL26 is coupled to the action of the UFL1 PTC remains to be seen.

7.5 - Conclusion

Collectively, the data presented in this thesis provides a basis for solving the high-resolution structures of two E3 ligase complexes, which will aid in understanding their biological functions and allow for structure-based therapeutics to be designed. This sets the groundwork for investigating the binding mechanism between SPOP and the Myddosome, in order to understand how exactly SPOP terminates Myddosome signalling. Understanding this mechanism of Myddosome negative regulation is crucial, since termination of Myddosome signalling is needed to allow the differentiation of haematopoietic stem cells into other blood cell lineages and to prevent auto-immune disease. The gaining of the high-resolution structure of the UFL1 ligase complex bound to the 60S ribosome has revealed how the single known ligase capable of ligating UFM1 acts to help rescue stalled ER-ribosomes, a fundamental process in normal and disease biology.

References

1. Goldstein, G., et al., *Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells.* Proceedings of the National Academy of Sciences of the United States of America, 1975. **72**(1): p. 11-15.
2. Goldknopf, I.L., et al., *Presence of protein a24 in rat-liver nucleosomes.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(12): p. 5492-5495.
3. Vijaykumar, S., et al., *Comparison of the 3-dimensional structures of human, yeast, and oat ubiquitin.* Journal of Biological Chemistry, 1987. **262**(13): p. 6396-6399.
4. Komander, D. and M. Rape, *The ubiquitin code*, in *Annual review of biochemistry, vol 81*, R.D. Kornberg, Editor. 2012. p. 203-229.
5. Ciechanover, A., et al., *Covalent affinity purification of ubiquitin-activating enzyme.* Journal of Biological Chemistry, 1982. **257**(5): p. 2537-2542.
6. Hershko, A., et al., *Components of ubiquitin-protein ligase system - resolution, affinity purification, and role in protein breakdown.* Journal of Biological Chemistry, 1983. **258**(13): p. 8206-8214.
7. Hann, Z.S., et al., *Structural basis for adenylation and thioester bond formation in the ubiquitin e1.* Proceedings of the National Academy of Sciences of the United States of America, 2019. **116**(31): p. 15475-15484.
8. Schulman, B.A. and J.W. Harper, *Ubiquitin-like protein activation by e1 enzymes: The apex for downstream signalling pathways.* Nature Reviews Molecular Cell Biology, 2009. **10**(5): p. 319-331.
9. Haas, A.L., P.M. Bright, and V.E. Jackson, *Functional diversity among putative e2 isozymes in the mechanism of ubiquitin-histone ligation.* Journal of Biological Chemistry, 1988. **263**(26): p. 13268-13275.
10. Wu, P.Y., et al., *A conserved catalytic residue in the ubiquitin-conjugating enzyme family (vol 23, pg 4876, 2004).* Embo Journal, 2007. **26**(17): p. 4051-4051.
11. Zheng, X. and T. Hunter, *Pink1, the first ubiquitin kinase.* Embo Journal, 2014. **33**(15): p. 1621-1623.
12. Kazlauskaite, A., et al., *Parkin is activated by pink1-dependent phosphorylation of ubiquitin at ser(65).* Biochemical Journal, 2014. **460**: p. 127-139.
13. Walser, F., et al., *Ubiquitin phosphorylation at thr12 modulates the DNA damage response.* Molecular Cell, 2020. **80**(3): p. 423-436.
14. Hepowit, N.L., et al., *Identification of ubiquitin ser57 kinases regulating the oxidative stress response in yeast.* Elife, 2020. **9**.
15. Ohtake, F., et al., *Ubiquitin acetylation inhibits polyubiquitin chain elongation.* Embo Reports, 2015. **16**(2): p. 192-201.
16. Swatek, K.N. and D. Komander, *Ubiquitin modifications.* Cell Research, 2016. **26**(4): p. 399-422.

17. Jacobson, A.D., et al., *The lysine 48 and lysine 63 ubiquitin conjugates are processed differently by the 26 s proteasome.* Journal of Biological Chemistry, 2009. **284**(51): p. 35485-35494.
18. Madiraju, C., et al., *K63 ubiquitination in immune signaling.* Trends in Immunology, 2022. **43**(2): p. 148-162.
19. Rahighi, S., et al., *Specific recognition of linear ubiquitin chains by nemo is important for nf-kappa b activation.* Cell, 2009. **136**(6): p. 1098-1109.
20. Komander, D., et al., *Molecular discrimination of structurally equivalent lys 63-linked and linear polyubiquitin chains.* Embo Reports, 2009. **10**(5): p. 466-473.
21. Rittinger, K. and F. Ikeda, *Linear ubiquitin chains: Enzymes, mechanisms and biology.* Open Biology, 2017. **7**(4).
22. Morris, J.R. and E. Solomon, *Brca1 : Bard1 induces the formation of conjugated ubiquitin structures, dependent on k6 of ubiquitin, in cells during DNA replication and repair.* Human Molecular Genetics, 2004. **13**(8): p. 807-817.
23. Wickliffe, K.E., et al., *K11-linked ubiquitin chains as novel regulators of cell division.* Trends in Cell Biology, 2011. **21**(11): p. 656-663.
24. Birsa, N., et al., *Lysine 27 ubiquitination of the mitochondrial transport protein miro is dependent on serine 65 of the parkin ubiquitin ligase.* Journal of Biological Chemistry, 2014. **289**(21): p. 14569-14582.
25. Geisler, S., et al., *Pink1/parkin-mediated mitophagy is dependent on vdac1 and p62/sqstm1.* Nature Cell Biology, 2010. **12**(2): p. 119-31.
26. Dikic, I., S. Wakatsuki, and K.J. Walters, *Ubiquitin-binding domains - from structures to functions.* Nature Reviews Molecular Cell Biology, 2009. **10**(10): p. 659-671.
27. Hershko, A., et al., *Proposed role of atp in protein breakdown - conjugation of proteins with multiple chains of the polypeptide of atp-dependent proteolysis.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1980. **77**(4): p. 1783-1786.
28. Pickart, C.M. and I.A. Rose, *Ubiquitin carboxyl-terminal hydrolase acts on ubiquitin carboxyl-terminal amides.* Journal of Biological Chemistry, 1985. **260**(13): p. 7903-7910.
29. Basar, M.A., D.B. Beck, and A. Werner, *Deubiquitylases in developmental ubiquitin signaling and congenital diseases.* Cell Death and Differentiation, 2021. **28**(2): p. 538-556.
30. Walden, M., et al., *Pseudo-dubs as allosteric activators and molecular scaffolds of protein complexes.* Biochemical Society Transactions, 2018. **46**: p. 453-466.
31. Walden, M., et al., *Metabolic control of brisc-shmt2 assembly regulates immune signalling.* Nature, 2019. **570**(7760): p. 194-199.
32. Huang, L., et al., *Structure of an e6ap-ubch7 complex: Insights into ubiquitination by the e2-e3 enzyme cascade.* Science, 1999. **286**(5443): p. 1321-1326.

