

**Exploiting the CRL4<sup>CRBN</sup> E3 ubiquitin  
ligase complex for the development  
of iTAG: a versatile targeted protein  
degradation system to enable  
functional evaluation and target  
validation.**

**Stephanos Nicolaou**

The Institute of Cancer Research  
University of London



This thesis is submitted for the degree of PhD

## Declaration

I declare that this thesis is an original report of my research, has been written by me and has not been submitted for any previous degree. The experimental work is almost entirely my own work; the collaborative contributions have been indicated clearly and acknowledged throughout the text. Due references have been provided on all supporting literatures and resources. I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

*Stephanos Nicolaou*

September 2021

## Abstract

Modulating target protein expression to probe for its relevance and function in a disease is critical in the drug discovery process of target validation. Current genetic methods and chemical-biology tools are widely employed for target protein modulation; however they exhibit several limitations in their use. The work presented in this thesis describes the development of a novel protein degradation method named iTAG (Inducible and TArgeted protein deGradation) based on thalidomide and its analogues [IMiDs (immunomodulatory imide drugs)/ CELMoDs (Cereblon E3 ligase modulation drugs)]. These compounds engage CRBN (cereblon), the substrate receptor of CUL4<sup>CRBN</sup> E3 ubiquitin ligase, and alter the enzyme target specificity leading to the recruitment and subsequent ubiquitination of neosubstrates. Up-to-date, examples of neosubstrates include GSPT1, CK1 $\alpha$ , Ikaros/IKZF1, Aiolos/IKZF3, and ZFP91. Mechanistically, neosubstrates interact with the IMiD/CELMoD-bound surface of CRBN through recognition motifs known as degrons. Using structural and sequence analysis, native and synthetic degron-containing domains (DCDs) (size ranging from 10 to 197 residues) were systematically explored and evaluated for their degradability when fused to a target protein. An optimal DCD (DCD23; 60 residues) was identified, that induces a rapid, potent, and selective degradation of nuclear, cytoplasmic, and mitochondrial proteins following IMiD/CELMoD treatment. DCD23 was also demonstrated to constitute a modular and versatile tool for degrading proteins when fused to either their N or C terminus. Furthermore, acute iTAG-mediated degradation of cMYC and KRAS[G12V] was showed to enable

the exploration of the downstream functions of these classical oncoproteins. In vivo, the iTAG system enables potent and reversible protein degradation in tumours, with acute loss of target protein by 4h following a single oral administration of the CELMoD CC-220. The iTAG system therefore constitutes a transformational modular and versatile tool to specifically degrade proteins of interest, which will enable systematic exploration of protein function across the human proteome.



## Acknowledgements

Firstly, I would like to thank the Institute of Cancer Research (ICR), Cancer Research UK (CRUK) and the Medical Research Council (MRC) for providing the funding to carry out my PhD project and for financially supporting me throughout my studies at the ICR.

A very special thanks to my primary supervisor Prof. Ian Collins and my secondary supervisor Dr Olivia Rossanese, my ex-primary supervisor Prof. Rajesh Chopra (left ICR April 2020) and my ex-associate supervisor Dr Amine Sadok (left ICR July 2019) for their mentorship, scientific training, and unwavering support. I am also immensely grateful to my associate supervisor Dr Habib Bouguenina for his mentoring, scientific training, and guidance throughout my PhD studies. His support and encouragement have been pivotal during my learning and my scientific progression and have shaped me as a scientist. A big thanks to my supervisors Prof. Ian Collins and Dr Habib Bouguenina for their comments and guidance in writing this thesis. I would also like to collectively thank all of my supervisors for being great role models during my PhD studies and for giving me the opportunity to work on a scientifically stimulating and challenging project and allowing me the independence to explore my own ideas.

I am also grateful for the support of all past and present members of the Translational Cancer Discovery (TCD) team and the Target Evaluation and Molecular Therapeutics (TEMT) team. In particular, I would like to thank Dr Hannah Wang for mentoring and scientifically training me during the early years

---

## Acknowledgements

of my PhD and for her support during difficult times. Moreover, a special thanks to Dr Laura Chan, Dr Nil Ege, Dr Esther Arwert, Dr Bradleigh Whitton, Dr Anjali Aurora, Dr Fernando Jr. Sialana, Dr Mara Tatari, Ilona Bernett and Xin Yin Yeo for their training in several experimental techniques and the many interesting discussions, both scientific and not, we had in the lab over the years. Many thanks to Dr Charlotte Pawlyn and her team for their support and help with various aspects of my PhD project.

In terms of collaborative work, I would like to express my gratitude to my colleagues at the ICR Cancer Therapeutics Unit, and in particular Dr Yann-Vaï Le Bihan, Dr Arjun Thapaliya, Dr Craig McAndrew and Dr Mark Stubbs for their scientific input and training through my PhD studies. Many thanks to our external collaborators, Prof. Trey Westbrook, and his team at the Baylor College of Medicine for carrying out the in vivo experiments. Thanks go to Dr Ian Tittley and Dr Steve Hobbs for assisting me with the flow cytometry equipment and teaching me molecular cloning, respectively. A special thank you to everyone who has helped me at the ICR over the past 4 years.

I would also like to give a massive thanks to my family, especially my dad, mum, and brother for supporting me throughout my PhD studies and for believing in me. A special thanks to my uncle Stelios and auntie Elsa for their support during the last month of my PhD, and to my auntie Androulla for mentoring and guiding me through the years. I would also like to say a big thank you to my partner Maria for her understanding and support throughout this time. Lastly, many thanks to my flatmate and friend Platon for his unwavering support and to all my close friends who have been part of this journey.

---

# Table of Contents

<b>Declaration</b> .....	2
<b>Abstract</b> .....	3
<b>Acknowledgements</b> .....	5
<b>Table of Contents</b> .....	7
<b>List of Figures</b> .....	10
<b>List of Tables</b> .....	14
<b>Abbreviations</b> .....	15
<b>1. INTRODUCTION</b> .....	27
1.1 <i>The process of drug discovery</i> .....	27
1.2 <i>Pitfalls of target validation in cancer drug discovery</i> .....	34
1.3 <i>Overview of genetic methods in target validation</i> .....	37
1.3.1 <i>siRNA and shRNA methods</i> .....	37
1.3.2 <i>CRISPR-based systems</i> .....	44
1.4 <i>Targeted protein degradation</i> .....	52
1.4.1 <i>The UPS (Ubiquitin-Proteasome System)</i> .....	53
1.4.2 <i>Chemical-biology tools exploiting the UPS</i> .....	66
1.5 <i>Rationale for developing a novel protein degradation tool</i> .....	98
1.5.1 <i>IMiDs/CELMoDs engage the CRL4<sup>CRBN</sup> E3 ubiquitin ligase and induce protein degradation</i> .....	100
1.5.2 <i>Exploiting CRL4<sup>CRBN</sup> E3 ubiquitin ligase to develop a novel targeted protein degradation system</i> .....	113
<b>2. MATERIALS AND METHODS</b> .....	116
2.1 <i>Cell lines and cell culture</i> .....	116
2.2 <i>Plasmids</i> .....	118
2.3 <i>Transient cell transfection</i> .....	133
2.4 <i>Virus production and cell line transduction</i> .....	134
2.5 <i>Drug storage and treatment</i> .....	136
2.6 <i>Cell Lysis</i> .....	137
2.7 <i>Electrophoresis and Western Blotting</i> .....	138

---

Table of Contents

2.8 Flow Cytometry.....	139
2.9 Cell Viability Assay.....	140
2.10 Recombinant protein production and purification.....	141
2.10.1 CRBN/ DDB1.....	141
2.10.2 DCD1.....	141
2.10.3 DCD19.....	142
2.10.4 DCD23.....	145
2.11 IMiD-based Fluorescence Polarisation assay.....	146
2.12 Aiolos peptide-based TR-FRET assay.....	147
2.13 In vivo degradation assay.....	148
2.14 Immunofluorescence.....	149
2.15 Statistical analysis.....	150
<b>3. RESULTS</b> .....	<b>152</b>
<b>Subchapter 3.1 : Identification of novel DCD motifs mediating acute and potent degradation of target proteins</b> .....	<b>152</b>
3.1.1 Design and generation of DCD1 to DCD9.....	153
3.1.2 Fusion of EGFP-DCD sequences to cMYC and stable expression in HMEC cells.....	157
3.1.3 DCD1 and DCD2 induced a potent decrease of cMYC-EGFP levels following CC-885 treatment.....	160
3.1.4 CC-885 treatment induces cellular toxicity.....	165
3.1.5 Design and generation of DCD10 to DCD22.....	167
3.1.6 Ikaros and Aiolos-based DCDs induced a potent decrease of cMYC-EGFP levels following IMiD/CELMoD treatment.....	171
3.1.7 Design and generation of DCD23 to DCD26.....	175
3.1.8 Chimeric DCDs induced a potent decrease in cMYC-EGFP levels following IMiD/CELMoD treatment.....	177
3.1.9 Acute degradation of cMYC-EGFP fused to DCD19 and DCD23 following IMiD/CELMoD treatment.....	181
3.1.10 Sub-micromolar concentrations of IMiDs/ CELMoDs induced a potent degradation of cMYC-EGFP fused to DCD19 and DCD23.....	182
3.1.11 DCDs efficiently engage CRBN/DDB1 in vitro.....	185
3.1.12 cMYC degradation is mediated by the UPS.....	191
3.1.13 cMYC degradation is reversible.....	193
3.1.14 Interim Discussion Results Subchapter 3.1.....	195
<b>Subchapter 3.2 : The iTAG system is a versatile tool for studying target protein function</b> .....	<b>197</b>

3.2.1 Degradation of N-terminally tagged cMYC following IMiD/CELMoD treatment allows to study functional consequences of nuclear target protein loss .....	198
3.2.2 CC-220 and CC-92480 exhibit a similar degradation profile of iTAG-tagged cMYC.....	201
3.2.3 iTAG is a versatile tool to degrade both cytoplasmic and nuclear proteins .....	205
3.2.4 Potent degradation of DCD23-tagged KRAS[G12V] and EZH2 following IMiD/CELMoD treatment.....	208
3.2.5 Degradation of N-terminally tagged KRAS[G12V] following IMiD/CELMoD treatment allows to study the functional consequences of cytoplasmic target protein loss.....	212
3.2.6 Degradation of iTAG-tagged EZH2 following IMiD/CELMoD treatment causes a decrease of H3K27me3 levels.....	214
<b>Subchapter 3.3 : The iTAG system constitutes a modular tool for both in vitro and in vivo target protein evaluation.....</b>	<b>217</b>
3.3.1 iTAG-induced degradation of target proteins is more potent and efficient compared to the dTAG system. ....	218
3.3.2 In vivo degradation of an iTAG-tagged target protein .....	222
<b>Subchapter 3.4 : Engineered mutant DCD23 motifs mediate acute and potent degradation of target proteins .....</b>	<b>227</b>
3.4.1 Design and generation of mutant DCD23 degrons .....	228
3.4.2 Mutant DCD23 motifs induced a potent and robust degradation of EGFP .....	230
3.4.3 Examining the cellular localisation of mutant DCD23 motifs fused to EGFP.....	234
3.4.4 Interim Discussion Results Subchapter 3.4 .....	244
<b>4. DISCUSSION.....</b>	<b>247</b>
<b>References.....</b>	<b>269</b>

