IN SILICO AND *IN VIVO* ANALYSIS OF OTU DEUBIQUITINASES OTUB1, OTUB2 AND OTULIN PROTEIN-PROTEIN INTERACTIONS

NUR WAHIDA BINTI ZULKIFLI

UNIVERSITI SAINS MALAYSIA

2023

IN SILICO AND *IN VIVO* ANALYSIS OF OTU DEUBIQUITINASES OTUB1, OTUB2 AND OTULIN PROTEIN-PROTEIN INTERACTIONS

by

NUR WAHIDA BINTI ZULKIFLI

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

September 2023

ACKNOWLEDGEMENT

First and foremost, I would like to offer this endeavour to our God Almighty for the wisdom He bestowed upon me, along with strength, peace of mind and good health that allowed me to complete this research. I would like to express my highest gratitude and heartiest thanks to my project supervisor, Ts. Dr. Nurulisa binti Zulkifle, for her infinite dedication, continuous support, valuable guidance, encouragement and patience in guiding me towards the completion of this long and challenging journey. She motivates me whenever I am down either during my research work or the writing process, and she is also willing to share her knowledge and expertise in this study. In addition, I would like to extend my warmest appreciation to my co-supervisor, Assoc. Prof. Dr. Hasni bin Arsad for his generosity and knowledge throughout the project, especially in computational technique. I am so grateful to them and may God Almighty bless them.

My heartfelt thanks to my beloved mom and dad for their unconditional love, understanding, compassion, financial support and prayers at every step of the way that I take in my life, who have always been there for me through the ups and downs and believed I could succeed. They also extended their support morally and emotionally. I would like to thank my siblings for always being there for me, especially in 2022, when we were going through a sad and tough time. They inspired me to pursue my own destiny and explore new avenues. This journey would not have been possible without them thus I dedicate this accomplishment to my entire family.

Last but not least, I would like to acknowledge the Ministry of Higher Education, Malaysia for supporting this project financially under their Fundamental Research Grant Scheme (FRGS) and for sponsoring my PhD study under MyBrain scholarship program. Apart from this, my deepest gratitude to my dear lab mates and friends for their encouragement, advice and valuable opinion, who helped me survive all the stress when I encountered difficulties. All praise to the Almighty for blessing me with supportive and kind people around me. I would like to thank all the staff members of the Advanced Medical and Dental Institute, particularly the Biomedical Science Department, for their assistance and services that made my research successful. I am very grateful to have met all of these individuals, who give me the strength to complete this study. This accomplishment would not have been possible without their contribution. Thank you very much.

TABLE OF CONTENTS

ACKNOWLEDGEMENT ..... **ii**

TABLE OF CONTENTS ..... **iv**

LIST OF TABLES ..... **x**

LIST OF FIGURES ..... **xiii**

LIST OF SYMBOLS ..... **xviii**

LIST OF ABBREVIATIONS ..... **xx**

LIST OF APPENDICES ..... **xxv**

ABSTRAK ..... **xxvi**

ABSTRACT ..... **xxviii**

CHAPTER 1 INTRODUCTION **1**

1.1 Background of study 1

1.2 Problem statement 3

1.3 Research aims and objectives 4

1.3.1 Specific objectives 4

1.4 Thesis layout 5

CHAPTER 2 LITERATURE REVIEW **6**

2.1 The ubiquitin system 6

2.1.1 History of ubiquitin 6

2.1.2 The ubiquitin molecules 7

2.1.3 Ubiquitin enzymatic cascade 8

2.1.4 Reversing ubiquitination 10

2.2 Deubiquitinating enzymes 12

2.2.1 Classification of DUBs 12

2.2.1(a) OTU deubiquitinase 13

2.2.1(b) Sequence alignment of OTUB1, OTUB2 and OTULIN 16

2.2.2 DUBs' implication on health and diseases 17

2.3 Protein-protein interaction 23

2.3.1 Methods for studying PPI 25

2.3.1(a) Detection of PPI 25

2.3.1(b) Characterisation of PPI 29

2.3.2 PPIs' implication on diseases 31

2.4 Human interactome and ubiquitome 32

CHAPTER 3 LANDSCAPE OF DEUBIQUITINATING ENZYME PROTEOME AND UBIQUITOME 34

3.1 Introduction 34

3.2 Materials and methods 34

3.2.1 Construction of DUB PPI network 34

3.2.2 Network and node analyses 35

3.2.3 Venn diagram for assessing DUB target specificity 35

3.2.4 Extraction of cancer and essential protein subnetworks from DUB interactome 36

3.2.5 Validation of network randomness 36

3.3 Result 36

3.3.1 List of DUBs 36

3.3.2 Network properties of DUBs interactome 37

3.3.3 Analysis of DUBs target proteins 40

3.3.3(a) Target specificity based on DUBs family 40

3.3.3(b) Ubiquitin components in DUBs interactome 41

3.3.3(c) Cancer protein subnetwork 44

3.4 Discussion 52

3.5 Conclusion 58

CHAPTER 4 CONSTRUCTION OF YEAST TWO-HYBRID BAIT AND PREY VECTORS 59

4.1 Introduction 59

4.2 Overview of methodology 60

4.3 Materials and methods 61

4.3.1 Materials 61

4.3.1(a) Equipment 61

4.3.1(b) Glassware and plasticware 61

4.3.1(c) Cell line and growth medium 62

4.3.1(d) Molecular biology reagent and resource 62

4.3.2 Methods 63

4.3.2(a) General methods 63

4.3.2(b) Generation of Y2H bait vector containing OTUB1, OTUB2 and OTULIN genes 66

4.3.2(c) Generation of Y2H prey vector containing cDNA library 73

4.3.2(d) Generation of Y2H prey vector containing TP53, UBE2D2, SIAH1, FAM168A and DAZAP2 genes 85

4.4 Result 89

4.4.1 Y2H bait vector containing OTUB1, OTUB2 and OTULIN 89

4.4.1(a) PCR-amplification of OTUB1, OTUB2 and OTULIN 89

4.4.1(b) TOPO cloning of PCR-amplified OTUB1, OTUB2 and OTULIN 90

4.4.1(c) Shuttling pENTR-OTUB1, pENTR-OTUB2 and pENTR-OTULIN into pDEST32 92

4.4.2 Y2H prey vector containing cDNA library 94

4.4.2(a) Synthesis of cDNA from HEK293 94

4.4.2(b) Establishment of cDNA library entry clones 96

4.4.2(c) Transferring pENTR-cDNA into pDEST22 101

4.4.3 Y2H prey vector containing TP53, UBE2D2, SIAH1, FAM168A and DAZAP2 106

4.4.3(a) PCR-amplification of TP53, UBE2D2, SIAH1, FAM168A and DAZAP2 106

4.4.3(b) TOPO cloning of PCR-amplified TP53, UBE2D2, SIAH1, FAM168A and DAZAP2 108

4.4.3(c) Shuttling pENTR-TP53, pENTR-UBE2D2, pENTR-SIAH1, pENTR-FAM168A and pENTR-DAZAP2 into pDEST22 110

4.5 Discussion 113

4.6 Conclusion 118

CHAPTER 5 YEAST TWO-HYBRID SCREENING OF OTUB1, OTUB2 AND OTULIN INTERACTION PARTNERS 119

5.1 Introduction 119

5.2 Overview of methodology 120

5.3 Materials and methods 121

5.3.1 Materials 121

5.3.1(a) Equipment 121

5.3.1(b) Glassware and plasticware 121

5.3.1(c) Yeast cell, selection growth medium and agar 121

5.3.1(d) Yeast two-hybrid reagents 122

5.3.2 Methods 122

5.3.2(a) General methods 122

5.3.2(b) Small-scale yeast transformation 125

5.3.2(c) Optimisation of His plate selection 129

5.3.2(d) Library-scale yeast transformation and Y2H screening 130

5.4 Result 134

5.4.1 3AT concentration for His3 inhibition 134

5.4.2 Interaction of control plasmids 136

5.4.3 Verification of MaV203 containing test bait plasmids 142

5.4.4 Library-scale yeast transformation 143

5.4.5 Reporter genes activation by transformants containing interaction partners 148

5.4.6 Isolation of prey plasmids containing potential interactors 163

5.4.7 Sequence-verified of prey plasmid DNA 167

5.5 Discussion 169

5.6 Conclusion 176

CHAPTER 6 STRUCTURE PREDICTION OF OTUB1 IN COMPLEX WITH PUTATIVE PARTNERS 177

6.1 Introduction 177

6.2 Materials and methods 178

6.2.1 Protein secondary structure analysis 178

6.2.2 Selection of protein's 3D crystal structure 178

6.2.3 Homology modelling of MKRN3 and ESD 178

6.2.4 Docking of OTUB1 with Y2H-identified partners 179

6.2.4(a) Preparation of 3D protein 179

6.2.4(b) Protein-protein docking simulation 179

6.3 Result 180

6.3.1 Protein secondary structure prediction 180

6.3.2 Protein 3D structure 186

6.3.3 Homology modelling of MKRN3 and ESD 188

6.3.4 Protein-protein docking 190

6.4 Discussion 194

6.5 Conclusion 199

CHAPTER 7 CONCLUDING REMARKS 200

7.1 OTUB1 protein interaction 200

7.2 Is OTUB1 inhibitor possible? 204

7.3 Future prospective 206

REFERENCES 208

APPENDICES

LIST OF PUBLICATIONS

LIST OF TABLES

Page

Table 3.1 Network and node analyses of the DUBs interactome.....39

Table 3.2 Proteins that are targeted by DUBs from four different families.....41

Table 3.3 Top 10 proteins in cancer protein subnetwork based on degree, BC and CC measure. 46

Table 3.4 Summary of network properties for DUB interactome, cancer and essential protein subnetworks. 50

Table 3.5 Comparison of the topological properties between cancer protein subnetwork and random networks. 51

Table 4.1 Antibiotic marker for every vector.....64

Table 4.2 List of gene-specific (a) and vector-specific (b) primers. 67

Table 4.3 Reaction mixture and cycling parameters of *Taq* PCR for amplifying OTUB1, OTUB2 and OTULIN. 67