33. Huibregtse, J.M., M. Scheffner, and P.M. Howley, *Cloning and expression of the cdna for e6-ap, a protein that mediates the interaction of the human papillomavirus e6 oncoprotein with p53.* Molecular and Cellular Biology, 1993. **13**(2): p. 775-784.
34. Weber, J., S. Polo, and E. Maspero, *Hect e3 ligases: A tale with multiple facets.* Frontiers in Physiology, 2019. **10**.
35. Schwarz, S.E., J.L. Rosa, and M. Scheffner, *Characterization of human hect domain family members and their interaction with ubch5 and ubch7.* Journal of Biological Chemistry, 1998. **273**(20): p. 12148-12154.
36. Ries, L.K., et al., *Analysis of ubiquitin recognition by the hect ligase e6ap provides insight into its linkage specificity.* Journal of Biological Chemistry, 2019. **294**(15): p. 6113-6129.
37. Mercier, P., et al., *Structure, interactions, and dynamics of the ring domain from human traf6.* Protein Science, 2007. **16**(4): p. 602-614.
38. Zheng, N., et al., *Structure of a c-cbl-ubch7 complex: Ring domain function in ubiquitin-protein ligases.* Cell, 2000. **102**(4): p. 533-539.
39. Petroski, M.D. and R.J. Deshaies, *Function and regulation of cullin-ring ubiquitin ligases.* Nature Reviews Molecular Cell Biology, 2005. **6**(1): p. 9-20.
40. Zheng, N., et al., *Structure of the cul1-rbx1-skp1-f box(skp2) scf ubiquitin ligase complex.* Nature, 2002. **416**(6882): p. 703-709.
41. Boh, B.K., P.G. Smith, and T. Hagen, *Neddylation-induced conformational control regulates cullin ring ligase activity in vivo.* Journal of Molecular Biology, 2011. **409**(2): p. 136-145.
42. Duda, D.M., et al., *Structural insights into nedd8 activation of cullin-ring ligases: Conformational control of conjugation.* Cell, 2008. **134**(6): p. 995-1006.
43. Saha, A. and R.J. Deshaies, *Multimodal activation of the ubiquitin ligase scf by nedd8 conjugation.* Molecular Cell, 2008. **32**(1): p. 21-31.
44. Aravind, L. and E.V. Koonin, *The u box is a modified ring finger - a common domain in ubiquitination.* Current Biology, 2000. **10**(4): p. R132-R134.
45. Hatakeyama, S. and K.I. Nakayama, *U-box proteins as a new family of ubiquitin ligases.* Biochemical and Biophysical Research Communications, 2003. **302**(4): p. 635-645.
46. Dove, K.K., et al., *Molecular insights into rbr e3 ligase ubiquitin transfer mechanisms.* Embo Reports, 2016. **17**(8): p. 1221-1235.
47. Beasley, S.A., V.A. Hristova, and G.S. Shaw, *Structure of the parkin in-between-ring domain provides insights for e3-ligase dysfunction in autosomal recessive parkinson's disease.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(9): p. 3095-3100.
48. Walden, H. and K. Rittinger, *Rbr ligase-mediated ubiquitin transfer: A tale with many twists and turns.* Nature Structural & Molecular Biology, 2018. **25**(6): p. 440-445.