---

## List of Figures

### INTRODUCTION

<b>Figure 1. 1</b> The process of drug discovery..	29
<b>Figure 1. 2</b> RNAi (RNA interference) cellular mechanism.	40
<b>Figure 1. 3</b> The CRISPR/Cas9 genome editing tool and its different application strategies.	47
<b>Figure 1. 4</b> The structure of ubiquitin (green, PDB code 1UBQ) with its lysine residues.	55
<b>Figure 1. 5</b> The Ubiquitin-Proteasome System.....	56
<b>Figure 1. 6</b> The different topologies of ubiquitination.....	57
<b>Figure 1. 7</b> Mechanism of ubiquitin transfer to a substrate by the three different classes of E3 ubiquitin ligases.	60
<b>Figure 1. 8</b> Schematic representation of an assembled cullin RING E3 ubiquitin ligase.....	63
<b>Figure 1. 9</b> Schematic diagram of the 26S proteasome.	65
<b>Figure 1. 10</b> The destabilising domain system.	67
<b>Figure 1. 11</b> The ligand-induced degradation (LID) system.....	70
<b>Figure 1. 12</b> The AID (Auxin Inducible Degron) system.	73
<b>Figure 1. 13</b> The SMASh tag system.....	80
<b>Figure 1. 14</b> HaloTag-based systems.....	84
<b>Figure 1. 15</b> The dTAG system.	87
<b>Figure 1. 16</b> The TRIM-Away system.	91
<b>Figure 1. 17</b> The AdPROM system.....	93
<b>Figure 1. 18</b> The deGradFP system.	97
<b>Figure 1. 19</b> Chemical structures of IMiDs thalidomide, lenalidomide and pomalidomide.....	101
<b>Figure 1. 20</b> The CRL4 <sup>CRBN</sup> E3 ubiquitin ligase.	104
<b>Figure 1. 21</b> Chemical structures of CELMoDs iberdomide (CC-220), mezigdomide (CC-92480), avadomide (CC-122), CC-90009 and CC-885. ...	107
<b>Figure 1. 22</b> CRL4 <sup>CRBN</sup> E3 ubiquitin ligase-mediated ubiquitination.	109
<b>Figure 1. 23</b> Binding mode of various known classes of neosubstrates on CRBN.	115

## MATERIALS AND METHODS

<b>Figure 2. 1</b> pLVX_TetOne_PURO lenti-vector map and multiple cloning site (MCS).....	125
<b>Figure 2. 2</b> pET_48b (+) vector map and MCS.....	126
<b>Figure 2. 3</b> pHAGE-PGK vector map.....	127
<b>Figure 2. 4</b> pFBDM vector map. ....	128
<b>Figure 2. 5</b> Generation of lentiviral backbone vectors encoding DCD23-tagged KRAS[G12V] and EZH2. ....	131
<b>Figure 2. 6</b> Generation of lentiviral backbone vector encoding untagged EZH2.. ....	131
<b>Figure 2. 7</b> Generation of lentiviral backbone vectors encoding EGFP tagged at either terminus with DCD23 and DCD23mut1-mut3.....	133
<b>Figure 2. 8</b> <i>In vitro</i> production and purification of DCD19.....	144

## RESULTS

### Subchapter 3.1 : Identification of novel DCD motifs mediating acute and potent degradation of target proteins

<b>Figure 3.1. 1</b> Structure and sequence-based design of DCD1 to DCD9 from CUL4 <sup>CRBN</sup> E3 ubiquitin ligase neosubstrates GSPT1 and CK1 $\alpha$ .....	154
<b>Figure 3.1. 2</b> DCD1 to DCD9 fusion to the C-terminus of cMYC-EGFP for functional evaluation of the various DCDs.. ....	158
<b>Figure 3.1. 3</b> Examining the expression of cMYC-EGFP fused to DCD1-DCD9 and their respective P2A controls.....	161
<b>Figure 3.1. 4</b> Workflow diagram of the experimental pipeline used to evaluate the various DCD constructs. ....	163
<b>Figure 3.1. 5</b> Functional evaluation of DCDs fused to the C-terminus of cMYC-EGFP. ....	165
<b>Figure 3.1. 6</b> Examining cell viability of various cell lines following IMiD/CELMoD treatment.....	166
<b>Figure 3.1. 7</b> Structure and sequence-based design of DCD10 to DCD22 from Aiolos, Ikaros and Eos.. ....	168

---

<b>Figure 3.1. 8</b> DCD10-DCD22 fusion to the C-terminus of cMYC-EGFP for functional evaluation of the various DCDs. ....	170
<b>Figure 3.1. 9</b> Functional evaluation of DCD10 to DCD22 fused to the C-terminus of cMYC-EGFP. ....	173
<b>Figure 3.1. 10</b> Loss of cMYC-EGFP-DCD19 following IMiD/CELMoD treatment is verified by immunoblotting. ....	174
<b>Figure 3.1. 11</b> Structure and sequence-based design of DCD23 to DCD26 from CUL4 <sup>CRBN</sup> E3 ubiquitin ligase neosubstrates Ikaros and ZFP91. ....	176
<b>Figure 3.1. 12</b> Functional evaluation of DCD23 to DCD26 fused to the C-terminus of cMYC-EGFP. ....	178
<b>Figure 3.1. 13</b> Loss of cMYC-EGFP-DCD23 following IMiD/CELMoD treatment is verified by immunoblotting. ....	179
<b>Figure 3.1. 14</b> Determining the earliest timepoint of IMiD/CELMoD-dependent target protein degradation. ....	182
<b>Figure 3.1. 15</b> Determining the minimum IMiD/CELMoD concentration required for target protein degradation. ....	184
<b>Figure 3.1. 16</b> In vitro measurement of various IMiD/CELMoD affinities for CRBN/DDB1 complex. ....	186
<b>Figure 3.1. 17</b> In vitro measurement of IMiD/CELMoD efficacy at recruiting DCDs to the CRBN/DDB1 complex. ....	190
<b>Figure 3.1. 18</b> DCD19 and DCD23-induced loss of target protein is dependent on the UPS system. ....	192
<b>Figure 3.1. 19</b> DCD-induced loss of target protein is reversible. ....	194

### **Subchapter 3.2 : The iTAG system is a versatile tool for studying target protein function**

<b>Figure 3.2. 1</b> Fusion of DCD19 and DCD23 to the N-terminus of cMYC induced its degradation following IMiD/CELMoD treatment. ....	199
<b>Figure 3.2. 2</b> Functional validation of cMYC loss following IMiD/CELMoD treatment. ....	200
<b>Figure 3.2. 3</b> Evaluation of CC-220 vs CC-92480-induced degradation of cMYC at early time-points. ....	202
<b>Figure 3.2. 4</b> Evaluation of the efficiency of cMYC degradation with prolonged CC-220 or CC-92480 treatment of cells. ....	204
<b>Figure 3.2. 5</b> Functional evaluation of DCD19 fused to either terminus of KRAS[G12V] and EZH2. ....	206
<b>Figure 3.2. 6</b> Fusion of DCD19 to either terminus of KRAS[G12V] and EZH2 induced their degradation following IMiD/CELMoD treatment. ....	207
<b>Figure 3.2. 7</b> Functional evaluation of DCD23 fused to either terminus of KRAS[G12V] and EZH2. ....	210



---

<b>Figure 3.2. 8</b> Fusion of DCD23 to either terminus of KRAS[G12V] and EZH2 induced their degradation following IMiD/CELMoD treatment.....	211
<b>Figure 3.2. 9</b> Functional validation of KRAS[G12V] degradation following IMiD/CELMoD treatment. ....	213
<b>Figure 3.2. 10</b> Functional validation of EZH2 degradation following IMiD/CELMoD treatment. ....	215

### **Subchapter 3.3 : The iTAG system constitutes a modular tool for both in vitro and in vivo target protein evaluation**

<b>Figure 3.3. 1</b> The iTAG system induced a more potent and robust degradation of multiple target proteins compared to the dTAG system. ....	221
<b>Figure 3.3. 2</b> Evaluation of iTAG-mediated degradation of Fluc in MDA-MB-231 cells.....	223
<b>Figure 3.3. 3</b> The iTAG system induced the selective degradation of a target protein in vivo.....	226

### **Subchapter 3.4 : Engineered mutant DCD23 motifs mediate acute and potent degradation of target proteins**

<b>Figure 3.4. 1</b> Structure and sequence-based design of DCD23mut1-mut3. ...	229
<b>Figure 3.4. 2</b> Functional evaluation of DCD23mut1-mut3 fused to the N- or C-terminus of EGFP.....	231
<b>Figure 3.4. 3</b> Fusion of DCD23 or DCD23mut1-mut3 to either terminus of EGFP induced its degradation following CC-220 treatment.....	234
<b>Figure 3.4. 4</b> Measuring nuclear/cytoplasmic localisation of N-terminally DCD23 and DCD23mut1-mut3 -tagged EGFP. ....	241
<b>Figure 3.4. 5</b> Measuring nuclear/cytoplasmic localisation of C-terminally DCD23 and DCD23mut1-mut3 -tagged EGFP. ....	243

## **DISCUSSION**

<b>Figure 4. 1</b> The iTAG system.. ....	250
---	-----

---

## List of Tables

### INTRODUCTION

<b>Table 1. 1</b> Summary of IMiD/CELMoD neosubstrate specificity. ....	110
---	-----

### MATERIALS AND METHODS

<b>Table 2. 1</b> Vectors commercially generated.....	122
<b>Table 2. 2</b> Vectors generated in-house.....	123
<b>Table 2. 3</b> Vectors generated by collaborators. ....	124
<b>Table 2. 4</b> Donor vectors used for PCR-amplification of target proteins' cDNA. .....	129
<b>Table 2. 5</b> Primary antibodies used for immunoblotting. ....	139
<b>Table 2. 6</b> Symbol meaning. ....	151

### RESULTS

<b>Table 3.1. 1</b> Summary of flow cytometry data on the degradation of all the DCDs tested. ....	180
<b>Table 3.1. 2</b> Summary of the data obtained in vitro and in cells. ....	189
<b>Table 3.3. 1</b> Summary table of quantified immunoblot band intensities representing the range of degradation in Log2 Fold Change relative to DMSO using either the dTAG or iTAG system.....	221