Table 4.4 Reaction mixture and cycling parameters of *Taq* colony PCR for verifying positive colonies containing pENTR-OTUB1, pENTR-OTUB2 and pENTR-OTULIN. 69

Table 4.5 Reaction mixture for LR recombinant reaction to generate bait expression vector pEXP32-OTUB1, pEXP32-OTUB2 and pEXP32-OTULIN, and prey expression vector pEXP22-TP53, pEXP22-UBE2D2, pEXP22-SIAH1, pEXP22-FAM168A and pEXP22-DAZAP2. 71

Table 4.6 Reaction mixture and cycling parameters of *Taq* colony PCR for verifying all expression clones and cDNA library entry clones. 73

Table 4.7 Reagent mixture for cDNA synthesis. 78

Table 4.8 Reaction mixture for BP recombinant reaction to generate pENTR-cDNA. 79

Table 4.9 Aliquots of BP and LR reaction samples and controls. 80

Table 4.10 Plating assay dilutions. 81

Table 4.11 Reaction mixture for LR recombinant reaction to generate pEXP22-cDNA. 83

Table 4.12 Reaction mixture and cycling parameters of *Taq* PCR for amplifying TP53, UBE2D2, SIAH1, FAM168A and DAZAP2..... 87

Table 4.13 Reaction mixture and cycling parameters of *Taq* colony PCR for verifying positive colonies containing pENTR-TP53, pENTR-UBE2D2, pENTR-SIAH1, pENTR-FAM168A and pENTR-DAZAP2. 88

Table 4.14 Concentration and purity of entry vectors pENTR-OTUB1, pENTR-OTUB2 and pENTR-OTULIN. 92

Table 4.15 Concentration and purity of expression vectors pEXP32-OTUB1, pEXP32-OTUB2 and pEXP32-OTULIN. 93

Table 4.16 The cfu of pENTR-cDNA, BP negative and BP positive controls. ... 98

Table 4.17 Qualitative and quantitative assessments of cDNA library entry and expression clones. 101

Table 4.18 The cfu of pEXP22-cDNA, LR negative and LR positive controls. 103

Table 4.19 The concentration and purity of the cDNA library expression vectors. 105

Table 4.20 Concentration and purity of entry vectors pENTR-TP53, pENTR-UBE2D2, pENTR-SIAH1, pENTR-FAM168A and pENTR-DAZAP2. 110

Table 4.21 Concentration and purity of expression vectors pEXP22-TP53, pEXP22-UBE2D2, pEXP22-SIAH1, pEXP22-FAM168A and pEXP22-DAZAP2. 112

Table 5.1 Amino acid preparation ingredients. 123

Table 5.2 List of ingredients used to make buffers. 123

Table 5.3 List of amino acids in SD broth and agar. 125

Table 5.4 Y2H set of control colonies.....127

Table 5.5 List of gene-specific (a) and vector-specific (b) primers. 131

Table 5.6 Reaction mixture and cycling parameters of *Taq* colony PCR for amplifying OTUB1, OTUB2 and OTULIN. 131

Table 5.7 Reaction mixture and cycling parameters of *Taq* colony PCR for verifying prey clones DNA. 134

Table 5.8 A summary of reporter gene phenotypes on four different selective media for every Y2H control plasmids. 141

Table 5.9 Reporter gene phenotypes on four different selective media of pEXP32-OTUB1, pEXP32-OTUB2 and pEXP32-OTULIN with pEXP22-cDNA interaction, respectively. 161

Table 5.10 Purification of prey plasmid from yeast cell, insert size and purification of prey plasmid from *E. coli*. 164

Table 5.11 BLAST result of the prey plasmid nucleotide sequence. 168

Table 6.1 Structural information of OTUB1 and its protein partners. 180

Table 6.2 The HDOCK docking results for OTUB1 in complex ubiquitin with LDHB, FOXM1, CENPU, MKRN3, ESD and POU2F1, respectively. 191

Table 6.3 Binding residues of OTUB1/Ub:UBE2D2 (Juang et al., 2012) and OTUB1 with Y2H-identifed partners LDHB, FOXM1, CENPU, MKRN3, ESD and POU2F1 from this study. 193

LIST OF FIGURES

Page

Figure 2.1 Ribbon diagram of human ubiquitin with a C-terminal Gly residue (labelled as C-term), Lys residue (labelled as Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63), and an N-terminal Met1 residue (labelled as N-term).....8

Figure 2.2 The simplified ubiquitination cascade and the diversity of ubiquitinated products.....11

Figure 2.3 Sequence alignment of OTUB1, OTUB2 and OTULIN.....17

Figure 2.4 Overview of the different roles of several DUBs in cancer.....22

Figure 2.5 A simple representation of a PPI network.....25

Figure 3.1 Protein-protein interaction network of human DUBs.....38

Figure 3.2 Subnetworks (or modules) of the top five most connected DUBs targets.....39

Figure 3.3 Specificity distribution of each DUB family.....40

Figure 3.4 DUB-E3 interactions.....42

Figure 3.5 A: DUB-E2 interactions. B: DUB-DUB interactions.....43

Figure 3.6 Subnetwork of cancer proteins in DUBs interactome.....45

Figure 3.7 Nodes in cancer protein subnetwork with the strongest characteristics based on plot of CC against BC.....47

Figure 3.8 Essential protein subnetwork in DUBs interactome.....49

Figure 4.1 Flowchart of Y2H vectors construction.....60

Figure 4.2 Agarose gel electrophoresis of OTUB1, OTUB2 and OTULIN amplification by PCR using gene-specific primers.....89

Figure 4.3 Agarose gel electrophoresis of entry clones containing OTUB1, OTUB2 and OTULIN genes, verified by colony PCR using gene-

specific primers for OTUB1 and OTUB2, while vector-specific primers for OTULIN. 91

Figure 4.4 Agarose gel electrophoresis of expression clones containing OTUB1, OTUB2 and OTULIN genes, verified by colony PCR using pDEST32 vector-specific primers pDEST32-F and pDEST32-R. 93

Figure 4.5 Agarose gel electrophoresis of total RNA purified by RNeasy Plus Mini Kit from HEK293 cells. 95

Figure 4.6 Agarose gel electrophoresis of poly A^+ mRNA isolated from total RNA. 95

Figure 4.7 The transformation plates of pENTR-cDNA, BP negative and BP positive controls on the LB agar containing 50 µg/ml kanamycin. ... 97

Figure 4.8 Agarose gel electrophoresis of cDNA library entry clones, pENTR-cDNA qualification by colony PCR using pDONR222 vector-specific primers M13F (-20) and M13R-pUC (-26). 100

Figure 4.9 The transformation plates of pEXP22-cDNA, LR negative and LR positive controls on the LB agar containing 100 µg/ml ampicillin.. 102

Figure 4.10 Agarose gel electrophoresis of cDNA library expression clones, pEXP22-cDNA qualification by colony PCR using pDEST22 vector-specific primers pDEST22-F and pDEST22-R. 105

Figure 4.11 Agarose gel electrophoresis of total RNA extracted by TRIzol from HEK293 cells. 106

Figure 4.12 Agarose gel electrophoresis of TP53, UBE2D2, SIAH1, FAM168A and DAZAP2, PCR-amplified using gene-specific primers listed in Table 4.2(a). 107

Figure 4.13 Agarose gel electrophoresis of entry clones containing TP53, UBE2D2, SIAH1, FAM168A and DAZAP2 genes, verified by colony PCR using gene-specific primers as listed in Table 4.2(a). . 109

Figure 4.14 Agarose gel electrophoresis of expression clones containing TP53, UBE2D2, SIAH1, FAM168A and DAZAP2 genes, verified by

colony PCR using pDEST22 vector-specific primers pDEST22-F and pDEST22-R. 111

Figure 4.15 The map of pDEST32 and pDETS22 empty vectors. 113

Figure 4.16 The specificity of *att* sites for site specific recombination reactions.. 114

Figure 5.1 Flowchart of Y2H screening. 120

Figure 5.2 The colonies of transformants containing test and control plasmids were patched on -LW agar plate. 128

Figure 5.3 Determining 3AT sensitivity. 135

Figure 5.4 Transformants characterisation and X-gal assay of Y2H prey vector containing UBE2D2 gene with empty pDEST32 (C7) and pEXP32-OTUB1 (C8), respectively. 137

Figure 5.5 Transformants characterisation and X-gal assay of Y2H prey vector containing SIAH1 gene with empty pDEST32 (C7) and pEXP32-OTUB2 (C8), respectively. 138

Figure 5.6 Transformants characterisation and X-gal assay of Y2H prey vector containing FAM168A gene with empty pDEST32 (C7) and pEXP32-OTUB2 (C8), respectively. 139

Figure 5.7 Transformants characterisation and X-gal assay of Y2H prey vector containing DAZAP2 gene with empty pDEST32 (C7) and pEXP32-OTULIN (C8), respectively. 140

Figure 5.8 Agarose gel electrophoresis of bait expression clones containing OTUB1, OTUB2 and OTULIN genes, verified by colony PCR. 142

Figure 5.9 Library-scale yeast transformation of pEXP32-OTUB1 and pEXP22-cDNA. 145

Figure 5.10 Library-scale yeast transformation of pEXP32-OTUB2 and pEXP22-cDNA. 146

Figure 5.11 Library-scale yeast transformation of pEXP32-OTULIN and pEXP22-cDNA. 147

Figure 5.12 Transformants characterisation and X-gal assay of pEXP32-OTUB1 and pEXP22-cDNA-a. 149

Figure 5.13 Transformants characterisation and X-gal assay of pEXP32-OTUB1 and pEXP22-cDNA-b. 150

Figure 5.14 Transformants characterisation and X-gal assay of pEXP32-OTUB1 and pEXP22-cDNA-c. 151

Figure 5.15 Transformants characterisation and X-gal assay of pEXP32-OTUB1 and pEXP22-cDNA-d. 152

Figure 5.16 Transformants characterisation and X-gal assay of pEXP32-OTUB1 and pEXP22-cDNA-e. 153

Figure 5.17 Transformants characterisation and X-gal assay of pEXP32-OTUB1 and pEXP22-cDNA-f. 154