49. Cotton, T.R. and B.C. Lechtenberg, *Chain reactions: Molecular mechanisms of rbr ubiquitin ligases.* Biochemical Society Transactions, 2020. **48**(4): p. 1737-1750.
50. Ye, Y. and M. Rape, *Building ubiquitin chains: E2 enzymes at work.* Nature Reviews Molecular Cell Biology, 2009. **10**(11): p. 755-764.
51. Kim, H.T., et al., *Certain pairs of ubiquitin-conjugating enzymes (e2s) and ubiquitin-protein ligases (e3s) synthesize nondegradable forked ubiquitin chains containing all possible isopeptide linkages.* Journal of Biological Chemistry, 2007. **282**(24): p. 17375-17386.
52. Wang, M. and C.M. Pickart, *Different hect domain ubiquitin ligases employ distinct mechanisms of polyubiquitin chain synthesis.* Embo Journal, 2005. **24**(24): p. 4324-4333.
53. Pintard, L., A. Willems, and M. Peter, *Cullin-based ubiquitin ligases: Cul3-btb complexes join the family.* Embo Journal, 2004. **23**(8): p. 1681-1687.
54. Nagai, Y., et al., *Identification of a novel nuclear speckle-type protein, spop.* Febs Letters, 1997. **418**(1-2): p. 23-26.
55. Zhuang, M., et al., *Structures of spop-substrate complexes: Insights into molecular architectures of btb-cul3 ubiquitin ligases.* Molecular Cell, 2009. **36**(1): p. 39-50.
56. Clark, A. and M. Burleson, *Spop and cancer: A systematic reviewd.* American Journal of Cancer Research, 2020. **10**(3): p. 704-726.
57. Nakazawa, M., et al., *Spop mutations in prostate cancer: Clinical and genomic features.* Journal of Clinical Oncology, 2021. **39**(6).
58. Blattner, M., et al., *Spop mutations in prostate cancer across demographically diverse patient cohorts.* Neoplasia, 2014. **16**(1): p. 14-20.
59. Li, C., et al., *Tumor-suppressor role for the spop ubiquitin ligase in signal-dependent proteolysis of the oncogenic co-activator src-3/aib1.* Oncogene, 2011. **30**(42): p. 4350-4364.
60. Aurilio, G., et al., *Androgen receptor signaling pathway in prostate cancer: From genetics to clinical applications.* Cells, 2020. **9**(12).
61. Bouchard, J.J., et al., *Cancer mutations of the tumor suppressor spop disrupt the formation of active, phase-separated compartments.* Molecular Cell, 2018. **72**(1): p. 19-36.
62. Guillamot, M., et al., *The e3 ubiquitin ligase spop controls resolution of systemic inflammation by triggering myd88 degradation.* Nature Immunology, 2019. **20**(9): p. 1196-1207.
63. Usher, E.T., et al., *Intrinsically disordered substrates dictate spop subnuclear localization and ubiquitination activity.* Journal of Biological Chemistry, 2021. **296**.
64. Zhang, Q., et al., *Multiple ser/thr-rich degrons mediate the degradation of ci/gli by the cul3-hib/spop e3 ubiquitin ligase.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(50): p. 21191-21196.
65. Ma, J., et al., *Spop mutation induces replication over-firing by impairing geminin ubiquitination and triggers replication*

catastrophe upon atr inhibition. Nature Communications, 2021. **12**(1).

66. Wang, D., et al., *Atm-phosphorylated spop contributes to 53bp1 exclusion from chromatin during DNA replication.* Science Advances, 2021. **7**(25).
67. Janouskova, H., et al., *Opposing effects of cancer-type-specific spop mutants on bet protein degradation and sensitivity to bet inhibitors.* Nature Medicine, 2017. **23**(9): p. 1046-1054.
68. Wang, C.B., Y. Pan, and B.L. Wang, *Suppressor of fused and spop regulate the stability, processing and function of gli2 and gli3 full-length activators but not their repressors.* Development, 2010. **137**(12): p. 2001-2009.
69. Pierce, W.K., et al., *Multiple weak linear motifs enhance recruitment and processivity in spop-mediated substrate ubiquitination.* Journal of Molecular Biology, 2016. **428**(6): p. 1256-1271.
70. Theurillat, J.P.P., et al., *Ubiquitylome analysis identifies dysregulation of effector substrates in spop-mutant prostate cancer.* Science, 2014. **346**(6205): p. 85-89.
71. Zhang, P.Z., et al., *Intrinsic bet inhibitor resistance in spop-mutated prostate cancer is mediated by bet protein stabilization and akt-mtorc1 activation.* Nature Medicine, 2017. **23**(9): p. 1055-1062.
72. Zhang, P.Z., et al., *Destruction of ddit3/chop protein by wild-type spop but not prostate cancer-associated mutants.* Human Mutation, 2014. **35**(9): p. 1142-1151.
73. Gan, W.J., et al., *Spop promotes ubiquitination and degradation of the erg oncoprotein to suppress prostate cancer progression.* Molecular Cell, 2015. **59**(6): p. 917-930.
74. Zhu, H.R., et al., *Spop e3 ubiquitin ligase adaptor promotes cellular senescence by degrading the senp7 desumoylase.* Cell Reports, 2015. **13**(6): p. 1183-1193.
75. Zhu, K., et al., *Spop-containing complex regulates setd2 stability and h3k36me3-coupled alternative splicing.* Nucleic Acids Research, 2017. **45**(1): p. 92-105.
76. Wu, F., et al., *Prostate cancer-associated mutation in spop impairs its ability to target cdc20 for poly-ubiquitination and degradation.* Cancer Letters, 2017. **385**: p. 207-214.
77. Luo, J., et al., *Spop promotes sirt2 degradation and suppresses non-small cell lung cancer cell growth.* Biochemical and Biophysical Research Communications, 2017. **483**(2): p. 880-884.
78. Zhang, L.L., et al., *Tumor suppressor spop ubiquitinates and degrades eg1n2 to compromise growth of prostate cancer cells.* Cancer Letters, 2017. **390**: p. 11-20.
79. Geng, C., et al., *Spop regulates prostate epithelial cell proliferation and promotes ubiquitination and turnover of c-myc oncoprotein.* Oncogene, 2017. **36**(33): p. 4767-4777.
80. Jin, X.F., et al., *Dysregulation of inf2-mediated mitochondrial fission in spop-mutated prostate cancer.* Plos Genetics, 2017. **13**(4).