---

## Abbreviations

### A

AAVS1	Adeno-associated virus integration site 1
ADMET	Absorption, Distribution, Metabolism, Excretion, Toxicity tolerance
AID	Auxin Inducible Degron
AID2	Auxin Inducible Degron version 2
AdPROM	Affinity-directed Protein Missile
AML	Acute myeloid leukaemia
AMP	Adenosine monophosphate
AmpR	Ampicillin resistance gene
APF-1	Adenosine triphosphate-dependent proteolytic factor-1
APOBEC1	Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 1
AREs	Auxin responsive elements
ARF	Auxin response factor
ATP	Adenosine triphosphate
A20/TNFAIP3	Tumour necrosis factor alpha-induced protein 3

### B

BCA	Bicinchoninic acid
BRD4	Bromodomain-containing protein 4
BP	Base pair
BSA	bovine serum albumin

**C**

CAMKII $\alpha$	Calcium- calmodulin–activated protein kinase II $\alpha$
CAR	Chimeric antigen receptor
Cas	Clustered Regularly Interspaced Short Palindromic-associated protein
Cdc42	Cell division control protein 42 homolog
CDK1	Cyclin-dependent kinase 1
CDK9	Cyclin-dependent kinase 9
CELMoD	Cereblon E3 ligase modulation drug
CFTR	Cystic fibrosis transmembrane conductance regulator
ChIP-seq	Chromatin immunoprecipitation sequencing
cIAP1	Cellular inhibitor of apoptosis protein 1
CK1 $\alpha$	Casein kinase isoform $\alpha$
Cppt/CTS	Central polypurine tract/central termination sequence
CRBN	Cereblon
CREB1	cAMP responsive element binding protein 1
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRa	Clustered Regularly Interspaced Short Palindromic Repeats activation
CRISPRi	Clustered Regularly Interspaced Short Palindromic Repeats interference
crRNA	Clustered Regularly Interspaced Short Palindromic Repeats-derived ribonucleic acid
CUL	Cullin
CYP21A2	Cytochrome P450 family 21 subfamily A member 2

**D**

---

DAPI	4',6-diamidino-2-phenylindole
DCAF	DDB1-CUL4 associated factor
dCas9	Dead Clustered Regularly Interspaced Short Palindromic-associated protein 9
DCD	Degron-containing domain
DCN1	Defective in cullin neddylation protein 1
DD	Destabilising Domain
DDB1	Damaged DNA binding protein 1
DDB2	Damaged DNA binding protein 2
deGradFP	degrade Green Fluorescent Protein
DGCR8	DiGeorge syndrome critical region gene 8
DHC1	Dynein heavy chain 1
DLBCL	Diffuse large B-cell lymphoma
DMEM	Dulbecco Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DOX	Doxycycline
dsRBP	Double-stranded ribonucleic acid-binding protein
dsRNA	Double-stranded ribonucleic acid
DSB	Double-strand break
DTT	Dithiothreitol
DUB	Deubiquitinating enzyme
<b>E</b>	
EED	Embryonic Ectoderm Development
EGF	Epidermal growth factor

---

---

EGFP	Enhanced green fluorescent protein
ERK	Extracellular signal-regulated kinase
ERK1	Extracellular signal-regulated kinase 1
EZH2	Enhancer of zeste 2 – polycomb repressive complex 2 subunit

**F**

f1 Ori	phage-derived origin of DNA replication
FACS	Fluorescence-activated cell sorting
FAM83D	Family With Sequence Similarity 83 Member D
FBDD	Fragment-based drug discovery
FCS	Foetal calf serum
FG	Ferrite-glycidyl methacrylate
FKBP12	FK506- and rapamycin binding protein
Fluc	Firefly luciferase
FP	Fluorescence polarisation

**G**

GFP	Green fluorescent protein
GluRIIA	Glutamate receptor IIA
GmR	Gentamicin resistance gene
GS	Glutamine synthetase
GR	Glucocorticoid receptor
gRNA	Guide ribonucleic acid
GSK-3 $\beta$	Glycogen synthase kinase 3 $\beta$
GSPT1	G1 to S Phase Transition 1

**H**

HA	Human influenza hemagglutinin
HBD	Helical bundle domain
HCV	Hepatitis C virus
HDAC1	Histone deacetylase 1
HDR	Homology directed repair
HECT	Homologous to the E6AP carboxyl terminus
HEK 293T	Human embryonic kidney 293T
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HERC4	HECT and RLD domain containing E3 ubiquitin protein ligase 4
HITI	Homology-independent targeted integration
HMEC	Human mammary epithelial cells
hpf	Hours-post-fertilization
HPRT1	Hypoxanthine phosphoribosyltransferase 1
HRV	human rhinovirus
h-TERT	Human telomerase reverse transcriptase
HTS	High-Throughput Screening
HyT13	Hydrophobic tag 13
H3K27me1	mono-methylated lysine 27 on histone H3
H3K27me2	di-methylated lysine 27 on histone H3
H3K27me3	tri-methylated lysine 27 on histone H3

**I**

IAA	Indole acetic acid
-----	--------------------

---

i-AID	Improved auxin-inducible degron
I $\kappa$ B $\alpha$	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
IKK	I $\kappa$ B kinase
IL-2	Interleukin-2
IMiD	Immunomodulatory imide drug
Ips	Induced pluripotent stem
IPTG	Isopropylthio- $\beta$ -galactoside
IRES	Internal ribosome entry site
IRF4	Interferon regulatory factor 4
iTAG	Inducible and Targeted protein deGradation

**J**

JAMM	Jab1/MPN/MOV34
------	----------------

**K**

KI	Knock-in
KO	Knock-out
KRAB	Krüppel-associated box

**L**

L-AdPROM	ligand-inducible AdPROM
LB	Lysogeny Broth
LID	Ligand-induced degradation
LTR	long terminal repeat

**M**



---

MCS	Multiple cloning site
MAPK	Mitogen-activated protein kinase
MAPK9	Mitogen-activated protein kinase 9
MAVS	Mitochondrial antiviral-signalling protein
MED14	Mediator of RNA polymerase II transcription subunit 14
MEK1/2	Mitogen-activated protein kinase kinase 1/2
MELK	Maternal embryonic leucine zipper kinase
Meis2	Myeloid ecotropic insertion site 2
MINDYs	Motif interacting with Ubiquitin-containing novel DUB family
miRNA	Micro-ribonucleic acid
MJDs	Machado-Josephin Domain proteases
MM	Multiple myeloma
MMEJ	Microhomology-mediated end-joining
MOI	Multiplicity of infection
mRNA	Messenger ribonucleic acid

**N**

NAA	1-naphthaleneacetic acid
NAE	NEDD8-activating enzyme
NEDD4	Neuronal precursor cell-expressed developmentally downregulated 4
NF-Kb	Nuclear factor Kb
NHEJ	Non-Homologous End Joining
NLS	Nuclear localisation signal
NMR	Nuclear Magnetic Resonance
NS3pro	Non-structural protein 3 protease

**O**

Ori Origin of DNA replication

OUT Ovarian Tumour Proteases

**P**

PAM Protospacer Adjacent Motif

PB1 Phox and Bem1

PBS Phosphate-buffered saline

PCR Polymerase chain reaction

PIKES Protein interaction kinetics and estimation of stoichiometries

PITCh Precise Integration into Target Chromosome

PK/PD Pharmacokinetic/ pharmacodynamic

PLK1 Polo-like kinase 1

PLZF Promyelocytic leukaemia zinc finger

PRC2 Polycomb repressive complex 2

pre-miRNA Precursor micro-ribonucleic acid

pri-miRNA Primary micro-ribonucleic acid

PROTACs PROteolysis Targeting Chimeras

PSD95 Postsynaptic density protein 95

PTEN Phosphatase and tensin homolog

PVDF Polyvinylidene Difluoride

5-Ph-IAA 5-phenyl-indole-3-acetic acid

**Q**

**R**

Rac1	Ras-related C3 botulinum toxin substrate 1
RBBP4	RB Binding Protein 4
RBBP7	RB Binding Protein 7
RBM39	RNA-binding protein 39
RBS	Ribosome binding site
RBX1/ROC1	RING box-domain 1/ Regulator of Cullins 1
RhoA	Ras homolog family member A
Rhox3	Reproductive homeobox 3
RING	Really interesting new gene
RIP1	Receptor interacting protein kinase 1
RISC	Ribonucleic acid-induced silencing complex
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RPMI 1640	Roswell Park Memorial Institute 1640
RRE	Rev response element
RT	Room Temperature

**S**

SALL4	Spalt-like transcription factor 4
SAR	Structure affinity relationship
SATI	Single homology Arm donor mediated intron-Targeting Integration
SCF	Skp1-cullin-Fbox
SD	Standard deviation

---

SDS	Sodium Dodecyl Sulfate
SEC	Super elongation complex
SGK3	Serum and glucocorticoid kinase-3
sgRNA	Single guide ribonucleic acid
shRNA	short-hairpin ribonucleic acid
SILAC	Stable isotope labelling of amino acids in cell culture
siRNA	small/short interfering ribonucleic acid
Slmb	supernumerary limbs
SMASh	Small Molecule-Assisted Shutoff
SMS	Spermine synthase
SMURF1	SMAD specific E3 ubiquitin protein ligase 1
SpCas9	Streptococcus pyogenes Cas9
ssODN	single-stranded DNA oligodeoxynucleotide
ssRNA	single-stranded ribonucleic acid
STR	Short tandem repeat DNA analysis
SV5	Simian virus 5
SUZ12	Suppressor Of Zeste 12 Protein Homolog
SYK	Spleen tyrosine kinase

**T**

TB	Terrific Broth
TBS-T	Tris- buffered saline containing 0.1% Tween20
TCEP	Tris(2-carboxyethyl)phosphine
TIR1	F-box transport inhibitor response 1
TLR3	Toll-like receptor 3

TNF $\alpha$	Tumour necrosis factor $\alpha$
tracrRNA	trans-activating Clustered Regularly Interspaced Short Palindromic Repeats-derived ribonucleic acid
TRAF4	Tumour necrosis factor receptor associated factor 4
TR-FRET	Time resolved fluorescence energy transfer
TRIF/TICAM1	TIR-domain containing adapter-inducing interferon- $\beta$

## U

UCHs	Ubiquitin C-terminal Hydrolases
ULK1	Unc-51 like autophagy activating kinase
UPS	Ubiquitin-Proteasome System
USPs	Ubiquitin-Specific Proteases
UTR	Untranslated region

## V

VCP	Valosin-containing protein
VHL	von-Hippel-Lindau
VPS34	Class III PI 3-kinase

## W

WPRE	Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element
WT	Wild type
WWP2	WW domain containing E3 ubiquitin protein ligase 2