Figure 5.18 Transformants characterisation and X-gal assay of pEXP32-OTUB1 and pEXP22-cDNA-g. 155

Figure 5.19 Transformants characterisation and X-gal assay of pEXP32-OTUB1 and pEXP22-cDNA-h. 156

Figure 5.20 Transformants characterisation and X-gal assay of pEXP32-OTUB2 and pEXP22-cDNA-a. 157

Figure 5.21 Transformants characterisation and X-gal assay of pEXP32-OTUB2 and pEXP22-cDNA-b. 158

Figure 5.22 Transformants characterisation and X-gal assay of pEXP32-OTUB2 and pEXP22-cDNA-c. 159

Figure 5.23 Transformants characterisation and X-gal assay of pEXP32-OTULIN and pEXP22-cDNA-a. 160

Figure 5.24 Agarose gel electrophoresis of prey plasmid DNA, verified by colony PCR. 166

Figure 5.25 The construction of two hybrid proteins and expression of reporter gene. 170

Figure 6.1 Secondary structure prediction of OTUB1 and LDHB. 181

Figure 6.2 Secondary structure prediction of FOXM1. 182
Figure 6.3 Secondary structure prediction of CENPU. 183
Figure 6.4 Secondary structure prediction of MKRN3 and ESD. 184
Figure 6.5 Secondary structure prediction of POU2F1. 185
Figure 6.6 The 3D structure of proteins from PDB and AlphaFold. 187
Figure 6.7 3D structure models of A: MKRN3 and B: ESD. 188
Figure 6.8 The overall and local model quality by ProSA evaluation analysis based on Z-score and plot of residue energies, respectively. 189
Figure 6.9 Ramachandran plot showing residues in the favoured (red), allowed (yellow), generously allowed (light yellow) and disallowed (white) regions. 189
Figure 6.10 The predicted structural docking conformation of the protein-protein complexes. 192
Figure 6.11 Structure of OTUB1 in complex with UBE2D2 in the presence of two ubiquitin molecules from Juang et al., (2012). 195
Figure 7.1 OTUB1 PPI. 203
Figure 7.2 OTUB1-interacting proteins but also targeted by other DUBs. 204
Figure 7.3 OTUB1/USP8 dual inhibitor ‘61’ (Tan et al., 2022). 205

LIST OF SYMBOLS

α	alpha
Å	angstrom
~	approximately
*	asterisk
β	beta
ΔG	binding energy
CO_2	carbon dioxide
cfu	colony-forming unit
℃	degree Celcius
K_D	dissociation constant
ΔH	enthalpy
ΔS	entropy
g	gram
∞	infinity
kcal/mol	kilocalorie per mole
kV	kilovolt
λ	lambda
L	liter
m	meter
µF	microfarad
µg	microgram
µl	microliter
µm	micrometer
µM	micromolar
mg	milligram
ml	milliliter
mm	millimeter
mM	millimolar
M	molar
ng	nanogram
nm	nanometer

nM	nanomolar
Ω	ohm
%	percentage
V	volt
v/v	volume/volume
w/v	weight/volume

LIST OF ABBREVIATIONS

AD	activation domain
AVT	active compound acevaltrate
AHS	adenine hemisulfate salt
ATP	adenosine triphosphate
AP-MS	affinity purification coupled with mass spectrometry
NH4OAc	ammonium acetate
APC/C	anaphase-promoting complex/cyclosome
AURKB	aurora kinase B
BIRC5	baculoviral IAP repeat containing 5
bp	base pairs
β-ME	beta-mercaptoethanol
BC	betweenness centrality
OBB	binding buffer
X-gal	bromo-4-chloro-3-indolyl-β-D-galactopyranoside
COSMIC	Catalogue Of Somatic Mutations In Cancer
CDC5L	cell division cycle 5 like
CENPU	centromere protein U
CFTR	CF transmembrane conductance regulator
CC	closeness centrality
CompPASS	Comparative Proteomic Analysis Software Suite
CUL	cullin
CCN	cyclin
CDKN2A	cyclin dependent kinase inhibitor 2A
Cys	cysteine
DAZAP2	DAZ-associated protein 2
CUL4-VPRBP	DDB1 and CUL4 associated factor 1
DNA	deoxyribonucleic acid
DNASSST	deoxyribonucleic acid, single stranded from salmon testes
DUB	deubiquitinating enzyme/deubiquitinase
DMSO	dimethyl sulfoxide
DSV	discovery studio visualizer

DISC1	DISC1 scaffold protein
DMWD	DM1 locus, WD repeat containing
DBD	DNA binding domain
DMEM	Dulbecco's Modified Eagle's Medium
QF	elution buffer
E. coli	*Escherichia coli*
EGFR	epidermal growth factor receptor
EMT	epithelial-mesenchymal transition
QBT	equilibrium buffer
ESD	esterase D
ESR1	estrogen receptor 1
EtBr	ethidium bromide
EDTA	ethylenediaminetetraacetic acid calcium disodium salt
EWSR1	EWS RNA binding protein 1
FAM168A	family with sequence similarity 168 member A
FBXW7	F-box and WD repeat domain containing 7
FBS	fetal bovine serum
FOX	forkhead box
GATA3	GATA binding protein 3
Gly	glycine
IC_{50}	half maximal inhibitory concentration
His	histidine
HEK293	human embryonic kidney 293
HER-2	human epidermal growth factor receptor 2
HCl	hydrogen chloride
MINT	IntAct Molecular INTeraction
IDT	Integrated DNA Technologies
IFN	interferon
IMEx	International Molecular Exchage
JAMMs	JAMM/MPN associated metalloproteases
JUN	Jun proto-oncogene, AP-1 transcription factor subunit
kDa	kilodalton
LDHB	lactate dehydrogenase B
Leu	leucine

LUBAC	linear ubiquitin chain assembly complex
LiAc	lithium acetate dihydrate
Lys	lysine
LB	lysogeny broth
MJDs	Machado-Joseph-disease proteases
$MgSO_4.7H_2O$	magnesium sulfate heptahydrate
MKRN3	makorin ring finger protein 3
MS	mass spectrometry
MDM2	MDM2 proto-oncogene
Met1	methionine
monoUb	monoubiquitination
MINDYs	motifs interacting with ubiquitin-containing novel DUB family
multi-monoUb	multi-monoubiquitination
NCBI	National Center for Biotechnology Information
DMF	N,N-dimethyl formamide
NSCLC	non-small cell lung cancer
NF-κB	nuclear factor kappa B
OGEE	Online Gene Essentiality
OD	optical density
OTUB1	OTU deubiquitinase, ubiquitin aldehyde binding 1
OTUB2	OTU deubiquitinase, ubiquitin aldehyde binding 2
OTULIN	OTU deubiquitinase with linear linkage specificity
ORAS	OTULIN-Related Autoinflammation Syndrome
OTUs	ovarian tumour proteases
Parkin	parkin RBR E3 ubiquitin protein ligase
PenStrep	Penicillin-Streptomycin
PGAM5	PGAM family member 5, mitochondrial serine/threonine protein phosphatase
PBS	phosphate buffered saline
PI3K/Akt	phosphatidylinositol 3-kinase/protein kinase B
P2/B2	plasmid lysis buffer
P3/B3	plasmid neutralisation buffer
P1/B1	plasmid resuspension buffer
PEG-3,350	poly(ethylene glycol) bioxtra, average mol wt 3,350
PCR	polymerase chain reaction

polyUb	polyubiquitination
KCTD10	potassium channel tetramerization domain containing 10
KCl	potassium chloride
POU2F1	POU class 2 homeobox 1
PPI	protein-protein interaction
PDB	protein data bank
ProSA	protein structure analysis
PSICQUIC	Proteomics Standard Initiative Common QUery InterfaCe
RAD23A	RAD23 homolog A, nucleotide excision repair protein
RAD23B	RAD23 homolog B, nucleotide excision repair protein
RING	really interesting new gene
Treg	regulatory T
RNA	ribonucleic acid
RNF168	ring finger protein 168
S. cerevisiae	*Saccharomyces cerevisiae*
SIAH1	siah E3 ubiquitin protein ligase 1
SCF-FBXO31	SKP1-cullin-F-box only protein 31
SMAD2/3	SMAD family member 2/3
SNAI2	snail family transcriptional repressor 2
SNW1	SNW domain containing 1
NaOAc	sodium acetate
NaCl	sodium chloride
NaOH	sodium hydroxide
$Na_2HPO_4.7H_2O$	sodium phosphate dibasic heptahydrate
$NaH_2PO_4.H_2O$	sodium phosphate monobasic monohydrate
SILAC	stable isotope labelling by amino acids in cell culture
S.O.C	Super Optimal broth with Catabolite repression
SD	Synthetic Dropout
3D	three-dimensional
TAB	TGF-beta activated kinase 1 (MAP3K7) binding protein
TGF-β	transforming growth factor-beta
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
TEN	Tris-EDTA-NaCl

Tris	trizma base
Trp	tryptophan
TRAF	tumour necrosis factor receptor-associated factor
TP53/p53	tumour protein 53
p63	tumour protein 63
p73	tumour protein 73
YWHAG	tyrosine 3- monooxygenase/tryptophan 5-monooxygenase activation protein gamma
Ub	ubiquitin
E1	ubiquitin-activating enzyme
UBD	ubiquitin binding domain
UCHs	ubiquitin carboxy-terminal hydrolases
E2	ubiquitin-conjugating enzyme
UBE2	ubiquitin conjugating enzyme E2
E3	ubiquitin ligase
USPs	ubiquitin-specific proteases
UniProt	UNIversal PROTein
UAS	upstream activating sequence
Ura	uracil
VEGFA	vascular endothelial growth factor A
QC/OW2	wash buffer
Wnt	Wingless-related integration site
YNBWAA	yeast nitrogen base without amino acids
Y1H	yeast one-hybrid
YPAD	yeast peptone adenine dextrose
YPD	yeast peptone dextrose
Y3H	yeast three-hybrid
Y2H	yeast two-hybrid
YSDOMS	yeast synthetic drop-out medium supplements
ZUFSPs	zinc finger with UFM1 specific peptidases
3AT	3-Amino-1,2,4-triazole
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
5FOA	5-fluoroorotic acid