81. Tan, Y.Y., et al., *Cullin 3(spop) ubiquitin e3 ligase promotes the poly-ubiquitination and degradation of hdac6.* Oncotarget, 2017. **8**(29): p. 47890-47901.
82. Geng, C., et al., *Androgen receptor is the key transcriptional mediator of the tumor suppressor spop in prostate cancer.* Cancer Research, 2014. **74**(19): p. 5631-5643.
83. Zhang, P., et al., *Endometrial cancer-associated mutants of spop are defective in regulating estrogen receptor-alpha protein turnover.* Cell Death & Disease, 2015. **6**.
84. Jin, X.F., et al., *Prostate cancer-associated spop mutations lead to genomic instability through disruption of the spop hipk2 axis.* Nucleic Acids Research, 2021. **49**(12): p. 6788-6803.
85. Zhang, J.F., et al., *Spop promotes nanog destruction to suppress stem cell traits and prostate cancer progression.* Developmental Cell, 2019. **48**(3): p. 329-344.
86. Ma, J., et al., *Spop promotes atf2 ubiquitination and degradation to suppress prostate cancer progression.* Journal of Experimental & Clinical Cancer Research, 2018. **37**.
87. Gang, X.K., et al., *Speckle-type poz protein suppresses lipid accumulation and prostate cancer growth by stabilizing fatty acid synthase.* Prostate, 2019. **79**(8): p. 864-871.
88. Shi, Q., et al., *Prostate cancer-associated spop mutations enhance cancer cell survival and docetaxel resistance by upregulating caprin1-dependent stress granule assembly.* Molecular Cancer, 2019. **18**(1).
89. Luo, Z.Z., et al., *Spop promotes cdca5 degradation to regulate prostate cancer progression via the akt pathway.* Neoplasia, 2021. **23**(10): p. 1037-1047.
90. Jiang, Q.W., et al., *Spop-mediated ubiquitination and degradation of pdk1 suppresses akt kinase activity and oncogenic functions.* Molecular Cancer, 2021. **20**(1).
91. Yuan, D., et al., *Spop attenuates migration and invasion of choriocarcinoma cells by promoting dhx9 degradation.* American Journal of Cancer Research, 2020. **10**(8): p. 2428-2445.
92. Kim, B., et al., *Breast cancer metastasis suppressor 1 (brms1) is destabilized by the cul3-spop e3 ubiquitin ligase complex.* Biochemical and Biophysical Research Communications, 2011. **415**(4): p. 720-726.
93. Su, S.Y., et al., *Spop and otud7a control ews-fli1 protein stability to govern ewing sarcoma growth.* Advanced Science, 2021. **8**(14).
94. Fan, Y., et al., *Erk1/2 inhibits cullin 3/spop-mediated prlz ubiquitination and degradation to modulate prostate cancer progression.* Cell Death and Differentiation, 2022. **29**(8): p. 1611-1624.
95. Errington, W.J., et al., *Adaptor protein self-assembly drives the control of a cullin-ring ubiquitin ligase.* Structure, 2012. **20**(7): p. 1141-1153.
96. Jumper, J., et al., *Highly accurate protein structure prediction with alphafold.* Nature, 2021. **596**(7873): p. 583-589.

97. Martin-Blanco, E., et al., *Puckered encodes a phosphatase that mediates a feedback loop regulating jnk activity during dorsal closure in drosophila.* Genes & Development, 1998. **12**(4): p. 557-570.
98. Fujimoto, K. and K.S. Polonsky, *Pdx1 and other factors that regulate pancreatic beta-cell survival.* Diabetes Obesity & Metabolism, 2009. **11**: p. 30-37.
99. Ostertag, M.S., et al., *The structure of the spop-pdx1 interface reveals insights into the phosphorylation-dependent binding regulation.* Structure, 2019. **27**(2): p. 327-334.
100. van Geersdaele, L.K., et al., *Structural basis of high-order oligomerization of the cullin-3 adaptor spop.* Acta Crystallographica Section D-Biological Crystallography, 2013. **69**: p. 1677-1684.
101. Gschweitl, M., et al., *A spopl/cullin-3 ubiquitin ligase complex regulates endocytic trafficking by targeting eps15 at endosomes.* Elife, 2016. **5**.
102. Boettcher, S. and M.G. Manz, *Regulation of inflammation- and infection-driven hematopoiesis.* Trends in Immunology, 2017. **38**(5): p. 345-357.
103. Pietras, E.M., et al., *Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal.* Nature Cell Biology, 2016. **18**(6): p. 607-618.
104. Zaretsky, A.G., J.B. Engiles, and C.A. Hunter, *Infection-induced changes in hematopoiesis.* Journal of Immunology, 2014. **192**(1): p. 27-33.
105. Dinarello, C.A., *Introduction to the interleukin-1 family of cytokines and receptors: Drivers of innate inflammation and acquired immunity.* Immunological Reviews, 2018. **281**(1): p. 5-7.
106. Wesche, H., et al., *Myd88: An adapter that recruits irak to the il-1 receptor complex (reprinted from immunity, vol 7, pg 837-847, 1997).* Journal of Immunology, 2013. **190**(1): p. 5-15.
107. Yamamoto, M. and S. Akira, *Tir domain--containing adaptors regulate tlr-mediated signaling pathways.* Nihon rinsho. Japanese journal of clinical medicine, 2004. **62**(12): p. 2197-203.
108. Lin, S.-C., Y.-C. Lo, and H. Wu, *Helical assembly in the myd88-irak4-irak2 complex in tlr/il-1r signalling.* Nature, 2010. **465**(7300): p. 885-890.
109. O'Neill, L.A.J. and A.G. Bowie, *The family of five: Tir-domain-containing adaptors in toll-like receptor signalling.* Nature Reviews Immunology, 2007. **7**(5): p. 353-364.
110. Loiarro, M., et al., *Mutational analysis identifies residues crucial for homodimerization of myeloid differentiation factor 88 (myd88) and for its function in immune cells.* Journal of Biological Chemistry, 2013. **288**(42): p. 30210-30222.
111. Cheng, H., et al., *Regulation of irak-4 kinase activity via autophosphorylation within its activation loop.* Biochemical and Biophysical Research Communications, 2007. **352**(3): p. 609-616.