## X

**Y**

YFP            Yellow fluorescent protein

**Z**

ZBTB16        Zinc finger and bric à brac, tramtrack, and broad 16

ZF             Zinc finger

ZFP145        Zinc finger protein 145

ZFP91         Zinc finger protein 91

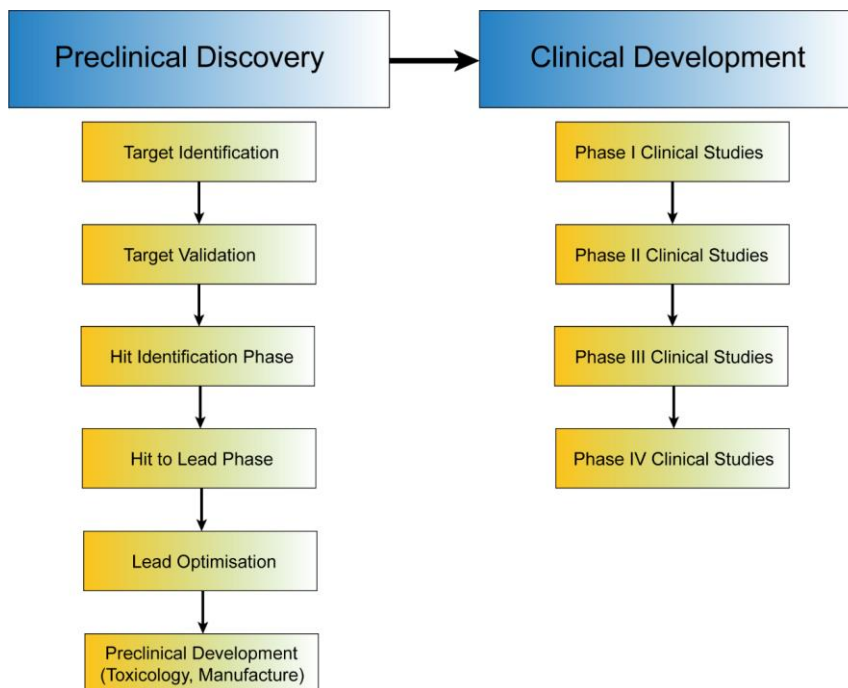
ZUP1          Zinc finger-containing ubiquitin peptidase 1

# 1. INTRODUCTION

## 1.1 *The process of drug discovery*

A drug discovery programme aims to tackle the unmet clinical need of a particular medical condition or disease through the identification of chemical or biological entities (e.g. peptides, antibodies), and subsequently, their development as novel therapeutic agents (Hughes *et al.*, 2011). The modern drug development process is comprised of preclinical and clinical stages (**Figure 1.1**). During the stage of clinical development, the safety and efficacy of a new drug candidate is examined in clinical trials. Clinical trials are typically classified into four phases (Phase I, II, III and IV) (Atkinson and Clark, 2016). In Phase I, several pharmacokinetic parameters and tolerance of the drug candidate are studied, generally in healthy volunteers; with the exception of many cytotoxic oncology drug candidates which are typically tested in cancer patients (Ananthakrishnan and Menon, 2013). Phase II studies are conducted to determine the initial efficacy and potential side effects of the drug candidate in small cohorts of healthy volunteers or patients, as is the case with oncology drugs. Phase III studies are large-scale clinical trials for safety and efficacy in large patient populations, usually randomised to compare the novel intervention to the standard of care, before the drug candidate is approved by the relevant drug regulatory agencies as a novel treatment for a particular disease. Phase IV studies mainly involve

monitoring of the new drug in the market for any long-term benefits or side-effects (Atkinson and Clark, 2016). The preclinical stage of drug discovery broadly includes the following key phases: 1) target identification, 2) target validation, 3) hit identification, 4) hit to lead, 5) lead optimisation, 6) selection of a candidate molecule to progress for preclinical evaluation and finally 7) preclinical development (toxicology, manufacture) (**Figure 1. 1**) (Hughes *et al.*, 2011). For the purposes of this thesis, the preclinical stage of drug discovery will form the main focus of the rest of this subchapter.





**Figure 1. 1 The process of drug discovery.** Summary of the key steps in target-based preclinical discovery prior to the four major phases of clinical development.

The identification of a novel drug target forms a critical primary step in the drug discovery process. A target constitutes a broad term and it could refer to anything from a gene, RNA (ribonucleic acid) or protein implicated with a disease state. Targets are often discovered in basic research, for example, by studying the normal or abnormal function of a protein, an aberrant signalling pathway or a mutation in a gene that is connected to a specific disease (Zheng, Thorne and Mckew, 2013). An ideal drug target would have been directly linked through multiple orthogonal experimental procedures with a pathophysiological condition and characterised as “druggable”; i.e. target is accessible to a drug molecule and upon binding, a therapeutic biological response is elicited. However, not all drug discovery programmes begin with an identified target. High-throughput cell-based phenotypic assays are commonly used to screen large compound libraries and identify lead compounds that ameliorate the disease phenotype through unprecedented drug mechanisms (Zheng, Thorne and Mckew, 2013; Heilker, Lessel and Bischoff, 2019). There are several key advantages to the use of phenotypic screens in drug discovery, as they both allow the identification of pharmacologically active molecules with favourable physicochemical properties, balanced between cellular permeability and hydrophilicity, and can also unearth novel biological pathways from the use of unbiased cellular systems (Ege *et al.*, 2021). Indeed, target-agnostic phenotypic screening has an advantage over target-based drug discovery of better capturing the complexity of biological systems, since inherent bias for specific targets is removed (Moffat, Rudolph and

Bailey, 2014; Croston, 2017). Nevertheless, phenotypic screening also entails considerable challenges in target deconvolution, validation and optimisation of hit compounds due to the hit compounds' polypharmacology which complicates optimisation of compound properties compared to compounds acting through a single molecular target (Moffat *et al.*, 2017; Heilker, Lessel and Bischoff, 2019). Contrary to phenotypic screening, target-based approaches are based on known molecular targets and are generally simpler and faster to develop and execute. The knowledge of a molecular target also facilitates the optimisation of hit compounds, as research efforts can utilise computational modelling, mutational analysis, biochemistry, binding kinetics, molecular pharmacology and crystallography to ascertain the interaction of a candidate drug and a target molecule, further enabling the development of efficient SAR (structure-affinity relationship) (Croston, 2017).

Once a target is identified, the contribution of that target in a particular disease setting has to be fully validated. Target validation is defined as the process in drug discovery by which the function and relevance of a target in a disease is assessed, prior to selection as a therapeutic candidate. Target validation tools range from *in vitro* assays to the use of animal models (Hughes *et al.*, 2011). Modulation of target expression to understand its relevance and function in a disease is an important step of target validation. To that extent, tools like RNAi (RNA interference) technologies and CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated protein) genome editing are widely used by the research community. Both genetic methods will be reviewed in depth in Introduction Subchapter 1.3. In short, RNAi technologies utilise RNA-like chemical oligonucleotides complimentary to the

mRNA (messenger ribonucleic acid) of a target gene (Boettcher and Mcmanus, 2015). Binding of these antisense nucleotides to a target mRNA prevents translation leading to gene downregulation. On the other hand, CRISPR/Cas genome editing is a technique based on an antiviral prokaryotic immune system that works by inducing double strand breaks at a desired location of a cell's genome, thus allowing existing genes to be removed and/or new ones to be added if a template DNA (deoxyribonucleic acid) is provided (Boettcher and Mcmanus, 2015). In summary, target validation forms a critical step in drug discovery, with tools such as RNAi and CRISPR/Cas genome editing employed to verify the therapeutic potential of a target before it proceeds down the drug development process.

Once a target has been identified and validated, compound libraries are screened to identify hit compounds during the hit identification phase. Hit compounds can be defined as molecules that are identified to interact with the target and are therefore potential candidates for drug discovery. HTS (High-Throughput Screening) is one screening paradigm, where large compound libraries are assayed for potential biological activity against a target implicated with a disease state (Armstrong, 1999). In HTS, screening can be in the format of high-throughput target-based assays where no previous knowledge is required of what pharmacophores could elicit a biological response following binding to a target (Hughes *et al.*, 2011). Typically, HTS compound libraries contain small molecular weight compounds that have been selected to obey certain chemical and physical properties that increase the likelihood of high oral absorption and exhibit a high degree of chemotype variability (Chessari and Woodhead, 2009; Lipinski *et al.*, 2001). On the contrary, in focused or knowledge-based screening,

small subsets of molecules from compound libraries based around chemical classes that have been previously shown to have activity against a target are selected and tested (Hughes *et al.*, 2011). An alternative approach in drug discovery is FBDD (fragment-based drug discovery), which involves the generation and testing of fragments, i.e. highly soluble, moderately lipophilic, and low molecular weight compounds. FBDD aims at identifying lead compounds that weakly interact with the drug target (in the  $\mu\text{M}$  to  $\text{mM}$  range) (Chessari and Woodhead, 2009; Kloe *et al.*, 2009). Although these initial lead compounds might weakly bind to the target, they will be further developed by medicinal chemistry optimisation (Whittaker *et al.*, 2010). Compared to more conventional HTS methods, fragment libraries cover a significantly broader chemical space which is reflected in the larger number of "hits" identified (Hajduk and Greer, 2007).

The hit to lead phase describes the process in drug discovery by which hit compounds are further optimised and combined with other chemical moieties to generate lead compounds of enhanced potency and selectivity (Hughes *et al.*, 2011). A lead compound is a chemical compound of suboptimal structure to fit the target binding site but, which has been demonstrated to exhibit pharmacological activity, likely to be therapeutic. During this stage of drug discovery, intensive and systematic SAR studies are performed around the hit compounds' core structure; in order to improve potency, selectivity and pharmacokinetic parameters (Hughes *et al.*, 2011). If structural information about the target is available, methodologies such as X-ray crystallography and NMR (Nuclear Magnetic Resonance) as well as computational modelling can be applied in developing a SAR (Hughes *et al.*, 2011; Ferreira and Andricopulo, 2019). In parallel, a series of studies aim at validating the activity of lead

compounds in in vivo animal disease models and in generating a PK/PD (pharmacokinetic/pharmacodynamic) profile. PK/PD studies at this point are generally carried out in mouse or rat models and are critical, as they allow for an early evaluation of both the therapeutic potential and the maximum half-life of lead compounds, their metabolism as well as any adverse toxicity effects (Hughes *et al.*, 2011).

Following from the hit to lead phase, lead optimisation forms the final phase in preclinical drug discovery. The object of this stage is to improve deficiencies in the structure of lead compounds whilst, maintaining favourable pharmacological and therapeutic properties (Hughes *et al.*, 2011). An important process at this stage, is the profiling of physicochemical properties of lead compounds as well as the characterisation of their ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity tolerance) properties (Ferreira and Andricopulo, 2019). In parallel, a critical component at this stage is the continued generation and testing of novel compounds as back-up molecules in case the lead compounds undergoing further preclinical characterisation fail. Compounds that meet the criteria of this phase can be selected to be further advanced as preclinical drug candidates and if successful, progressed to clinical trials as explained above. During preclinical development, selected candidate compounds undergo rigorous toxicology testing and also reliable methods for large scale synthesis of the drug molecules are developed (Hughes *et al.*, 2011). Finally, a drug candidate that successfully passes both the preclinical and clinical stages of drug discovery can be marketed as a novel therapeutic and in some diseases even becoming the new standard of care.