LIST OF APPENDICES

Appendix A	A paper on understanding human deubiquitinases target specificity by network-based analysis towards their development as therapeutics target
Appendix B	List of DUBs

ANALISA INTERAKSI PROTEIN-PROTEIN OTU *DEUBIQUITINASE* OTUB1, OTUB2 DAN OTULIN SECARA *IN SILICO* DAN *IN VIVO*

ABSTRAK

Dalam era pasca genomik, data proteomik dan interaktomik adalah sumber penting untuk memahami asas molekul bagi kepelbagaian fungsi sel. Penyelidikan dalam bidang ini masih berkembang, dan setakat ini sekurang-kurangnya 90% protein telah beranotasi. Untuk melengkapkan peta proteom dan interaktom, banyak interaksi lain yang perlu dikenal pasti. Oleh itu, kajian ini bertujuan untuk menyumbang pada pengembangan data interaktom manusia. Dalam kajian ini memfokuskan pada interaksi protein-protein (PPI) bagi enzim *deubiquitinating* (DUBs) dengan menggunakan beberapa pendekatan *in silico* dan *in vivo*. Pertama, Cytoscape versi 3.9.1 telah digunakan untuk membina interaktom DUBs dengan menggunakan data daripada pangkalan data IMEx dan dianalisis menggunakan pelbagai algoritma teori graf. Interaktom DUBs terdiri daripada 3,406 nod dan 4,982 *edge*, manakala subrangkaian protein kanser yang diekstrak daripada interaktom mendedahkan bahawa DUBs adalah penting dalam biologi kanser. Dengan ini kajian interaktom menyimpulkan bahawa, OTUB1 merupakan salah satu protein yang mempunyai ciri rangkain yang kuat, manakala OTUB2 mempunyai homolog yang hampir sama dengan OTUB1, dan OTULIN sebagai ahli keluarga yang terbaru, telah dipilih untuk penyaringan yis dua-hibrid (Y2H) dengan perpustakaan cDNA manusia yang diperolehi daripada HEK293. Vektor umpan dan mangsa perpustakaan cDNA dijana menggunakan teknologi *Gateway*, di mana gen diklon ke dalam vektor TOPO/pDONR222 dahulu sebelum diklon ke dalam vektor pDEST32/22. Penjujukan bagi vektor umpan telah disahkan dan vektor mangsa mempunyai kualiti dan kuantiti

yang mencukupi untuk mewakili cDNA manusia dengan jumlah klon 6.8×10^{6} cfu. Kedua-dua vektor telah ditranformasi ke dalam strain MaV203 *S. cerevisiae* untuk penyaringan Y2H dan mengenal pasti protein yang diandaikan berinteraksi. Penyaringan ini telah mengenal pasti LDHB, FOXM1, ESD, MKRN3, CENPU dan POU2F1 sebagai interaksi bagi OTUB1, di mana tiga yang pertama telah dilaporkan pada kajian lain. Untuk menganalisis antara muka yang menjadi pengikat di antara OTUB1 dan interaksi yang dikenal pasti dari Y2H, struktur 3D diperolehi sama ada daripada pangkalan data PDB atau dengan pemodelan homologi. Seterusnya, server HDOCK digunakan dalam kajian pengedokan molekul untuk meramalkan konformasi struktur OTUB1 yang berinteraksi dengan protein yang dikenal pasti dari Y2H. Keputusan mengesahkan bahawa OTUB1 berkedudukan pada konformasi aktif, di mana heliks αA, α5, α7 dan penghubung α9- α10 terlibat dalam pengikatan dengan protein yang dikenal pasti dari Y2H. Tambahan pula, kompleks OTUB1:LDHB mempunyai afiniti pengikatan yang tertinggi dengan ΔG = -23.7 kcal/mol dan K_D pada 25.0 ℃ = 4.1×10^{-18} M. Sebagai kesimpulan, analisis rangkaian PPI DUBs, pengesanan rakan kongsi interaksi OTUB1 yang novel dan pencirian kompleks OTUB1 dalam kajian ini telah memberikan gambaran yang jelas tentang PPI DUBs dan mewujudkan asas yang kukuh untuk kajian interaksi yang lebih banyak pada masa hadapan.

IN SILICO AND *IN VIVO* ANALYSIS OF OTU DEUBIQUITINASES OTUB1, OTUB2 AND OTULIN PROTEIN-PROTEIN INTERACTIONS

ABSTRACT

In this post-genomic era, proteomic and interactomic data are important sources for understanding the molecular basis of cell functional diversity. Research in these fields is still progressing, with at least 90% of proteins annotated so far and many more interactions to be uncovered to complete the proteome and interactome maps. The purpose of this study is to contribute to the expansion of human interactome data by focusing on the deubiquitinating enzymes (DUBs) protein-protein interaction (PPI) using several *in silico* and *in vivo* approaches. First, a DUBs interactome was built on Cytoscape ver 3.9.1 using data from IMEx database and analysed using various graph theory algorithms. The DUBs interactome consisted of 3,406 nodes and 4,982 edges, whereas a cancer protein subnetwork extracted from the interactome revealed that DUBs are significant in cancer biology. Concluding the interactome study, OTUB1, one of the proteins with strong network characteristics, along with its closest homologue OTUB2, and the newest family member OTULIN, were selected for yeast two-hybrid (Y2H) screening against human cDNA library derived from HEK293. Y2H bait and cDNA library prey vectors were generated using Gateway technology, first in donor vector TOPO/pDONR222 before being shuffled into destination vector pDEST32/22. The bait vector is sequence-verified, and the prey vector is of sufficient quality and quantity to represent human cDNA with total clones of 6.8×10^6 cfu. Both vectors were transformed into MaV203 *S. cerevisiae* strain for Y2H screening to identify putative interacting proteins. The screening identified LDHB, FOXM1, ESD, MKRN3, CENPU and POU2F1 as OTUB1 interactors, with the first three have already

been reported elsewhere. To analyse the binding interface between OTUB1 and its Y2H-identified interactors, the 3D structures were obtained either by directly retrieved from PDB database or by homology modelling. The HDOCK server was used for molecular docking studies to predict the structural conformation of OTUB1 in association with Y2H-identified interacting proteins. The results confirmed that OTUB1 is positioned at the intended active conformation, in which the helices αA, α5, α7 and the α9-α10 linker involved in binding with the Y2H-identified proteins. Furthermore, OTUB1:LDHB complex is observed to possess the highest binding affinity with ΔG = -23.7 kcal/mol and K_D at 25.0 ℃ = 4.1×10^{-18} M. To conclude, the analysis of DUBs PPI network, the detection of OTUB1's novel interaction partners and the characterisation of OTUB1 complexes in this study have given an illuminating insight on DUBs PPI and established a strong foundation for many more interaction studies in future.

CHAPTER 1

INTRODUCTION

1.1 Background of study

Understanding the complex relationship between a cell's genotype and phenotype is a major goal of molecular life science research. In order to attain this goal, omics research has grown in importance, starting with the Human Genome Project that attempted to sequence and map the genome across human cells. Upon conclusion of this project, the International Human Genome Sequencing Consortium reported approximately 20,000 protein-coding genes (Collins et al., 2004), a figure regarded to be lower than pre-genomic estimations and highlighting the fact that the organism's functional diversity is not exactly proportional to the number of protein-coding genes. In 2010, HUPO's Human Proteome Project was initiated with the intention of annotating all proteins encoded by the human genome. It is now in its penultimate stage, with 18,407 (93.2%) of the 19,750 predicted proteins coded in the human genome having been annotated (Omenn et al., 2022).

However, believing that cell phenotypic information can be deduced after mapping the entire human protein is erroneous due to the fact that proteins interact with one another to function. Identification of physical interactions between biological macromolecules, such as protein-DNA, protein-RNA and protein-protein, is essential for comprehending the function of gene products as well as the global organisation and interplay of various molecular machineries within the cell. Thus, the study of molecular interaction, termed 'interactomics', should be considered the pinnacle of omics research. Interactomics, particularly protein interaction studies, is a powerful tool for functional characterisation of proteins and provides a fundamental understanding of the whole proteome. Leveraging protein interactions can infer protein function because for

a protein with unknown function, its function can be determined by looking at its interaction partner with annotated function, a phrase dubbed ‘guilt by association’.

Currently, nearly 88% of proteins reviewed in UniProt/Swiss-Prot having interaction data (Dimitrakopoulos et al., 2022). With such a large amount of information, interactome research has shifted to focus on utilising protein-protein interaction (PPI) data for novel scientific discoveries such as drug development, while also identifying novel interaction for newly annotated protein. The study on interactomics is still ongoing, but it is now focusing on understanding the proteome and interactome of specific sets of conditions, for example cancer interactome or ubiquitin pathway interactome, also known as ubiquitome. Ubiquitin pathway is a post-translational modification of protein that governs practically every function in human cells. Since its inception in the 1980s, research in the ubiquitin field has advanced rapidly, and its major role in protein degradation, non-degradative functions and non-canonical interactions among its components have been extensively explored. A term called ‘ubiquitomics’ was coined in 2007 to refer to the study of a set of proteins that are modified by ubiquitin (Tomlinson et al., 2007). The most recent ubiquitin component to be the subject of interest is the deubiquitinating enzymes (DUBs), the enzyme responsible for reversing the ubiquitination process. During the beginning of our study, there was only one large-scale study of DUBs interactome reported, which used MS-based analysis and a specifically built software, Comparative Proteomic Analysis Software Suite (CompPASS) to assign confidence measurements to interactions from parallel non-reciprocal proteomic datasets (Sowa et al., 2009). Another DUBs interaction landscape report that adopted MS-based chemoproteomic approach emerged in the middle of our study, which increased our understanding of human DUBs and its importance in cell regulation (Pinto-Fernández et al., 2019).