112. Ferrao, R., et al., *Irak4 dimerization and trans-autophosphorylation are induced by myddosome assembly.* Molecular Cell, 2014. **55**(6): p. 891-903.
113. Vollmer, S., et al., *The mechanism of activation of irak1 and irak4 by interleukin-1 and toll-like receptor agonists.* Biochemical Journal, 2017. **474**: p. 2027-2038.
114. Gupta, P., et al., *Pu.1 and partners: Regulation of haematopoietic stem cell fate in normal and malignant haematopoiesis.* Journal of Cellular and Molecular Medicine, 2009. **13**(11-12): p. 4349-4363.
115. Walsh, M.C., J. Lee, and Y. Choi, *Tumor necrosis factor receptor-associated factor 6 (traf6) regulation of development, function, and homeostasis of the immune system.* Immunological Reviews, 2015. **266**(1): p. 72-92.
116. Lee, Y.S., et al., *Smad6-specific recruitment of smurf e3 ligases mediates tgf-beta 1-induced degradation of myd88 in tlr4 signalling.* Nature Communications, 2011. **2**.
117. Lee, B.-C., et al., *Deubiquitinase cyld acts as a negative regulator for bacterium nthi-induced inflammation by suppressing k63-linked ubiquitination of myd88.* Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(2): p. E165-E171.
118. Li, Q., et al., *Spop promotes ubiquitination and degradation of myd88 to suppress the innate immune response.* Plos Pathogens, 2020. **16**(5).
119. Jin, X.F., et al., *Crl3-spop ubiquitin ligase complex suppresses the growth of diffuse large b-cell lymphoma by negatively regulating the myd88/nf-kappa b signaling.* Leukemia, 2020. **34**(5): p. 1305-1314.
120. Hu, Y.-H., et al., *Spop negatively regulates toll-like receptor-induced inflammation by disrupting myd88 self-association.* Cellular & Molecular Immunology, 2021. **18**(7): p. 1708-1717.
121. Komatsu, M., et al., *A novel protein-conjugating system for ufm1, a ubiquitin-fold modifier.* Embo Journal, 2004. **23**(9): p. 1977-1986.
122. Sasakawa, H., et al., *Solution structure and dynamics of ufm1, a ubiquitin-fold modifier 1.* Biochemical and Biophysical Research Communications, 2006. **343**(1): p. 21-26.
123. Banerjee, S., M. Kumar, and R. Wiener, *Decrypting ufmylation: How proteins are modified with ufm1.* Biomolecules, 2020. **10**(10).
124. Yoo, H.M., et al., *Modification of asc1 by ufm1 is crucial for er alpha transactivation and breast cancer development.* Molecular Cell, 2014. **56**(2): p. 261-274.
125. Lemaire, K., et al., *Ubiquitin fold modifier 1 (ufm1) and its target ufbp1 protect pancreatic beta cells from er stress-induced apoptosis.* Plos One, 2011. **6**(4).
126. Zhang, Y., et al., *Transcriptional regulation of the ufm1 conjugation system in response to disturbance of the endoplasmic reticulum homeostasis and inhibition of vesicle trafficking.* Plos One, 2012. **7**(11).

127. Rubio, M.D., et al., *Dysfunction of the ubiquitin proteasome and ubiquitin-like systems in schizophrenia.* Neuropsychopharmacology, 2013. **38**(10): p. 1910-1920.
128. Lu, Y., et al., *Identification of potential markers for type 2 diabetes mellitus via bioinformatics analysis.* Molecular Medicine Reports, 2020. **22**(3): p. 1868-1882.
129. Yoo, H.M., et al., *Modification of er alpha by ufm1 increases its stability and transactivity for breast cancer development.* Molecules and Cells, 2022. **45**(6): p. 425-434.
130. Tatsumi, K., et al., *A novel type of e3 ligase for the ufm1 conjugation system.* Journal of Biological Chemistry, 2010. **285**(8): p. 5417-5427.
131. Peter, J.J., et al., *A non-canonical scaffold-type e3 ligase complex mediates protein ufmylation.* Embo Journal, 2022. **41**(21).
132. Wu, J.C., et al., *A novel c53/lzap-interacting protein regulates stability of c53/lzap and ddrgk domain-containing protein 1 (ddrgk1) and modulates nf-kappa b signaling.* Journal of Biological Chemistry, 2010. **285**(20): p. 15126-15136.
133. Liu, G., et al., *Nmr and x-ray structures of human e2-like ubiquitin-fold modifier conjugating enzyme 1 (ufc1) reveal structural and functional conservation in the metazoan ufm1-uba5-ufc1 ubiquination pathway.* Journal of Structural and Functional Genomics, 2009. **10**(2): p. 127-136.
134. Ishimura, R., et al., *The ufm1 system regulates er-phagy through the ufmylation of cyb5r3.* Nature communications, 2022. **13**(1): p. 7857-7857.
135. Qin, B., et al., *Ufl1 promotes histone h4 ufmylation and atm activation.* Nature Communications, 2019. **10**.
136. Liu, J., et al., *Ufmylation maintains tumour suppressor p53 stability by antagonizing its ubiquitination.* Nature Cell Biology, 2020. **22**(9): p. 1056-1063.
137. Walczak, C.P., et al., *Ribosomal protein rpl26 is the principal target of ufmylation.* Proceedings of the National Academy of Sciences of the United States of America, 2019. **116**(4): p. 1299-1308.
138. Wang, L.H., et al., *Ufmylation of rpl26 links translocation-associated quality control to endoplasmic reticulum protein homeostasis.* Cell Research, 2020. **30**(1).
139. Simsek, D., et al., *The mammalian ribo-interactome reveals ribosome functional diversity and heterogeneity.* Cell, 2017. **169**(6): p. 1051-1065.
140. Kapp, L.D. and J.R. Lorsch, *Gtp-dependent recognition of the methionine moiety on initiator trna by translation factor eif2.* Journal of Molecular Biology, 2004. **335**(4): p. 923-936.
141. Benne, R. and J.W.B. Hershey, *Mechanism of action of protein-synthesis initiation-factors from rabbit reticulocytes.* Journal of Biological Chemistry, 1978. **253**(9): p. 3078-3087.
142. Korneeva, N.L., et al., *Mutually cooperative binding of eukaryotic translation initiation factor (eif) 3 and eif4a to human eif4g-1.* Journal of Biological Chemistry, 2000. **275**(52): p. 41369-41376.