## **1.2 Pitfalls of target validation in cancer drug discovery**

Academic laboratories are large contributors to the identification of disease-causing targets (Kaelin, 2017). Putative disease-driving targets published from academia are often further explored by pharmaceutical and biotechnology companies as potential novel therapeutic targets which, when modulated by a chemical compound or a biological agent (e.g. peptide, antibody) would lead to a favourable clinical outcome (Kaelin, 2017). However, several reports recently published have addressed the lack of reproducibility and insufficient robustness in findings from such exploratory studies (Prinz, Schlange and Asadullah, 2011; C. Glenn Begley and Lee M. Ellis, 2012; Errington *et al.*, 2014; Kaelin, 2017; Miyakawa, 2020). Some of these reports state that in the field of cancer research, only about 11% of published data on putative target proteins could be validated by pharmaceutical and biotechnology labs (Prinz, Schlange and Asadullah, 2011; C. Glenn Begley and Lee M. Ellis, 2012; Miyakawa, 2020). As a consequence, even though pharmaceutical and biotechnology companies are pursuing in-depth biological understanding of putative therapeutic targets at the target validation phase, attrition rates during early clinical development remain particularly high (Hughes *et al.*, 2011; Prinz, Schlange and Asadullah, 2011; C. Glenn Begley and Lee M. Ellis, 2012; Miyakawa, 2020).

Evidently, due to this lack of reproducibility and insufficient robustness of exploratory data, the process of target validation in drug discovery has been

becoming increasingly important in assessing the therapeutic potential of a target in a disease; prior to initiating a drug discovery program and heavily investing in assay development, high-throughput screening campaigns, lead optimization and animal testing (Prinz, Schlange and Asadullah, 2011). Currently, in vitro/ in vivo target manipulation using genetic methods, such as RNAi and CRISPR/Cas genome editing, or chemical compounds that inhibit or modulate protein abundance are widely employed to validate targets (Kaelin, 2017). However, the use of these methods in target validation harbour some common pitfalls, as are explained by Nobel laureate William Kaelin in his recent review (Kaelin, 2017). To begin with, William Kaelin states that one of the main problems is that most target validation phenotypic assays are often configured as “down” rather than “up” assays (Kaelin, 2017). In his review, down assays are defined as those that measure decreased readouts, e.g. decreased cell proliferation or viability, whilst up assays are designed for measuring an increase in a readout, e.g. increased cell death or enzymatic activity (Kaelin, 2017). William Kaelin argues that down assays are limited by off-target effects due to perturbations of cellular homeostasis and that biological insights obtained from up assays suffer significantly less from biologically uninteresting or false-positive results, as it is more difficult to improve the performance of a complex system, such as the cellular milieu (Kaelin, 2017). Since most target validation assays are designed as down assays, it is almost impossible to disentangle whether the effects of target perturbation are due to on-target-effects, or off-target activity that decreases cell viability, or a combination of both. Therefore, to distinguish observed phenotypes that result from perturbations of the target under investigation and attribute a causal relationship, it is important to incorporate up

assays or to include “rescue” experiments in target validation assays (i.e. introduction of an expression vector harbouring target cDNA; washout of the compound inhibiting target protein function or modulating protein abundance) (Kaelin, 2017). Similarly, validating in vitro assay observations with in vivo experiments would further strengthen conclusions regarding the role of the target in a particular disease setting and would also increase the likelihood that these conclusions are correct and robust (Kaelin, 2017).

As stated above, current genetic methods such as RNAi and CRISPR/Cas genome editing are widely used in the drug discovery stage of target validation (Kaelin, 2017). The following section gives a review of these genetic methods and outlines the main advantages and disadvantages in their use as target validation tools.



## **1.3 Overview of genetic methods in target validation**

### **1.3.1 siRNA and shRNA methods**

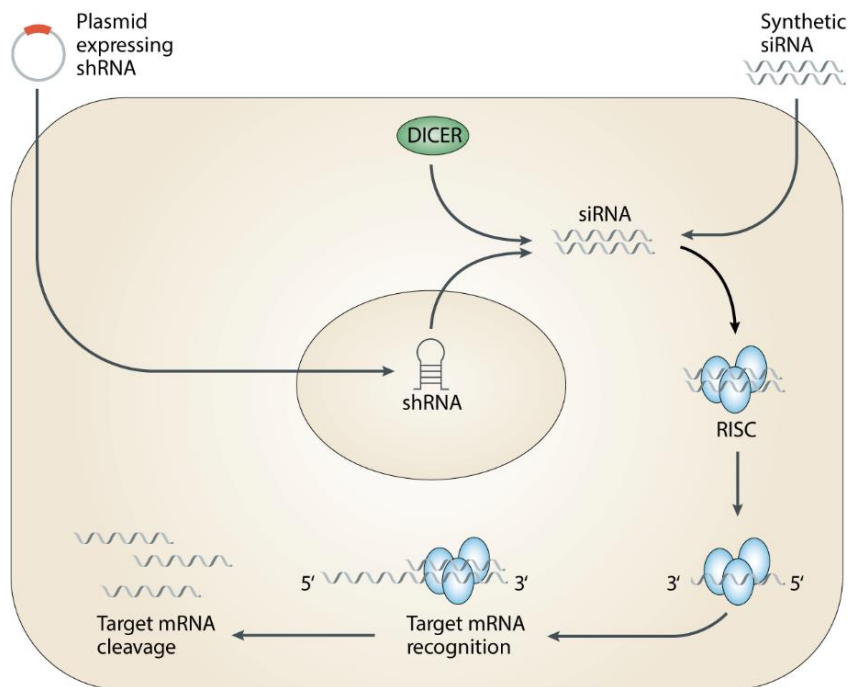
Found in three life kingdoms (*Fungi*, *Plantae* and *Animalia*), RNAi constitutes an evolutionarily conserved cellular mechanism for sequence-specific suppression of gene expression at the post-transcriptional level (Iorns *et al.*, 2007). RNAi is postulated to have evolved as a defence mechanism in cells against viruses or transposon elements, which utilise dsRNA (double-stranded ribonucleic acid) molecules for hijacking the cellular machinery and propagating their genomes (Iorns *et al.*, 2007). In RNAi, a dsRNA molecule triggers the specific cleavage of a complementary mRNA molecule via an endogenous cellular machinery (Watts, Deleavey and Damha, 2008). The phenomenon of RNAi was first observed in the early 1990s by Richard Jorgensen and colleagues, where they exogenously overexpressed a chimeric chalcone synthase into petunia plant (*Petunia hybrida*) and observed a block in the biosynthesis of anthocyanin, resulting in altered colouration patterns of petunia petals (Napoli, Lemieux and Jorgensen, 1990). However, the mechanism of RNAi was only described back in 1998 by Craig Melo, Andrew Fire and colleagues, by injecting dsRNA molecules against a specific set of genes in the nematode worm *Caenorhabditis Elegans* (Fire *et al.*, 1998). The group identified that the use of dsRNA against those genes produced a pronounced decrease in the mRNA

levels of each respective target and that this loss recapitulated the phenotypes observed with previous loss-of-function mutants (Fire *et al.*, 1998). Shortly thereafter, another group in 2001, discovered that synthetic short (21 nucleotides) dsRNA molecules could be applied for gene knockdown in mammalian cells (Elbashir *et al.*, 2001). Currently, this intrinsic RNAi cellular machinery is widely exploited by the scientific community, in order to study the functions of target proteins in mammalian cells, as it allows for gene knockdown by introducing to the cells exogenous non-coding RNA molecules (Watts, Deleavey and Damha, 2008).

In mammalian cells, the process of RNAi can be regulated via either miRNA (micro-ribonucleic acid) or siRNA (small/short interfering ribonucleic acid) molecules, which are characterised by the double-stranded nature of their precursors (Carthew and Sontheimer, 2009). miRNAs primarily act as regulators of endogenous genes, whilst, siRNAs are implicated in genome integrity by mounting a response against foreign RNA from either viruses, transposons or transgenes (Carthew and Sontheimer, 2009). miRNAs are endogenous small non-coding RNA-molecules (~20-30 nucleotides in length) that base-pair with sequence motifs in the 3'-UTR (three-prime untranslated region) of target mRNAs with perfect or near perfect complementarity and inhibit their translation or act to promote their degradation (Cai *et al.*, 2009). Certain miRNAs have also been described to negatively regulate translational expression of target mRNAs by binding at their 5'-UTR (five-prime untranslated region) (Cai *et al.*, 2009). miRNAs are encoded from polycistronic transcripts (more than 1000 nucleotides long) called pri-miRNAs (primary miRNAs) composed of double-stranded hairpins with 5' and 3' overhangs (Wilson and Doudna, 2013). siRNAs are non-

coding dsRNA molecules (~20-30 nucleotides in length) derived from long dsRNA molecules that can either arise from endogenous genomic loci, be introduced directly into the cytoplasm by transfection methods or taken up as foreign nucleic acids from the environment (viruses, transposons, transgenes) (**Figure 1. 2**) (Valencia-sanchez *et al.*, 2006; Carthew and Sontheimer, 2009). The microprocessor complex, which is primarily constituted by the RNAase III Drosha and the dsRNA-binding protein DGCR8 (DiGeorge syndrome critical region gene 8), crops pri-miRNAs into ~65-70 nucleotide long stem-loop structures called pre-miRNAs (precursor-miRNAs) (Wilson and Doudna, 2013). Association of pre-miRNAs with transport facilitators Exportin-5 and RanGTP shuttles them to the cytoplasm, where the processing pathways converge for both miRNAs and siRNAs (**Figure 1. 2**). Then both precursors are further trimmed by a Dicer enzyme down to ~21-25 nucleotides long dsRNA molecules suitable for loading to an Argonaute protein. A dsRBP (dsRNA-binding protein) can aid Dicer enzyme in the process and along with an Argonaute protein they constitute a minimal RISC (RNA-induced silencing complex) (**Figure 1. 2**) (Wilson and Doudna, 2013). Once the dsRNA helix is presented to an Argonaute protein the dsRNA molecule is separated into two ssRNA (single-stranded RNA) molecules, the passenger strand that is discarded and the guide strand which remains bound to Argonaute. The RISC can base-pair with ssRNA molecules, such as mRNA, that are complementary to the Argonaute-bound guide strand (**Figure 1. 2**). Binding to the target initiates via nucleotides 2-6 of the guide strand, which are referred to as the seed sequence and in cases where the Argonaute protein bears catalytic activity this binding results to target cleavage, thereby preventing translation (Wilson and Doudna, 2013). One of the key

differences between miRNA and siRNA-mediated transcriptional regulation is that an siRNA will only bind to a single mRNA target through a perfect complementarity of base-pair matching whilst, miRNAs can inhibit the translation of various mRNA sequences through imperfect pairing (Wilson and Doudna, 2013).