Since ubiquitination is essential for cell function, its deregulation is always associated with pathological conditions such as cancer, neurological diseases and immunological disorders, to name a few. This sparked a great deal of interest in investigating the therapeutics potential among the ubiquitination components including ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin ligase (E3), DUB and protease. The E3 and proteases have long been the subject of pharmaceutical intervention, but little progress has been made in developing DUB inhibitors, possibly due to a lack of understanding of DUB interactions. Inhibiting an enzyme, particularly one engaged in many substrates, may have unintended consequences for other pathways regulated by that enzyme. In fact, many of the early generation DUB inhibitors were discovered to be multitargeted in retrospect (Altun et al., 2011; Ritorto et al., 2014). Thus, a complete understanding of that potential DUB's interaction network is crucial to avoid perturbation of other cellular functions. More importantly, structural information is also critical, since three-dimensional (3D) structure plays an important role in understanding protein activities, determining the active sites and studying binding interfaces. This kind of analysis provides valuable information for drug design studies.

1.2 Problem statement

In between the two large-scale studies of DUBs interactome, there are also many individual studies that have identified physical interactions of certain DUBs and their interaction partners. A compilation of this information is required, along with a comprehensive analysis of the network topological behaviour, since a reference map of DUBs interactome could provide a deeper, more mechanistic understanding of cellular functions. Aside from the known interactions, it is also vital to increase the coverage of

the DUBs interactome, as it is the foundation to enable significant progress in mapping the complete human protein interactome. This has prompted a need to continue identifying novel interactions for DUBs, as well as assessing their 3D structure, not only towards achieving the completeness of human interaction map, but also to enable rapid development in disease therapeutic intervention targeting DUBs' PPI.

1.3 Research aims and objectives

The general objective of this study is to analyse, identify and characterise selected DUBs protein-protein interaction utilising *in silico* and *in vivo* approaches. We intend to expand the coverage of the DUBs interactome and gain insight for DUBs development as therapeutic drug for (1) the most understudied DUBs and (2) the most prominent DUBs with the most promising role in diseases. These aims will be achieved by the specific objectives below:

1.3.1 Specific objectives

1) Construct and analyse the DUBs known interactome using a network-based approach.
2) Establish Y2H bait and prey vectors using Gateway technology.
3) Screen for novel interaction partners of selected DUBs (OTUB1, OTUB2 and OTULIN) using Y2H.
4) Predict the structural conformation of OTUB1 in complex with interaction partners detected from Y2H screening.

1.4 Thesis layout

Chapter One contains an introduction to the work, a problem statement, research aims and objectives, and chapter descriptions.

Chapter Two summarises past research in this topic and provides a brief description of the proteins and methodologies employed throughout the thesis.

Chapter Three describes the construction and analysis of the DUBs interactome and its cancer protein subnetwork, with a goal of exploring the interaction pattern and key proteins and highlight the importance of DUBs in cancer biology.

Chapter Four presents the approach to establish yeast two-hybrid (Y2H)'s bait expression plasmids containing OTUB1, OTUB2 and OTULIN and prey expression plasmids containing cDNA library derived from HEK293 cells.

Chapter Five explains the Y2H screening between OTUB1, OTUB2 and OTULIN against human cDNA library.

Chapter Six describes the protein-protein docking simulation for predicting the structural conformation of OTUB1 in complex with Y2H-identified interacting partners.

Chapter Seven concludes the study's findings with future prospectives.

CHAPTER 2

LITERATURE REVIEW

2.1 The ubiquitin system

2.1.1 History of ubiquitin

In the past, the majority of intracellular proteins were considered to be immortal and long-lived. This notion remained unchallenged until the late 1970s, when Ciechanover et al., (1978) found a heat-stable, low-molecular-weight protein allegedly involved in ATP-dependent proteolysis. Two years later, the same research group discovered that the molecule, now designated as ubiquitin, degraded protein by covalently attaching to proteolytic substrates in a process called ubiquitination (Ciechanover et al., 1980). Hershko and colleagues then went on to identify and characterise the enzymes responsible for attaching ubiquitin to the target protein, named ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin ligase (E3) (Ciechanover et al., 1982; Hershko et al., 1983, 2000). Meanwhile, the protease responsible for the destruction of ubiquitin-tagged proteins was identified in the mid-1980s as a high-molecular-weight protein complex called the proteasome (Arrigo et al., 1988; Hough et al., 1986, 1987). Since these key discoveries, research into the function of ubiquitination in dynamic biological processes has accelerated. Ubiquitin, which debuted as just a protein degradation tag, is now thought to govern nearly every function in the cells (Hoeller and Dikic, 2009; Nakamura, 2018; Wang and Wang, 2021), including inflammatory response, endocytic trafficking, transcriptional regulation, signal transduction, and cell cycle progression (Ashida et al., 2014; Hershko and Ciechanover, 1998), hence maintaining cell homeostasis and life activities (Deng et al., 2020).

2.1.2 The ubiquitin molecules

Ubiquitin is a remarkably conserved small protein consisting of 76 amino acids, with a molecular mass around 8.5 kDa (Figure 2.1) (D'Arcy et al., 2015; Deák and Boros, 2015; Hegde et al., 2012; Xu and Jaffrey, 2013). The ubiquitin protein is found in all eukaryotes but appears to be lacking in prokaryotes, such as archaeal and eubacterial species (Hegde, 2010; Hochstrasser, 2000). Ubiquitin is encoded by four distinct genes: RPS27A, UBA52, UBC and UBB. The RPS27A and UBA52 are monomeric ubiquitin-ribosomal fusion genes that expressed one ubiquitin unit fused to a ribosomal protein, while UBC and UBB are polyubiquitin genes, which harbour 3-4 and 9-10 tandem repeats of ubiquitin coding units, respectively (Baker and Board, 1991; Finley et al., 1989; Wiborg et al., 1985).

Ubiquitin can occur either free or covalently conjugated to other substrate proteins in the cytoplasm and nucleus (Chau et al., 1989; Ubiquitination Cascade Pathway, 2016). Pool of free ubiquitin occurs with exposed diglycine (Gly-Gly) C-terminus tail whereas conjugated ubiquitin is covalently linked to substrate proteins via an isopeptide bond between its C-terminus and either substrate protein's N-terminus or the ε-amino group of a lysine (Lys) residue (Dye and Schulman, 2007). There are seven Lys residues that functioned as the conjugation sites among the 76 amino acids that constitute ubiquitin: Lys6, Lys11, Ly27, Lys29, Lys33, Lys48 and Lys63 (Peng et al., 2003). The ubiquitin protein adopted a compact globular fold of 5-stranded beta-sheet, a short 3_{10} helix, and a 3.5-turn alpha-helix. The majority of substrate proteins carry ubiquitin binding domains (UBD) that engage with the hydrophobic patch formed on the surface of ubiquitin's beta-sheet by Leu8, Ile44 and Val70 (Randles and Walters, 2018).

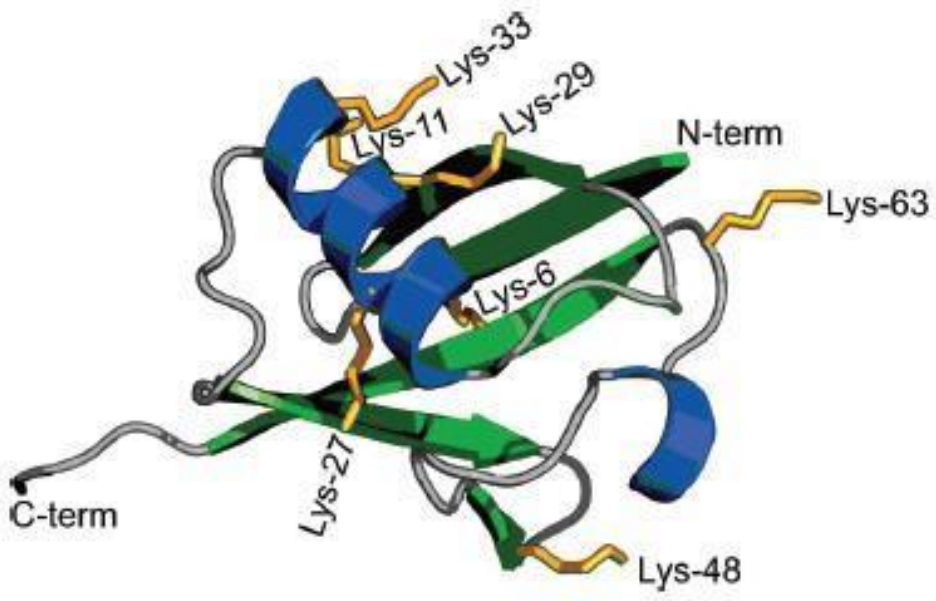

Figure 2.1 Ribbon diagram of human ubiquitin with a C-terminal Gly residue (labelled as C-term), Lys residue (labelled as Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63), and an N-terminal Met1 residue (labelled as N-term). The diagram was adapted from Deák and Boros, (2015) and modified.

2.1.3 Ubiquitin enzymatic cascade

Protein ubiquitination consists of three main steps: activation, conjugation and ligation, which are carried out by different enzymes named E1, E2 and E3 that work sequentially in a cascade. The human genome encodes two E1 (Uba1 and Uba6), around 40 E2s, and greater than 600 E3 enzymes (Deshaies and Joazeiro, 2009). During the ubiquitination process, E1 activates and adenylates the ubiquitin C-terminal Gly residue by adenosine triphosphate (ATP) hydrolysis, followed by the formation of a thioester bond between the ubiquitin C-terminal Gly residue and the E1 active site cysteine (Cys) residue. The activated ubiquitin is then transported from the E1 to the Cys residue of the E2 by forming a thioester bond in an ATP-dependent reaction. The thioester-linked ubiquitin is either delivered straight from the E2 to the substrate protein using E3 as a scaffold, or the E3 first forms a thioester bond with the ubiquitin before transferring the ubiquitin to the substrate protein (Figure 2.2(A)) (Ashida et al., 2014; Callis, 2014; D'Arcy et al., 2015).