143. Keiper, B.D., W.N. Gan, and R.E. Rhoads, *Protein synthesis initiation factor 4g.* International Journal of Biochemistry & Cell Biology, 1999. **31**(1): p. 37-41.
144. Algire, M.A., D. Maag, and J.R. Lorsch, *P-i release from elf2, not gtp hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation.* Molecular Cell, 2005. **20**(2): p. 251-262.
145. Yi, S.-H., et al., *Conformational rearrangements upon start codon recognition in human 48s translation initiation complex.* Nucleic Acids Research, 2022. **50**(9): p. 5282-5298.
146. Pestova, T.V., et al., *The joining of ribosomal subunits in eukaryotes requires eif5b.* Nature, 2000. **403**(6767): p. 332-335.
147. Lapointe, C.P., et al., *Eif5b and eif1a reorient initiator trna to allow ribosomal subunit joining.* Nature, 2022. **607**(7917): p. 186-190.
148. Lee, J.H., et al., *Initiation factor eif5b catalyzes second gtp-dependent step in eukaryotic translation initiation.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(26): p. 16689-16694.
149. Hingerty, B., R.S. Brown, and A. Jack, *Further refinement of structure of yeast transfer-rna phe.* Journal of Molecular Biology, 1978. **124**(3): p. 523-534.
150. Negrutskii, B.S. and A.V. El'skaya, *Eukaryotic translation elongation factor 1 alpha: Structure, expression, functions, and possible role in aminoacyl-trna channeling*, in *Progress in nucleic acid research and molecular biology, volume 60*, K. Moldave, Editor. 1998. p. 47-78.
151. Graifer, D. and G. Karpova, *Interaction of trna with eukaryotic ribosome.* International Journal of Molecular Sciences, 2015. **16**(4): p. 7173-7194.
152. Caulfield, T. and B. Devkota, *Motion of transfer rna from the a/t state into the a-site using docking and simulations.* Proteins-Structure Function and Bioinformatics, 2012. **80**(11): p. 2489-2500.
153. Schmeing, T.M., et al., *Structural insights into the roles of water and the 2 ' hydroxyl of the p site trna in the peptidyl transferase reaction.* Molecular Cell, 2005. **20**(3): p. 437-448.
154. Simonovic, M. and T.A. Steitz, *A structural view on the mechanism of the ribosome-catalyzed peptide bond formation.* Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms, 2009. **1789**(9-10): p. 612-623.
155. Polacek, N., et al., *The critical role of the universally conserved a2602 of 23s ribosomal rna in the release of the nascent peptide during translation termination.* Molecular Cell, 2003. **11**(1): p. 103-112.
156. Amort, M., et al., *An intact ribose moiety at a2602 of 23s rrna is key to trigger peptidyl-trna hydrolysis during translation termination.* Nucleic Acids Research, 2007. **35**(15): p. 5130-5140.
157. Moazed, D. and H.F. Noller, *Intermediate states in the movement of transfer-rna in the ribosome.* Nature, 1989. **342**(6246): p. 142-148.