**Figure 1. 2 RNAi (RNA interference) cellular mechanism.** siRNAs (small-interfering RNAs) can be chemically synthesized and introduced directly into cells using transfection reagents or electroporation. Alternatively, plasmid-encoded shRNAs (short-hairpin RNAs) are processed by the RNase III-like enzyme, DICER, into siRNA duplexes that are 21–28 nucleotides in length with dinucleotide 3' overhangs. Both siRNAs and processed shRNAs are then incorporated into the multiprotein endoribonuclease RISC (RNA-induced silencing complex). A helicase within RISC

unwinds duplex siRNA allowing its antisense strand to bind mRNA through perfect complementarity of base-pair matching. An RNase within RISC (e.g. argonaute) degrades the target mRNA by cleavage, which results in silenced gene expression and reduced protein production. (Taken from Iorns *et al.*, 2007)

In terms of utilizing RNAi to study the function of target proteins, chemically synthesized double-stranded siRNAs are primarily employed due to their specificity towards a single mRNA target and the fact that they do not induce an inhibitory interferon response which can result in the global shutdown of protein synthesis (Iorns *et al.*, 2007). siRNAs can be directly introduced into cells by transfection methods or endogenously generated by transducing cells with vectors expressing shRNA (short-hairpin RNA) molecules, precursors of siRNA (**Figure 1. 2**) (Iorns *et al.*, 2007). shRNAs are artificial non-coding ssRNA molecules that fold into a hairpin-shaped structure, encompassing a stem region of dsRNA and a loop region of ssRNA (Paddison *et al.*, 2002). Even though their structure resembles that of pre-miRNAs, shRNAs are either introduced into cells by transfection methods or expressed within cells by virally-integrated vectors (lentiviral and retroviral-based) (Watts, Deleavey and Damha, 2008). Once inside the cell, shRNA molecules are also processed in the cytoplasm by the Dicer enzyme to form siRNAs which associate with RISC and modulate transcriptional regulation of target mRNAs (**Figure 1. 2**) (Mikuma *et al.*, 2004). Contrary to siRNAs which transiently knockdown target expression, genetically-expressed shRNAs can be employed for inducing stable and heritable gene silencing (Paddison *et al.*, 2002).

One of the major advantages of RNAi technology is that the molecular machinery described above is ubiquitously expressed in mammalian cells, thus

not requiring prior genetic manipulation of target cell line (Boettcher and Mcmanus, 2015). Indeed, a simple siRNA transfection can result in the downregulation of a target mRNA, in a relatively cost-effective process that allows to study the effects of loss-of-target function. Moreover, large libraries of siRNA and shRNA molecules have been developed since the discovery of RNAi by various groups which had transformed at the time target identification and validation in drug discovery (Iorns *et al.*, 2007). Nevertheless, since the application of RNAi technology in the study of target function, several reports have raised the issue of off-target effects, mediated in part from binding to non-target mRNA molecules through partial complementarity (Jackson *et al.*, 2006; Sigoillot and King, 2011). In addition, the RNAi technology has been shown to cause non-sequence specific off-target effects by displacing endogenous miRNAs from the RISC complex, thus leading to an altered transcriptome and consequently altered phenotypes (Khan *et al.*, 2009). Nonetheless, it should be stated that over the years several machine learning algorithms have been developed, which allowed for the design of efficient and accurate siRNAs for target protein knockdown, but despite their implementation, there are still drawbacks (Shabalina, Spiridonov and Ogurtsov, 2006; He *et al.*, 2017). Another issue of RNAi technologies is the incomplete knockdown of a target protein, which could convolute observed phenotypes and lead to erroneous conclusions regarding the function of the target protein (Sigoillot and King, 2011; Boettcher and Mcmanus, 2015). Lastly, the use of RNAi technologies to knockdown targets in vivo, in a cell specific and inducible manner remains a challenging process. For systemically-administered siRNA molecules, adsorption, distribution, metabolism, and excretion are significant obstacles (Paroo and Corey, 2004).

Indeed, siRNA molecules exhibit a relatively large molecular weight and high negative charge which leads to poor cellular uptake, and a number of other biological barriers, including rapid blood clearance due to degradation by ribonucleases and renal extraction, uptake by the reticuloendothelial system and aggregation with serum proteins (Paroo and Corey, 2004; Kim *et al.*, 2009). Moreover, the administration of siRNA in vivo has been shown to elicit toxicities, such as unwanted activation of the innate immune system via both off-target silencing and the induction of interferon responses (Kim *et al.*, 2009). Even though shRNAs are not affected by these limitations, as they are retrovirally integrated into the cell genome and therefore endogenously expressed in the cells of interest in vivo, they have been demonstrated to induce toxicity effects from their prolonged expression and suffer from unintended off-target effects in vivo (Kim *et al.*, 2009). Indeed, in a study of targeting *Rhox3* (reproductive homeobox 3) gene using shRNA molecules in male germ cells in vivo, it was shown that shRNA molecules were disrupting spermatogenesis independently of knocking down *Rhox3*, highlighting the need for controls when performing shRNA-mediated knockdown-studies and in extend employing the RNAi technology (Song *et al.*, 2015).

### **1.3.2 CRISPR-based systems**

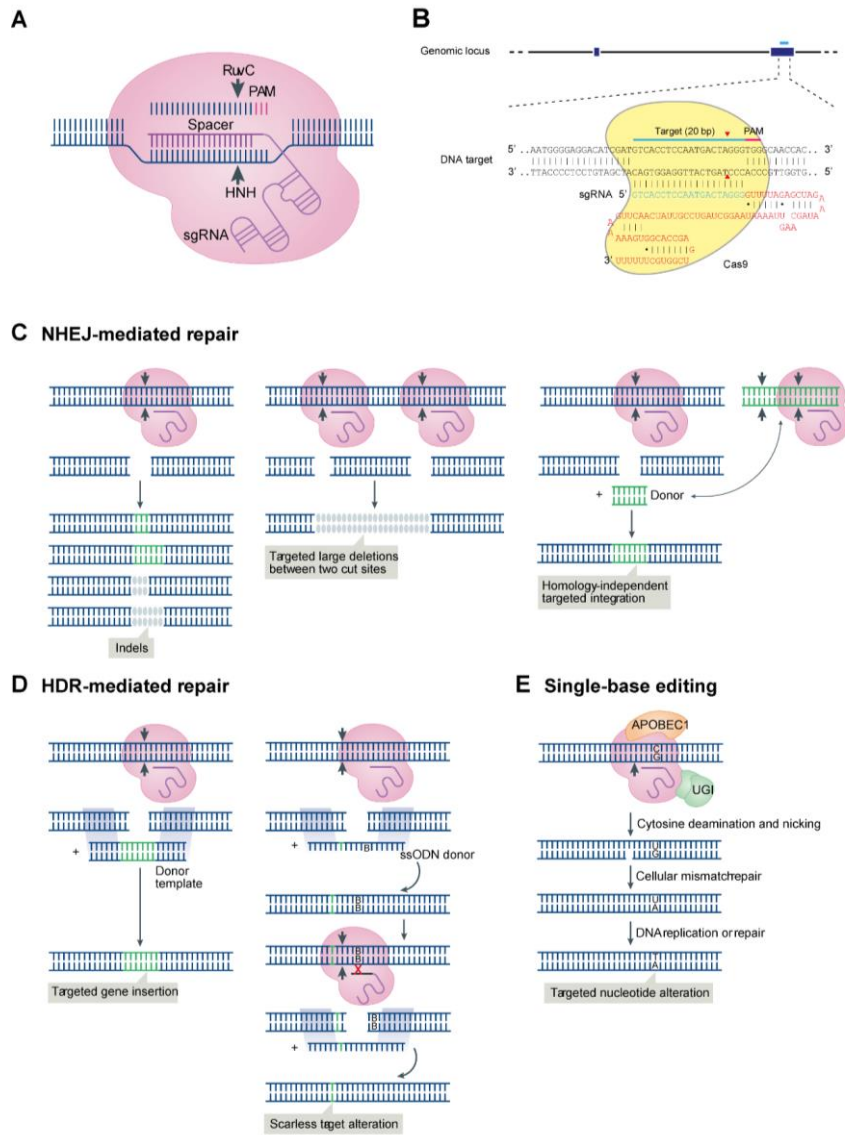
CRISPR/Cas is an RNA-guided genome editing technology that can be employed to generate cuts at the genomic locus of a target protein, resulting in either a loss or gain of function (Fennell *et al.*, 2014). The CRISPR/Cas system was initially discovered in bacteria and archaea as an adaptive immune system, mediating protection against foreign invading DNA, such as that of a bacteriophage. In nature, the cascade of events initiates when short fragments of foreign DNA from either plasmids or viruses are inserted by Cas endoribonucleases into the bacteria's or archaea's chromosome as new spacer sequences next to repetitive sequences (30-40 bp long), known as CRISPR (Wiedenheft, Sternberg and Doudna, 2012; Charpentier and Marraffini, 2014; Rath *et al.*, 2015). The genomic regions harbouring these CRISPR repeat-spacer arrays are termed CRISPR loci. From these loci, the repeat-spacer arrays are transcribed into long primary transcripts which are subsequently processed into short crRNAs (CRISPR-derived RNAs) by Cas endoribonucleases (Wiedenheft, Sternberg and Doudna, 2012). These crRNAs then associate with Cas endoribonucleases forming RNA-protein complexes and are used as guides to recognise invading DNA sequences. Following base-pair binding of the crRNA-Cas complex to an invading DNA sequence, the Cas endoribonuclease cleaves both strands of the invading DNA sequence within the region that is complementary to the crRNA, thus mediating immunity to the bacteria or archaea (Wiedenheft, Sternberg and Doudna, 2012).