The ubiquitin can be attached to the substrate protein through three different attachment modes: monoubiquitination (monoUb), multi-monoubiquitination (multi-monoUb) or polyubiquitination (polyUb) (Figure 2.2(B)) (Hegde et al., 2012; Stone, 2016; Ubiquitination Cascade Pathway, 2016). The conjugation of a single ubiquitin to the substrate protein is referred to as a monoUb, while the attachment of more than one ubiquitin at multiple sites of the same substrate protein leads to the formation of a multi-monoUb. The C-terminal Gly residue of the additional ubiquitin can be covalently linked to any of the seven Lys residues or N-terminal methionine (Met1) residue of the previous ubiquitin that was already linked on the substrate protein, forming a different linkage of polyubiquitin or a linear polyubiquitin chain, respectively (Asaoka and Ikeda, 2015; Callis, 2014; D'Arcy et al., 2015; Hegde, 2010; Stone, 2016). The formation of a mixed or branched polyubiquitin chain is the result of the attachment of two ubiquitins to two different Lys residues of the ubiquitin that is already attached to the substrate protein (Ye and Rape, 2009).

The protein that has been tagged with ubiquitin influences the structure and particular functions of the protein based on the ubiquitin chain topologies, such as altering the binding partners recruitment, protein interactions activity, trafficking, localisation, activity and stability (Eletr and Wilkinson, 2014; French et al., 2021; Pickart and Eddins, 2004; Stone, 2016), thereby regulating different cellular processes and determining the diversity of downstream consequences to the protein (Callis, 2014; Liao et al., 2022; Ye and Rape, 2009). Proteins ligated with Lys48-linked polyubiquitin chains are generally degraded by proteasome (Chau et al., 1989; Hershko and Ciechanover, 1998; Stone, 2016; Ye and Rape, 2009) while other types of ubiquitin chains are more typically associated with non-proteolytic functions such as Lys63 that is commonly destined in signal transduction, endocytosis and DNA repair pathways

(Deng et al., 2000; Haglund and Dikic, 2005; Hofmann and Pickart, 1999; Kawadler and Yang, 2006; Liao et al., 2022). However, Kim et al., (2007) and Saeki et al., (2009) demonstrated that the protein with Lys63-linked polyubiquitin chains can also function as a signal for proteasomal degradation.

2.1.4 Reversing ubiquitination

An important aspect of protein ubiquitination is its reversibility, mediated by the action of a specialised enzyme known as a deubiquitinase or deubiquitinating enzyme (DUB). DUBs either prevent ubiquitin from attaching to the target protein or cleave the isopeptide bond between the C-terminal Gly of ubiquitin and the Lys residue of the substrate protein, thus opposing E3 action and function (D'Arcy et al., 2015; Isono and Nagel, 2014; Komander et al., 2009; Liao et al., 2022; Turcu et al., 2009).

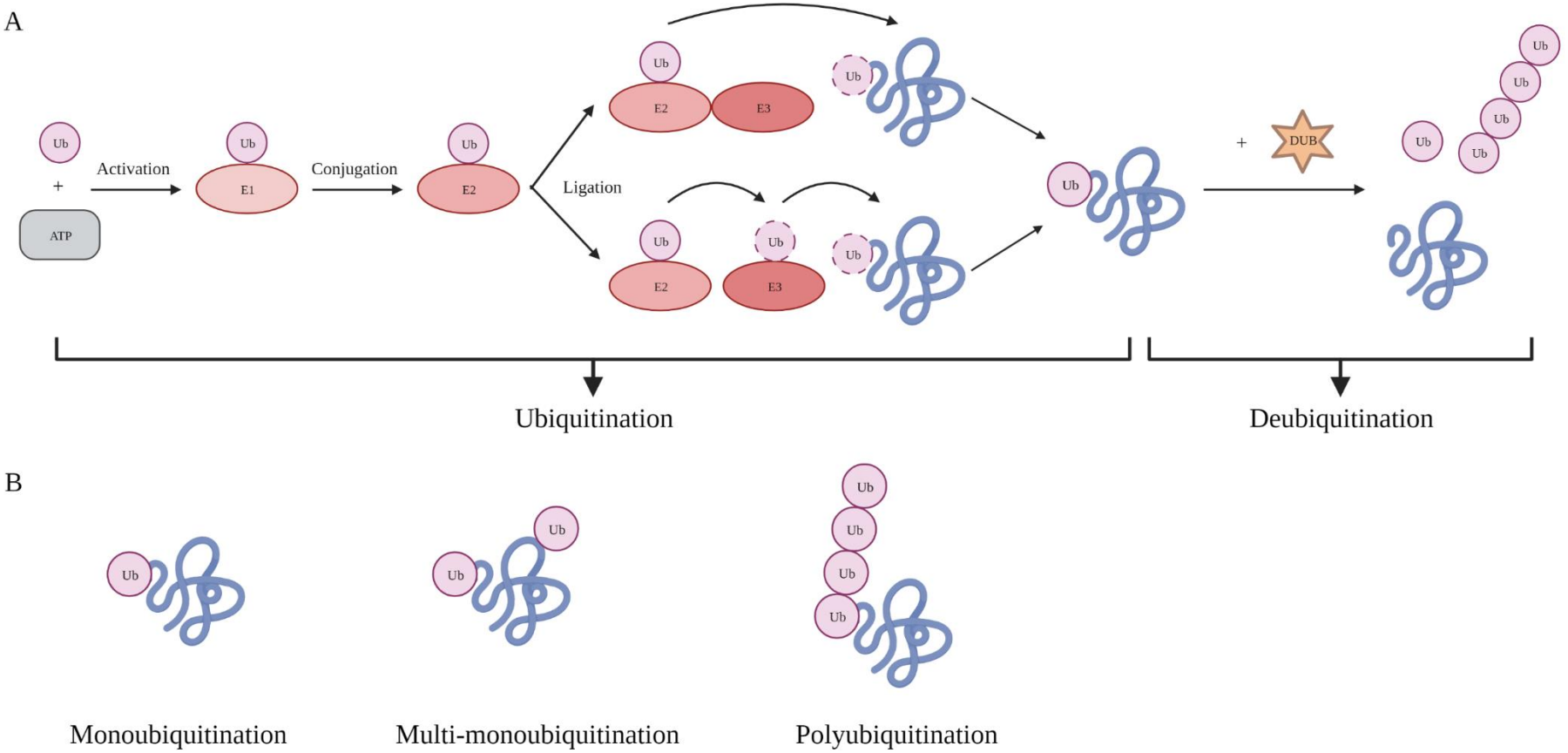

Figure 2.2 The simplified ubiquitination cascade and the diversity of ubiquitinated products. A: Ubiquitin is activated by E1 and transferred to the Cys residue of E2 before the covalent attachment of ubiquitin to the Lys residue of the substrate protein that mediated by E3. The ligation process is either directly from the E2 to the substrate protein or sequentially from the E2 to the E3 and then to the substrate protein. The DUBs that recognise ubiquitinated proteins and remove the ubiquitin tags from the protein. B: The different chain types of ubiquitin modifications: monoUb, multi-monoUb and polyUb. The illustration was created by BioRender.com.

2.2 Deubiquitinating enzymes

DUBs are a large family of proteases that act to hydrolyse the covalent bond between the single ubiquitin or polyubiquitin chains from the substrate. Overall, the general role of DUBs can be summarised in four main categories: ubiquitin processor, remover, editor and recycler. When the ubiquitin molecule is expressed by genes RPS27, UBA52, UBC or UBB, it is created as precursors, which are either precursors made up of 3-10 single ubiquitin linked together (when produced by UBC and UBB) or precursors coupled to ribosomal proteins L40 or S27a (in the case of expression by RPS27 and UBA52 genes). Here, DUBs serve as the ubiquitin maturation processors for the newly expressed ubiquitin (Komander et al., 2009). Next, DUBs can remove ubiquitin chains attached to substrate protein prior to modulating the substrate protein activities. The DUBs also operate as ubiquitin chain editors of mixed polyubiquitin chains by removing one chain linkage type prior to the elongation of a second chain. The removal and remodelling of the polyubiquitin chains affects the fate of the substrate protein (Stone, 2016). Additionally, as part of a process known as ubiquitin recycling, DUBs are in charge of cutting unanchored polyubiquitin chains into free single ubiquitin, preserving the pool of free ubiquitin (Asaoka and Ikeda, 2015; D'Arcy et al., 2015; Eletr and Wilkinson, 2014; Hermanns et al., 2018; Isono and Nagel, 2014).

2.2.1 Classification of DUBs

The human genome encodes approximately 101 putative DUBs, which may be classified into two major classes: cysteine proteases and metalloproteases. The former consisted of six families categorised based on their catalytic domains: motifs interacting with ubiquitin-containing novel DUB family (MINDYs), Machado-Joseph-disease proteases (MJDs), ovarian tumour proteases (OTUs), ubiquitin carboxy-terminal

hydrolases (UCHs), ubiquitin-specific proteases (USPs) and the most recently added zinc finger with UFM1 specific peptidases (ZUFSPs), discovered in 2018 (Hermanns et al., 2018). Despite the fact that each family has its own distinct folds, the catalytic triad or diad residues of the diverse family stay neatly superimposed when bound to ubiquitin (Komander et al., 2008; Komander and Barford, 2008). A histidine (His) side chain is necessary in cysteine proteases for nucleophilic attack of a Cys residue on isopeptide bonds. To polarise the His residue, a third residue is usually required, and in most cysteine protease DUBs this residue is an aspartate or asparagine. The last family is JAMM/MPN associated metalloproteases (JAMMs) that belongs to the metalloproteases class with zinc residues as active sites (Keusekotten et al., 2013; Komander et al., 2009; Kwasna et al., 2018; Maurer and Wertz, 2016; Rehman et al., 2016; Wang and Wang, 2021).

2.2.1(a) OTU deubiquitinase

The human genome encodes 16 DUBs of the OTU family that play critical roles in signalling cascades such as innate immunity and cell cycle control (Du et al., 2020). OTU domains can hydrolyse isopeptides even in the absence of aspartate or asparagine residue needed for polarisation of catalytic histidine (Komander and Barford, 2008). The first OTU proteins to be recognised as DUBs were OTUB1 and OTUB2, both showed the ability to cleave ubiquitin-GFP and tetraubiquitin *in vitro* (Balakirev et al., 2003; Borodovsky et al., 2002). Since then, the OTU family has expanded and been phylogenetically subclassified into the Otubains (OTUB1 and OTUB2), the A20-like OTUs (OTUD7A, OTUD7B, TNFAIP3, VCPIP1 and ZRANB1), the OTUDs (OTUD1, OTUD2, OTUD3, OTUD4, OTUD5, OTUD6A, OTUD6B and ALG13), and the OTULIN subfamilies (Kayagaki et al., 2007; Keusekotten et al., 2013).