158. Frank, J., et al., *The process of mrna-trna translocation.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(50): p. 19671-19678.
159. Bertram, G., et al., *Terminating eukaryote translation: Domain 1 of release factor erf1 functions in stop codon recognition.* Rna, 2000. **6**(9): p. 1236-1247.
160. Conard, S.E., et al., *Identification of erf1 residues that play critical and complementary roles in stop codon recognition.* Rna, 2012. **18**(6): p. 1210-1221.
161. Frolova, L., A. Seit-Nebi, and L. Kisselev, *Highly conserved niks tetrapeptide is functionally essential in eukaryotic translation termination factor erf1.* Rna, 2002. **8**(2): p. 129-136.
162. Frolova, L.Y., et al., *Mutations in the highly conserved ggq motif of class 1 polypeptide release factors abolish ability of human erf1 to trigger peptidyl-trna hydrolysis.* Rna, 1999. **5**(8): p. 1014-1020.
163. Frolova, L., et al., *Eukaryotic polypeptide chain release factor erf3 is an erf1- and ribosome-dependent guanosine triphosphatase.* Rna, 1996. **2**(4): p. 334-341.
164. Cheng, Z., et al., *Structural insights into erf3 and stop codon recognition by erf1.* Genes & Development, 2009. **23**(9): p. 1106-1118.
165. Song, H.W., et al., *The crystal structure of human eukaryotic release factor erf1 - mechanism of stop codon recognition and peptidyl-trna hydrolysis.* Cell, 2000. **100**(3): p. 311-321.
166. Juszkiewicz, S. and R.S. Hegde, *Initiation of quality control during poly(a) translation requires site-specific ribosome ubiquitination.* Molecular Cell, 2017. **65**(4): p. 743-750.
167. Gamble, C.E., et al., *Adjacent codons act in concert to modulate translation efficiency in yeast.* Cell, 2016. **166**(3): p. 679-690.
168. Lu, J. and C. Deutsch, *Electrostatics in the ribosomal tunnel modulate chain elongation rates.* Journal of Molecular Biology, 2008. **384**(1): p. 73-86.
169. Sundaramoorthy, E., et al., *Znf598 and rack1 regulate mammalian ribosome-associated quality control function by mediating regulatory 40s ribosomal ubiquitylation.* Molecular Cell, 2017. **65**(4): p. 751-760.
170. Shao, S., K. von der Malsburg, and R.S. Hegde, *Listerin-dependent nascent protein ubiquitination relies on ribosome subunit dissociation.* Molecular Cell, 2013. **50**(5): p. 637-648.
171. Pisareva, V.P., et al., *Dissociation by pelota, hbs1 and abce1 of mammalian vacant 80s ribosomes and stalled elongation complexes.* Embo Journal, 2011. **30**(9): p. 1804-1817.
172. Shoemaker, C.J., D.E. Eyler, and R. Green, *Dom34:Hbs1 promotes subunit dissociation and peptidyl-trna drop-off to initiate no-go decay.* Science, 2010. **330**(6002): p. 369-372.
173. Rendon, O.Z., et al., *Vms1p is a release factor for the ribosome-associated quality control complex.* Nature Communications, 2018. **9**.

174. Verma, R., et al., *Vmsl and ankzf1 peptidyl-trna hydrolases release nascent chains from stalled ribosomes.* Nature, 2018. **557**(7705): p. 446-451.
175. Wang, L., et al., *Ufmylation of rpl26 links translocation-associated quality control to endoplasmic reticulum protein homeostasis.* Cell Research, 2020. **30**(1).
176. Grollman, A.P., *Inhibitors of protein biosynthesis .2. Mode of action of anisomycin.* Journal of Biological Chemistry, 1967. **242**(13): p. 3226-3233.
177. Fresno, M., A. Jimenez, and D. Vazquez, *Inhibition of translation in eukaryotic systems by harringtonine.* European Journal of Biochemistry, 1977. **72**(2): p. 323-330.
178. Schneider-Poetsch, T., et al., *Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin.* Nature Chemical Biology, 2010. **6**(3): p. 209-217.
179. Winter, G., *Xia2: An expert system for macromolecular crystallography data reduction.* Journal of Applied Crystallography, 2010. **43**: p. 186-190.
180. Emsley, P., et al., *Features and development of coot.* Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 486-501.
181. Liebschner, D., et al., *Macromolecular structure determination using x-rays, neutrons and electrons: Recent developments in phenix.* Acta Crystallographica Section D-Structural Biology, 2019. **75**: p. 861-877.
182. Williams, C.J., et al., *Molprobity: More and better reference data for improved all-atom structure validation.* Protein Science, 2018. **27**(1): p. 293-315.
183. Zivanov, J., et al., *New tools for automated high-resolution cryo-em structure determination in relion-3.* Elife, 2018. **7**.
184. Rohou, A. and N. Grigorieff, *Ctffind4: Fast and accurate defocus estimation from electron micrographs.* Journal of Structural Biology, 2015. **192**(2): p. 216-221.
185. Punjani, A., et al., *Cryosparc: Algorithms for rapid unsupervised cryo-em structure determination.* Nature Methods, 2017. **14**(3): p. 290-296.
186. Zheng, S.Q., et al., *Motioncor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy.* Nature Methods, 2017. **14**(4): p. 331-332.
187. Pettersen, E.F., et al., *Ucsf chimerax: Structure visualization for researchers, educators, and developers.* Protein Science, 2021. **30**(1): p. 70-82.
188. Cuneo, M.J., et al., *Higher-order spop assembly reveals a basis for cancer mutant dysregulation.* Molecular Cell, 2023. **83**(5): p. 731-745.
189. Zhong, E.D., et al., *Cryodrgn: Reconstruction of heterogeneous cryo-em structures using neural networks.* Nature Methods, 2021. **18**(2): p. 176-185.