Thus far, on the basis of the Cas gene content six types of CRISPR systems have been identified in bacteria and archaea, with the type II system being one of the best characterised (Ran *et al.*, 2013; Pickar-Oliver and Gersbach, 2019). The CRISPR/Cas genome editing technology is primarily based on the type II CRISPR/Cas9 system which consists of the Cas9 endoribonuclease and a CRISPR locus that encodes crRNA and an auxiliary tracrRNA (trans-activating crRNA) (Ran *et al.*, 2013). The tracrRNA base-pairs with crRNA to form a functional gRNA (guide RNA) with a 5' end complementarity to the target gene sequence (Gilbert *et al.*, 2013; Qi *et al.*, 2013). The CRISPR/Cas9 system is ideal for genome editing for two main reasons, first, the Cas9 provides very efficient blunt double-stranded DNA cleavage and second, a minimal set of components are required to form a functional system (Charpentier and Marraffini, 2014). In 2012, Jennifer Doudna's and Emmanuelle Charpentier's laboratories demonstrated that fusion of the 3' end of crRNA to the 5' end of tracrRNA to generate an sgRNA (single guide RNA) can also direct sequence-specific DNA cleavage by the Cas9 (**Figure 1. 3 A-B**) (Jinek *et al.*, 2012). Since then, the CRISPR/Cas9 system has been demonstrated to edit the genome of eukaryotic cells and was further developed into a genome editing tool for studying the functions of target genes in cells, both in vitro and in vivo (Le Cong *et al.*, 2013; Mali *et al.*, 2013; Hsu, Lander and Zhang, 2014). A widely-used Cas9 endonuclease is the SpCas9 (*Streptococcus pyogenes* Cas9) (Ran *et al.*, 2013). An sgRNA will bind to the SpCas9 and upon binding will induce a conformational change to the endoribonuclease, which will assume its active form (**Figure 1. 3 A-B**). Activated SpCas9 binds to genomic sequences that match its PAM (Protospacer Adjacent Motif) sequence, a three base nucleotide sequence (5'-

NGG-3' for Cas9) that is present downstream of sgRNA target gene complimentary region (Jinek *et al.*, 2012; Sternberg *et al.*, 2014). Other less-used Cas9 orthologues may have different PAM sequences such as 5'-NGGNG-3' in *Streptococcus thermophilus* and 5'-NNNNGATT-3' in *Neisseria meningitidis* (where N denotes any DNA base), with these PAM sequence variations being advantageous in certain cases (Ran *et al.*, 2013). Following binding to PAM, Cas9 unwinds nucleotide bases immediately upstream of the PAM motif to facilitate pairing of the sgRNA complimentary region to the target genomic locus (**Figure 1. 3 A-B**). A perfect match between sgRNA and target gene sequence will lead to the activation of RuvC and HNH nuclease domains of Cas9 to induce a double strand cut after the third nucleotide base upstream of the PAM site (Anders *et al.*, 2014).

Eukaryotes predominantly repair these double strand breaks through the error-prone NHEJ (non-homologous end joining) pathway which leads to insertions or deletions of nucleotides at the targeted genomic locus, knocking out the gene of interest (**Figure 1. 3 C**) (Charpentier and Marraffini, 2014; Pickar-Oliver and Gersbach, 2019). Alternatively, if a repair template is provided, the HDR (homology directed repair) pathway is activated, inserting the DNA fragment in the targeted genomic site (**Figure 1. 3 D**) (Bennardo *et al.*, 2008; Pickar-Oliver and Gersbach, 2019). This repair template could be encoding a fluorescent protein or an epitope tag so that when inserted into protein-coding genes, the endogenously expressed fusion proteins could be monitored to study their function in native cellular settings (Pickar-Oliver and Gersbach, 2019). Moreover, in one adaptation of the CRISPR/Cas9 system a D10A mutation in the



**Figure 1. 3 The CRISPR/Cas9 genome editing tool and its different application strategies.**

(A-B) Schematic representations of the *S. pyogenes* Cas9 protein bound to a sgRNA (single guide RNA) molecule via its RuvC and HNH domains. The spacer sequence (20 nucleotides long) of the sgRNA has sequence complementarity to a specific genomic locus adjacent to a 5' PAM (Protospacer Adjacent Motif). Recruitment of Cas9 to that genomic locus results in the generation

of a blunt DNA double-strand break three base pairs upstream of the PAM. **(C-E)** DNA double-strand breaks generated by the CRISPR/Cas9 genome editing tool can be repaired by different DNA repair pathways. **(C)** Error-prone NHEJ (non-homologous end joining)-mediated repair results in nucleotide insertions or deletions. The use of two sgRNAs to target two different genomic loci can result in large deletions during NHEJ repair. Alternatively, addition of a donor DNA sequence that is independently targeted for cutting, can result in its homology-independent genomic integration at the single cut site. **(D)** The addition of either a double-stranded or ssODN (single-stranded DNA oligodeoxynucleotide) with 5' and 3' overhangs complementary to the genomic locus targeted by the Cas9 can result in its genomic integration via the homology-directed repair pathway. Silent mutations along with other intended nucleotide alterations can be introduced in the donor DNA sequence which prevent subsequent target site recognition by the Cas9 protein and formation of NHEJ-mediated indels. **(E)** Cytidine deaminases fused to the Cas9 protein, such as APOBEC1, can be utilized for single nucleotide conversions, either C → T or a G → A. The fusion of two UGIs (uracil glycosylase inhibitors) to the Cas9 nickase increases base-editing efficiency. (Taken from Ran *et al.*, 2013; Pickar-Oliver and Gersbach, 2019)

RuvC or a H840A mutation in the HNH nuclease domains of *SpCas9* are generated. These mutations prevent double-stranded DNA cleavage and instead mediate nicking of the DNA to yield single-stranded breaks, which are preferentially repaired by the high-fidelity HDR pathway (Ran *et al.*, 2013). *SpCas9*<sup>D10A</sup> cleaves the gRNA-targeting strand, while *SpCas9*<sup>H840A</sup> cleaves the non-targeted strand. When used with two sgRNAs targeting different regions of a gene of interest, these Cas9 nickases can also be combined to induce DSB (double-strand break), which can lower the probability of off-target editing (Ran *et al.*, 2013; Pickar-Oliver and Gersbach, 2019). Furthermore, these Cas9 nickases have also been fused with cytidine deaminases for direct conversion of single nucleotides. For instance fusion of cytidine deaminase APOBEC1 (Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 1) to a Cas9 nickase can achieve targeted C → T or G → A nucleotide conversions to generate point-

mutants of target genes for functional evaluation (**Figure 1.3 E**) (Komor *et al.*, 2016; Pickar-Oliver and Gersbach, 2019).

Besides genome editing, the CRISPR/Cas9 technology has also been effectively applied for regulating expression of target genes. In these adaptations a catalytically inactive dead Cas9 (dCas9, double D10A and H840A mutant) is used, which cannot cleave DNA sequences but can still be guided by a sgRNA to the target sequence (Qi *et al.*, 2013; Adli, 2018). The dCas9 can be fused to transcription activators, such as p65 or VP64 (four repeats of the herpes simplex VP16 activation domain), for programmed transcription activation termed CRISPRa (CRISPR activation). Alternatively, the dCas9 can be solely used to bind DNA sequences mediating steric interference and blocking recruitment of the RNA polymerase, in a process termed CRISPRi (CRISPR interference), or fused to transcription repressors, such as KRAB (Krüppel-associated box) which repress transcription by inducing changes in chromatin structure (Gilbert *et al.*, 2013; Qi *et al.*, 2013; Pickar-Oliver and Gersbach, 2019). Lastly, the CRISPR/Cas9 system can be employed to study the regulation of gene expression on an epigenetic level. For example, fusion of the catalytic core of the human acetyltransferase p300 to dCas9 can target gene promoters or enhancers for acetylation of histone H3 Lys27, leading to gene activation (Pickar-Oliver and Gersbach, 2019).

One of the major advantages of the CRISPR/Cas genome editing tool is the versatility and programmability of the Cas endoribonuclease to be targeted to a large variety of DNA motifs, by simply co-expressing a target specific sgRNA. In contrast to RNAi technology, which induces incomplete and/or transient knockdown of gene expression, CRISPR/Cas9-mediated gene KO (knock-out)

can be permanent and complete, thus making it invaluable in genetic studies (Boettcher and Mcmanus, 2015). Moreover, genome-wide libraries for CRISPR KO, CRISPR activation and CRISPR interference have been developed, i.e. the Cancer Dependency Map (DEpMap), which have enabled the systematic interrogation of gene function (Sanson *et al.*, 2018; Behan *et al.*, 2019; Dempster *et al.*, 2019). However, since the application of the CRISPR/Cas9 technology in the study of gene function, several reports have raised the issue of off-target effects. Amongst the first reports was a study in 2013, demonstrating in a human cell-based EGFP disruption assay that the Cas9-sgRNA duplex can induce significant off-target mutagenesis in three different human cell lines (Fu *et al.*, 2013). Moreover, in two 2014 reports, dCas9 was found to interact with a large number of off-target sites in a ChIP-seq (chromatin immunoprecipitation sequencing) assay (Kuscu *et al.*, 2014; Wu *et al.*, 2014). Furthermore, it has been shown that repair of double-strands induced by the CRISPR/Cas genome-editing tool can lead to large genomic deletions and complex chromosomal rearrangements in mouse embryonic stem cells and mouse hematopoietic progenitors (Kosicki, Tomberg and Bradley, 2018). In addition, another study demonstrated by targeting 17 genomic sites in the mouse genome with the CRISPR/Cas9 tool that it can cause insertions and/or genomic deletions (9bp-600bp) with a higher prevalence at repeat sequences in the mouse genome (Shin *et al.*, 2017). Another disadvantage in the use of CRISPR/Cas genome editing tool are the long clonal selection steps required to establish daughter cell lines with homozygous loss of target genes, which can lead to loss of cell heterogeneity and/or cellular adaptation to the loss of the target gene (Olive *et al.*, 2018). As a result the behaviour of the selected subclonal population may

differ to that of the parental population for reasons unrelated to the target gene KO. For example, KO cell lines are used in cancer research to determine the relevance of a target gene in tumorigenicity, compared to the parental cell line. If however, the selected subclonal population exhibits an inherently different tumorigenic potential to the parental cell line or rewires the cell machinery to adapt to the loss of target gene, then erroneous conclusions would be made that the target gene is important in driving the tumour phenotype.

## **1.4. Targeted protein degradation**

The development of chemical compounds that modulate protein degradation via the UPS (Ubiquitin-Proteasome System) has brought great excitement to the field of drug discovery; providing a novel therapeutic approach to target difficult to drug or previously-thought undruggable proteins (Crews, 2010; Collins *et al.*, 2017; Fuchs, 2017; Neklesa, Winkler and Crews, 2017). The past two decades, chemical-biology tools have been developed based on these chemical compounds to acutely and selectively degrade via the UPS target proteins either directly, fused to degron motifs or modified protein domains (T. Wu *et al.*, 2020). Compared to genetic methods, these chemical-biology tools offer several advantages for the study of a target protein's functions. First, they allow for the controlled, rapid, and reversible degradation of a target protein. Indeed, the acute depletion of a target protein is one of the major advantages that chemical-biology tools offer, as this is important in preventing phenotypes arising from molecular compensation and cellular adaptation; as is observed with the CRISPR/Cas genome editing tool, due to clonal selection steps (Olive *et al.*, 2018). Equally important, is the reversibility of target protein depletion compared to CRISPR/Cas-mediated gene KO methods, which facilitates the study of the physiological effects following a rescue in a target protein's levels. Lastly, chemical-biology tools allow to overcome some of the challenges of applying genetic tools like RNAi in the in vivo setting, by allowing the cell specific degradation of a target protein in an inducible manner. The following sections first, review the components constituting the UPS and the regulatory



mechanisms that underpin it, and second, offer a critical perspective on current well-established chemical-biology tools and their applicability as research tools to interrogate complex biological problems.