The OTUs exhibit a striking propensity for cleaving polyubiquitin chains with particular linkage types, as reviewed by Du et al., (2020). In summary, OTUD2 prefers Lys11-linked chains while OTUD3 prefers Lys6-linked diubiquitin chains. Meanwhile, TNFAIP3 prefers to cleave Lys11 and Lys48 chains. ZRANB1 is Lys29 and Lys33 specific whereas OTUB1 favours Lys48 linkages, OTUD7A/OTUD7B prefers Lys11 linkages, and OTULIN is Met1 specific (Du et al., 2020). Interestingly, OTUD1's Lys63 specificity depends on its C-terminal UIM domain (Mevissen et al., 2013) while in the case of ZRANB1, an Ank (ankyrin repeat) UBD domain is required for its specificity towards Lys29- and Lys33-linked proteins (Licchesi et al., 2012).

2.2.1(a)(i) OTUB1

OTU deubiquitinase, ubiquitin aldehyde binding 1 (OTUB1) is a pioneer member in OTU DUBs with 271 amino acid-length sequence and molecular mass of 31.28 kDa. Two distinct ubiquitin-binding sites are shown in the crystal structure of OTUB1. During the deubiquitination process, OTUB1 preferentially cleaves the Lys48 polyubiquitin chains using the active centre to catalyse the substrate reaction (Wiener et al., 2013). OTUB1 catalytic domain consisted of three parts: Cys91, His265, and Asp267. In the absence of ubiquitin, the His265 is located too far to interact directly with the Cys91 but the His265 moved towards a close proximity with Cys91 in the presence of ubiquitin (Edelmann et al., 2009).

OTUB1 has gained a great deal of interest in cancer research. Since its discovery as a p53 interactor, it has emerged as an essential regulator in disease-related pathways (Sun et al., 2012). Numerous studies have reported a direct association between OTUB1 and various cancer-related proteins and plays a role in the development of various malignancies such as glioma, gastric adenocarcinoma and breast, colorectal, lung, ovarian and prostate cancers (Saldana et al., 2019). For example, OTUB1 interacts with

FOXM1, a transcription factor that promotes cell cycle progression and epithelial-mesenchymal transition (EMT). In breast cancer, OTUB1 was discovered to reduce the cytotoxic effects of drug epirubicin by deubiquitinating and stabilising FOXM1, hence promoting cell proliferation in a manner dependent on the FOXM1 and OTUB1 deubiquitination activity (Karunarathna et al., 2015). Other than that, OTUB1 can also suppress the ubiquitination of several proteins such as phospho-SMAD2/3 and p53 by attaching to and blocking the E2 ubiquitin-conjugating enzymes independent of its catalytic activity (Herhaus et al., 2013; Juang et al., 2012).

2.2.1(a)(ii) OTUB2

OTU deubiquitinase, ubiquitin aldehyde binding 2 (OTUB2) is the closest structural homologue of OTUB1 that does not share similar interaction partners as OTUB1, due to the lack of N-terminal region that allows OTUB1 to interact with its substrates (Edelmann et al., 2009). OTUB2 is also smaller in size, with 234 aa and molecular mass of 27.21 kDa. Its involvement in cancer is also not comparable to OTUB1 in that it was reported more later and slower than OTUB1. Nonetheless, its significance in cancer is becoming more prominent, for instance, it was shown to be overexpressed in non-small cell lung cancer (NSCLC) and promoted its development (Li et al., 2019). Elevated expression of OTUB2 was also observed in both liver cancer tumour tissues and cell lines, indicating a poor prognosis for individuals with liver cancer (Gu et al., 2020). Furthermore, it was also reported that the suppression of OTUB2 inhibited the development of papillary thyroid carcinoma (Ma and Sun, 2019). Other than that, OTUB2 performs various biological functions, including supporting the DNA repair pathway (Johmura et al., 2016; Kato et al., 2014), and negatively regulating virus-triggered interferon (IFN) induction pathways and cellular antiviral

response via interaction with the tumour necrosis factor receptor-associated factor (TRAF) protein family, TRAF3 and TRAF6 (Li et al., 2010).

2.2.1(a)(iii) OTULIN

OTU deubiquitinase with linear linkage specificity (OTULIN) is a protein consisting of 352 aa and possesses a molecular weight of approximately 40.26 kDa that preferentially removes linear (Met1-linked) polyubiquitin chains from substrates and functions as a regulator of innate immune response and angiogenesis (Hrdinka et al., 2016; Keusekotten et al., 2013; Rivkin et al., 2013). OTULIN is involved in the negative regulation of nuclear factor kappa B (NF-κB) signalling (Xu et al., 2018) by counteracting linear ubiquitin chain assembly complex (LUBAC) and so inhibiting the activation of NF-κB signalling (Damgaard et al., 2020; Li et al., 2022; Zhao et al., 2020). Furthermore, OTULIN loss-of-function mutation promotes a severe autoinflammatory condition referred to as OTULIN-Related Autoinflammation Syndrome (ORAS) that is normally inhibited by OTULIN-LUBAC interaction in the absence of mutation (Fiil and Gyrd-Hansen, 2016; Li et al., 2022; Zhou et al., 2016). Furthermore, Damgaard et al., (2020) discovered that a deficiency in OTULIN caused spontaneous and progressive steatotic liver disease in an ORAS patient between the age of 10-13 months. OTULIN deficiency is observed to be the cause of metabolic alterations, apoptosis and inflammation in the liver, hence, it is considered as an essential DUB for maintaining liver homeostasis and preventing the development of liver disease (Damgaard et al., 2020).

2.2.1(b) Sequence alignment of OTUB1, OTUB2 and OTULIN

Figure 2.3 illustrates the alignment of OTUB1, OTUB2 and OTULIN amino acid sequences, with the OTU domain highlighted within a boxed region, indicating that all three proteins are from the same subfamily. Significant differences can be observed,

especially in the N-terminal region that contribute to varying interaction partners or binding affinities, potentially influencing the specificity of PPIs. The aligned sequences show the identical amino acid identity as represented by the asterisk (*), as well as amino acids with similar properties at specific positions (: and .), indicating that these regions have similar biochemical properties or functions, thus providing valuable insights into the roles and relationships of the proteins.

Figure 2.3 Sequence alignment of OTUB1, OTUB2 and OTULIN.

2.2.2 DUBs' implication on health and diseases

DUBs can regulate protein stability by releasing ubiquitin molecules from substrate proteins, cancelling signals for protein degradation and other non-proteolytic activities (Clague et al., 2012; Komander et al., 2009; Sowa et al., 2009; Stone, 2016). Deconjugation of ubiquitin from ubiquitinated substrates is critical not only in physiological condition but also in the regulation of numerous cellular events and

biological processes that underpin the development of disease. DUBs dysregulation or malfunction has been implicated in diseases such as cancers, immunological diseases and neurological disorders, usually resulting from aberrant signalling within the cell (Fraile et al., 2012; Kowalski and Juo, 2012).

Extensive research has elucidated the role of DUBs in the central nervous system (Ristic et al., 2014; Todi and Paulson, 2011). Oxidative stress and DUB activity modulation have been specifically associated with certain neurological disorders such as alzheimer's disease, parkinson's disease, and epilepsy (Chakraborty and Ziviani, 2020; Das et al., 2020; Kowalski and Juo, 2012; Paemka et al., 2015). Overexpression of USP11 increases the risk of alzheimer's disease in women, which USP11 deubiquitinates microtubule-associated protein tau and decreases proteasomal degradation of tau, leading in increased tauopathy and vulnerability to alzheimer's disease (Paul, 2022). By knocking down USP11 promotes tau clearance while physically preventing acetylation and decreasing phosphorylation of tau (Yan et al., 2022). Parkin ubiquitinates mitochondrial components promotes mitochondria turnover through lysosome-mediated mitophagy, defective mitophagy and the resulting accumulation of defective mitochondria contribute to enhanced oxidative stress, which is thought to underlie parkinson's disease (Hauser and Hastings, 2013; Narendra and Youle, 2011). DUB that opposes Parkin's function was identified as USP30, which is associated with mitochondria and acts as an antagonist of Parkin-mediated mitophagy (Bingol et al., 2014; Durcan and Fon, 2015). Besides that, USP30 is considered to counteract the clearance of damaged mitochondria in cases of mitochondrial dysfunction due to deficiencies in Parkin, leading to a build-up of metabolically and energetically deficient cells (Bingol et al., 2014). Meanwhile, depletion of USP8 was observed to delay the translocation of Parkin onto depolarised mitochondria, along with

a delay in the clearance of mitochondria, and USP8 was discovered to be capable of removing Lys6-linked ubiquitin chains from Parkin *in vitro* (Durcan et al., 2014). USP15 and ATXN3 have been identified as a Parkin-interacting proteins that colocalise with mitochondria (Cornelissen et al., 2014) and counteract Parkin autoubiquitination (Durcan et al., 2012), respectively.

The imbalance between FOXP3 and GATA3 is a significant factor contributing to the pathogenic alteration in a key component of immune system, regulatory T (Treg) cells in asthma patients (Chen et al., 2018). Increased expression of USP21 was discovered and confirmed to regulate the stability of both FOXP3 and GATA3 (Li et al., 2016; Zhang et al., 2013). USP21 has also been implicated in regulating liver fibrosis in schistosomiasis patients, indicating its potential role in the Treg cell-mediated regulation of immune interactions between *Schistosoma* and its host (Zhang et al., 2021). Additionally, previous studies have indicated that the depletion of USP21 enhances the immune defense against certain viruses, resulting in reduced viral replication both *in vitro* and *in vivo* (Chen et al., 2017; Fan et al., 2014; Wu et al., 2021). A20 is a ubiquitin-editing enzyme that has been suggested to limit activation of NF-κB, a transcription factor mediating inflammatory and innate immune signalling pathways by removing the Lys63-linked ubiquitin chains from the ubiquitinated substrates and then conjugating them to the Lys48-linked polyubiquitin chains to trigger proteasomal degradation (De et al., 2014; Feoktistova et al., 2020; Hymowitz and Wertz, 2010; Shembade and Harhaj, 2012). Besides, mutations in several DUBs, including USP18 and A20, have been found to be associated with multiple sclerosis, a chronic inflammatory demyelinating disease (De Jager et al., 2009; Malhotra et al., 2011, 2013).