190. Wagner, T., et al., *Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em.* Communications Biology, 2019. **2**.
191. Punjani, A. and D.J. Fleet, *3dflex: Determining structure and motion of flexible proteins from cryo-em.* Nature Methods, 2023. **20**(6): p. 860-870.
192. Sanchez-Garcia, R., et al., *Deepemhancer: A deep learning solution for cryo-em volume post-processing.* Communications Biology, 2021. **4**(1).
193. Jomaa, A., et al., *Mechanism of signal sequence handover from nac to srp on ribosomes during er-protein targeting.* Science, 2022. **375**(6583): p. 839-844.
194. Paluda, A., et al., *Ubiquitin and a charged loop regulate the ubiquitin e3 ligase activity of ark2c.* Nature Communications, 2022. **13**(1).
195. Crooks, G.E., et al., *Weblogo: A sequence logo generator.* Genome Research, 2004. **14**(6): p. 1188-1190.
196. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega.* Molecular Systems Biology, 2011. **7**.
197. Myers, E.W. and W. Miller, *Optimal alignments in linear-space.* Computer Applications in the Biosciences, 1988. **4**(1): p. 11-17.
198. Wang, Z., et al., *A peptide binder of e3 ligase adaptor spop disrupts oncogenic spop-protein interactions in kidney cancer cells.* Chinese Journal of Chemistry, 2021. **39**(2): p. 274-280.
199. Ostertag, M.S., et al., *Structural insights into bet client recognition of endometrial and prostate cancer-associated spop mutants.* Journal of Molecular Biology, 2019. **431**(11): p. 2213-2221.
200. Ngo, V.N., et al., *Oncogenically active myd88 mutations in human lymphoma.* Nature, 2011. **470**(7332): p. 115-9.
201. Bohers, E., et al., *Targetable activating mutations are very frequent in gcb and abc diffuse large b-cell lymphoma.* Genes Chromosomes & Cancer, 2014. **53**(2): p. 144-153.
202. Kridel, R., et al., *Cell of origin of transformed follicular lymphoma.* Blood, 2015. **126**(18): p. 2118-2127.
203. Marzahn, M.R., et al., *Higher-order oligomerization promotes localization of spop to liquid nuclear speckles.* Embo Journal, 2016. **35**(12): p. 1254-1275.
204. Mirdita, M., et al., *Colabfold: Making protein folding accessible to all.* Nature Methods, 2022. **19**(6): p. 679-682.
205. Moncrieffe, M.C., et al., *Myd88 death-domain oligomerization determines myddosome architecture: Implications for toll-like receptor signaling.* Structure, 2020. **28**(3): p. 281-289.
206. Clabbers, M.T.B., et al., *Myd88 tir domain higher-order assembly interactions revealed by microcrystal electron diffraction and serial femtosecond crystallography.* Nature Communications, 2021. **12**(1).
207. McGilvray, P.T., et al., *An er translocon for multi-pass membrane protein biogenesis.* Elife, 2020. **9**.

208. Yokoyama, T., et al., *Hcv ires captures an actively translating 80s ribosome.* Molecular Cell, 2019. **74**(6): p. 1205-1214.
209. Shanmuganathan, V., et al., *Structural and mutational analysis of the ribosome-arresting human xbp1u.* Elife, 2019. **8**.
210. Liu, J., et al., *A critical role of ddrgk1 in endoplasmic reticulum homoeostasis via regulation of ire1 alpha stability.* Nature Communications, 2017. **8**.
211. Fei, J., et al., *Coupling of ribosomal l1 stalk and trna dynamics during translation elongation.* Molecular Cell, 2008. **30**(3): p. 348-359.
212. Valle, M., et al., *Locking and unlocking of ribosomal motions.* Cell, 2003. **114**(1): p. 123-134.
213. Korostelev, A., et al., *Crystal structure of a 70s ribosome-trna complex reveals functional interactions and rearrangements.* Cell, 2006. **126**(6): p. 1065-1077.
214. Wilson, D.N., S. Arenz, and R. Beckmann, *Translation regulation via nascent polypeptide-mediated ribosome stalling.* Current Opinion in Structural Biology, 2016. **37**: p. 123-133.
215. de Loubresse, N.G., et al., *Structural basis for the inhibition of the eukaryotic ribosome.* Nature, 2014. **513**(7519): p. 517-22.
216. Seefeldt, A.C., et al., *The proline-rich antimicrobial peptide onc112 inhibits translation by blocking and destabilizing the initiation complex.* Nature Structural & Molecular Biology, 2015. **22**(6): p. 470-5.
217. Wekselman, I., et al., *The ribosomal protein ul22 modulates the shape of the protein exit tunnel.* Structure, 2017. **25**(8): p. 1233-1241.
218. Hansen, J.L., et al., *Structural insights into peptide bond formation.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(18): p. 11670-11675.
219. Arango, D., et al., *Direct epitranscriptomic regulation of mammalian translation initiation through n4-acetylcytidine.* Molecular Cell, 2022. **82**(15): p. 2797-2814.
220. Brown, A., et al., *Structural basis for stop codon recognition in eukaryotes.* Nature, 2015. **524**(7566): p. 493-496.
221. Shao, S.C., et al., *Structure and assembly pathway of the ribosome quality control complex.* Molecular Cell, 2015. **57**(3): p. 433-444.
222. Li, W.F., et al., *Selective inhibition of human translation termination by a drug-like compound.* Nature Communications, 2020. **11**(1).
223. Matheisl, S., et al., *Structure of a human translation termination complex.* Nucleic Acids Research, 2015. **43**(18): p. 8615-8626.
224. Koliopoulos, M.G., et al., *Functional role of trim e3 ligase oligomerization and regulation of catalytic activity.* Embo Journal, 2016. **35**(11): p. 1204-1218.
225. Teng, F., et al., *Cryo-em structure of the klhl22 e3 ligase bound to an oligomeric metabolic enzyme.* Structure (London, England : 1993), 2023.

226. Balaji, V. and T. Hoppe, *Regulation of e3 ubiquitin ligases by homotypic and heterotypic assembly.* F1000Research, 2020. **9**.
227. Motshwene, P.G., et al., *An oligomeric signaling platform formed by the toll-like receptor signal transducers myd88 and irak-4.* Journal of Biological Chemistry, 2009. **284**(37): p. 25404-25411.
228. Latty, S.L., et al., *Activation of toll-like receptors nucleates assembly of the myddosome signaling hub.* Elife, 2018. **7**.