Furthermore, DUBs regulate numerous cancer proteins and are the direct antagonists of many oncogenic or tumour-suppressive E3 ligases (Hoeller and Dikic,

2009). In the regulation of p53 stability and activity, an E3 ligase MDM2 ubiquitinates and destabilises p53, whereas DUBs OTUB1, USP7 and USP10 deubiquitinated and stabilised it (Sun et al., 2012). DUB CYLD plays a predominant role in the negative regulation of the NF-κB activation, contributing to its tumour suppressor function (Lork et al., 2017; Sun, 2010). Meanwhile, OTUD5 is identified as a negative regulator of type I IFN production by removing Lys63-linked polyubiquitin chains of tumour protein TRAF3 (Cho et al., 2021; Kayagaki et al., 2007). Elevated DUB expressions have been observed in a variety of cancers, such as glioma, endometrial cancer, ovarian cancer and breast cancer (Sivakumar et al., 2020). In breast cancer, increased expression levels of COPS5, OTUD6B, UCHL5, USP7 and VCPIP1 were reported (He et al., 2017). As a result, DUBs are currently in the spotlight as promising therapeutic targets in diseases ranging from oncology to neurodegeneration (An et al., 2022).

The escalating interest towards DUBs as potential therapeutic targets, especially in context like cancer, underscores their extensive functional diversity. This diversity significantly influences the regulation of various biological processes, including cell cycle control, DNA repair, chromatin remodelling and several signalling pathways that are frequently altered in cancer (Hussain et al., 2009; Komander et al., 2009; Turcu et al., 2009). Figure 2.4 shows a visual summary illustrating the diverse roles of several DUBs (pink ovals) implicated in cancer and involved in distinct cellular pathways. In the diagram, ubiquitin in green indicates Lys48-linked chains targeting proteins to the proteasome, whereas ubiquitin in blue corresponds to non-Lys48-linked chains. Numerous DUBs, including USP7, USP11, USP13, USP19, USP37, USP39, USP44, USP50 and BAP1 have important roles in cell cycle progression. Simultaneously, USP1, USP3, USP11, USP16, USP28, USP47, BRCC3 and OTUB1 are involved in DNA damage repair. Besides, USP3, USP7, USP16, USP21, USP22, UCHL5,

MYSM1, BAP1 and BRCC3 are participated in chromatin remodelling by deubiquitinating histones or other chromatin-related substrates. DUBs that play a role in p53 regulation include USP2, USP4, USP5, USP7, USP10 and USP29, while A20, OTUD7B, OTUD5, CYLD, USP2, USP4, USP11, USP15 and USP21 are involved in NF-κB signalling. USP8, USP18, STAMBP and PSMD7 interfere in receptor tyrosine kinase trafficking. Meanwhile, USP4, USP15, USP34 and ZRANB1 are associated with Wingless-related integration site (Wnt) signalling. Finally, USP9X, STAMBPL1 and UCHL5 play roles in regulating the transforming growth factor-beta (TGF-β) pathway.

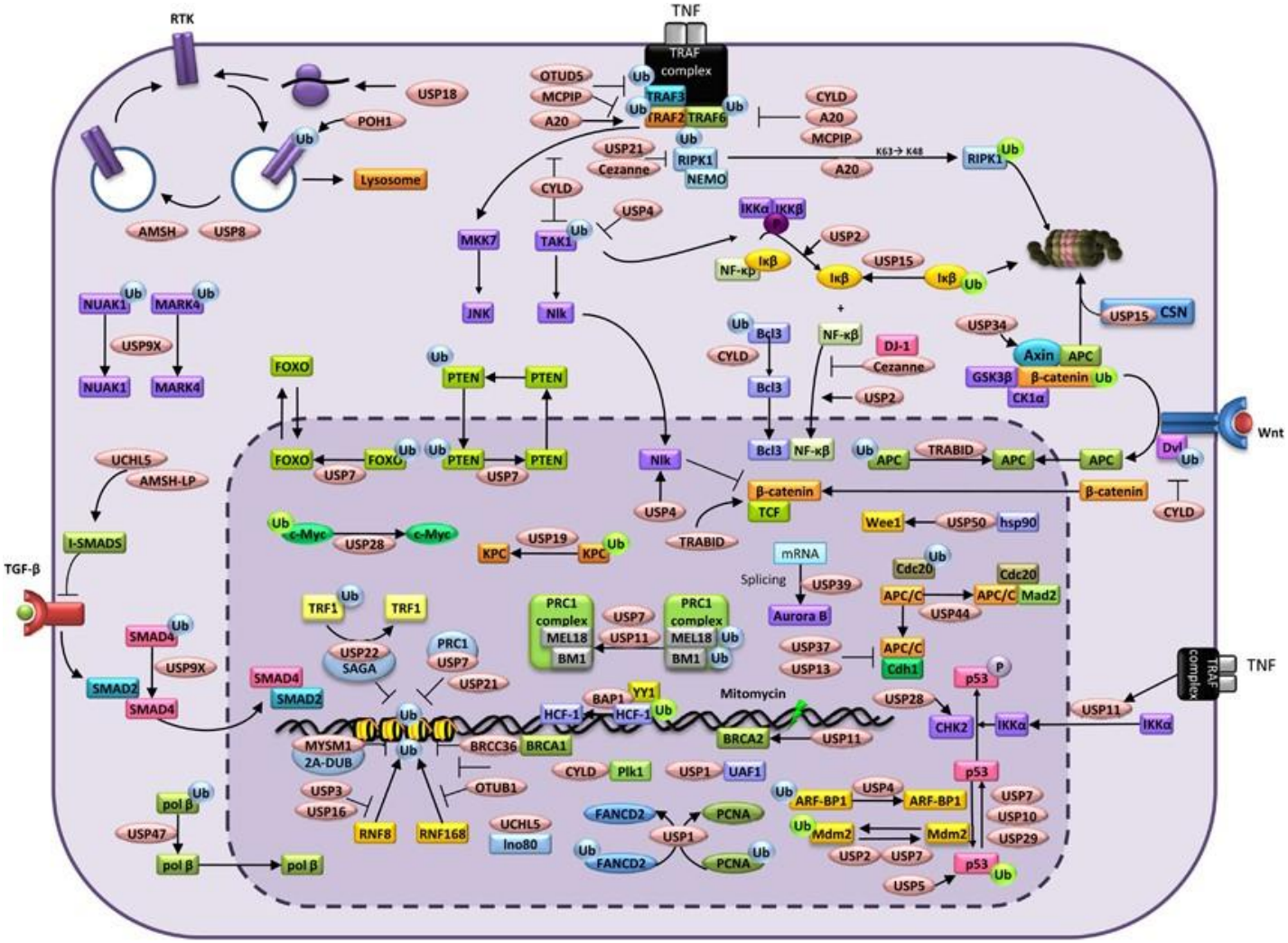

Figure 2.4 Overview of the different roles of several DUBs in cancer. The diagram was adapted from Fraile et al., (2012).

2.3 Protein-protein interaction

Proteins are complex molecules that were identified as the primary agents of biological function that determined the phenotype of all organisms (Safari-Alighiarloo et al., 2014). They play a central role in biological processes and perform essential functions and regulation throughout the system of human body, including initiating cell signalling cascades, operating as transporters on the cell membrane, catalysing biochemical reactions and serving as antibodies, to name a few (Browne et al., 2010; Jaimovich, 2010; Perkins et al., 2010). In many of these tasks, proteins accomplished their biological roles through interacting and collaborating with other proteins (Browne et al., 2018; Ideker and Sharan, 2008; Lage, 2014; Rao et al., 2014; Sarajlić et al., 2013; Sevimoglu and Arga, 2014; Vidal, 2005). It has been estimated that over 80% of proteins are not naturally functional in isolated forms but in complexes (Berggård et al., 2007), in which they have interactions with not only proteins but also other molecules such as DNA, RNA, chemical or lipid (Gonzalez and Kann, 2012; Makino and Gojobori, 2007; Pedamallu and Posfai, 2010; Peng et al., 2017; Pratt et al., 2016; Yu et al., 2020). Therefore, identifying and characterising the full repertoire of these cellular machinery and the interplay between them is of utmost relevance for a comprehensive understanding of a living cell functionality (Jaimovich, 2010).

Protein interaction is essentially the study of how proteins work with one another (Koh et al., 2012) and one of the significances of studying protein interaction is to deduce the function of protein within the cell (Rao et al., 2014). Phizicky and Fields, (1995) proposed that unknown protein functionality could be determined by investigating the interaction of unknown proteins with a known protein target, whose function is already revealed. This is because the physical interaction between proteins

is one of the strongest indications of functional association (Bergholdt et al., 2007). In other words, proteins implicated in the same phenotype are almost certainly belong to the same functional module (Gandhi et al., 2006).

PPIs can be categorised in several ways based on the differing structural and functional features (Nooren and Thornton, 2003). They can be homo- or heterooligomeric, based on the type of their components: obligate or nonobligate, judged by their stability: transient or permanent, as measured by their binding constant (Zhang et al., 2009). A particular PPI could be a combination of these three pairs. Usually, the transient interactions represent signalling pathways whereas permanent interactions indicate a stable protein complex.

Over the past two decades, the emergence of numerous technologies have enabled large-scale high-throughput experiments for detecting novel PPI and resulted in an explosion in the volume of PPI data (Ding and Kihara, 2019). Because visualising this data in tabular format is increasingly unfeasible, mathematical representations interactions between proteins PPI are performed in the form of graphs (also more commonly known as networks), in which proteins are represented as nodes, and edges that connected the nodes are evidence for functional correlations between the nodes (Figure 2.5) (Bergholdt et al., 2007; Browne et al., 2010; Lage, 2014; Peng et al., 2017; Sarajlić et al., 2013). The interpretation of PPI networks, usually involving many tools derived from graph theory, is a key step to understand the represented system.