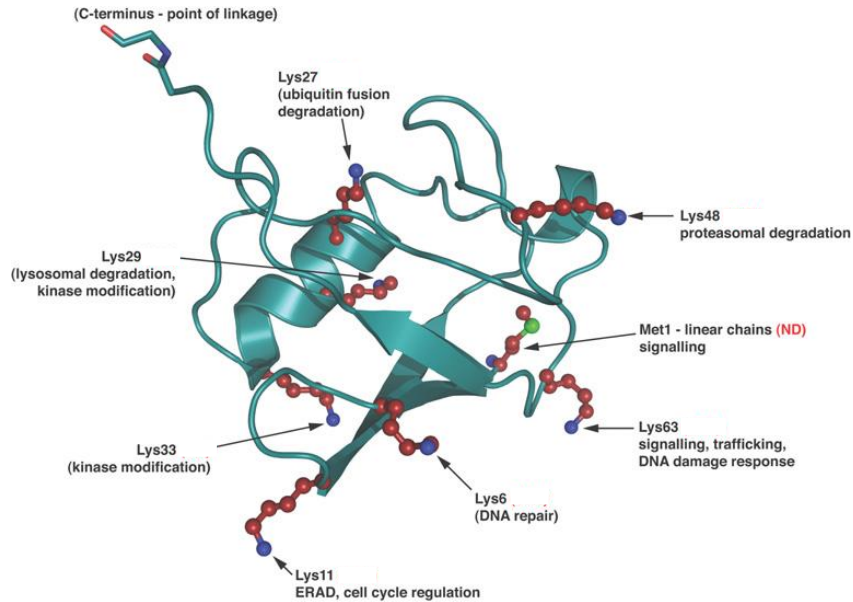
### **1.4.1 The UPS (Ubiquitin-Proteasome System)**

Constituting the primary proteolytic route for short-lived, misfolded, and damaged proteins, the UPS is one of the two major proteolytic systems in eukaryotic cells; the other being the autophagy-lysosome pathway (Dikic, 2017; Kwon and Ciechanover, 2017). The UPS was discovered in the 1980s by the seminal work of Aaron Ciechanover, Avram Hershko, Irwin Rose and colleagues when investigating the processes regulating degradation of denatured globin in reticulocyte lysates (Hershko *et al.*, 1980, 1983). Their work demonstrated that proteins were targeted for degradation via the covalent addition of multiple molecules of a small protein termed ATP (adenosine triphosphate)-dependent proteolytic factor-1 (APF-1), which was later identified as ubiquitin (Hershko *et al.*, 1980, 1983; Ciechanover, Finley and Varshavsky, 1984; Ganoth *et al.*, 1988). Since then, the importance of UPS-mediated degradation of proteins in regulating cellular processes, including cell cycle progression, cell survival, proliferation, apoptosis, and maintaining homeostasis has been well-documented (Dikic, 2017; Kwon and Ciechanover, 2017).

As stated, central to the functions of the UPS system, is ubiquitin, a small (8.6 kDa, 76 residues) and highly conserved protein amongst eukaryotic

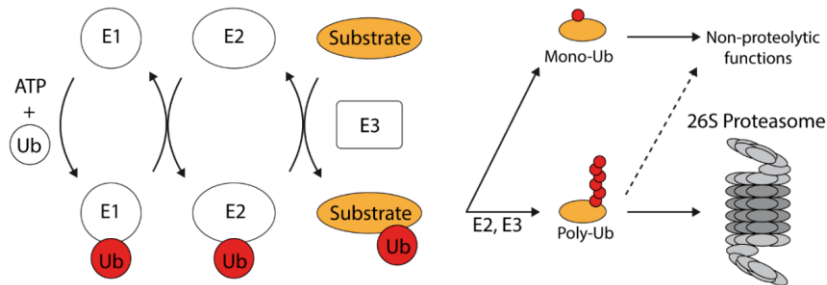
organisms. Ubiquitin assumes a highly stable  $\beta$ -grasp fold with a flexible six-residue C-terminal tail that ends in a conserved di-glycine motif (**Figure 1. 4**) (Komander and Rape, 2012). Its conjugation to a substrate protein occurs via an enzymatic post-translational modification process termed ubiquitination/ubiquitylation (Kwon and Ciechanover, 2017). Ubiquitination involves the cooperative action of three enzymes: the E1 ubiquitin-activating enzyme, the E2 ubiquitin-conjugating enzyme and the E3 ubiquitin ligase (Pickart and Eddins, 2004). In the human proteome there are two E1 ubiquitin-activating enzymes, ~40 E2 ubiquitin-conjugating enzymes and ~600 E3 ubiquitin ligases (Chen and Sun, 2009). In eukaryotes, the cascade of events initiates with an E1 ubiquitin-activating enzyme binding to ubiquitin and utilising ATP to create a thioester bond at the C-terminus of ubiquitin followed by release of an AMP (adenosine monophosphate) molecule (**Figure 1. 5**) (Pickart and Eddins, 2004). Subsequently, conformational changes in E1 and E2 enzymes, lead to the formation of an intermediary complex where activated ubiquitin is now transferred from the E1 ubiquitin-activating enzyme to the E2 ubiquitin-conjugating enzyme through a transthioesterification reaction (Pickart and Eddins, 2004). Subsequently, the ubiquitin-bound E2 ubiquitin-conjugating enzyme interacts with an E3 ubiquitin ligase enzyme promoting the transfer of the ubiquitin moiety to a lysine residue of a substrate protein (**Figure 1. 5**).

E3 ubiquitin ligases function by catalysing the formation of an isopeptide bond between the C-terminal glycine of ubiquitin and a lysine residue of the substrate (Komander and Rape, 2012). The attachment of a single ubiquitin moiety to a lysine residue of a substrate is referred to as monoubiquitination, whilst the addition of multiple single ubiquitin molecules to different lysine



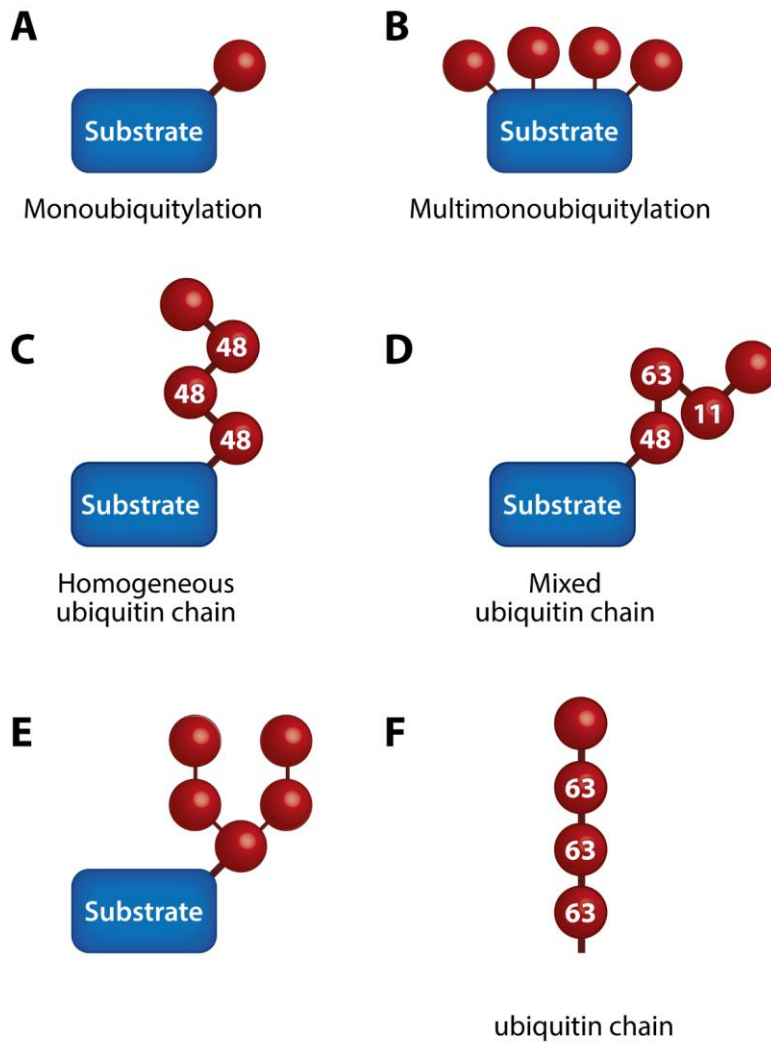
**Figure 1. 4 The structure of ubiquitin (green, PDB code 1UBQ) with its lysine residues.** These key seven lysine (red, with blue nitrogen atoms) residues and Met1 reside on different surfaces of ubiquitin. The speculative roles of each type of lysine or Met1 linkage are also indicated. The C-terminal diglycine motif is involved in isopeptide bond formation between different ubiquitin molecules (red, carbon atoms; blue, nitrogen atoms; green, sulphur atoms). (Taken from Komander, 2009)

residues is termed multi-monoubiquitination (**Figure 1. 6 A-B**) (Komander and Rape, 2012). Following the attachment of the first ubiquitin to a substrate, the conjugation process can be repeated to assemble a polyubiquitin chain. Ubiquitin chain formation or polyubiquitination involves the formation of isopeptide bonds between the C-terminal glycine (Gly76) of the incoming ubiquitin and either the N-terminus (Met1) or one of seven lysine residues (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 and Lys63) of a substrate-attached ubiquitin, which are oriented in



**Figure 1. 5 The Ubiquitin-Proteasome System.** Ubiquitin is activated in an ATP-dependent reaction by an E1 ubiquitin-activating enzyme and subsequently transferred to an E2 ubiquitin-conjugating enzyme via a transthioesterification reaction. The ubiquitin-charged E2 ubiquitin-conjugating enzyme then binds to an E3 ubiquitin ligase and the ubiquitin is transferred to a substrate. The substrate can be ubiquitinated with a single ubiquitin molecule (mono-Ub) or polyubiquitinated (poly-Ub) with several ubiquitin molecules forming long chains. The mode of ubiquitination determines the cellular fate of the substrate which is either altered in a non-proteolytic manner or send to the 26S proteasome to be degraded. (Taken from Sarikas, Hartmann and Pan, 2011)

space towards distinct directions (Komander and Rape, 2012). Polyubiquitin chains have been identified to be as short as two ubiquitins or encompassing more than ten ubiquitin molecules conjugated together (Komander and Rape, 2012). In addition, polyubiquitin chains are characterised as homogeneous when the same ubiquitin residue is modified during the elongation process, e.g. Lys48-linked or Lys63-linked chains (**Figure 1. 6 C**). The type of ubiquitin assembly on a given substrate conveys distinct structural and functional information, which is critical, as it determines the fate of that substrate in the cellular milieu. For instance, monoubiquitination has been shown to be implicated in DNA repair, regulation of histone function, gene expression, protein degradation and receptor



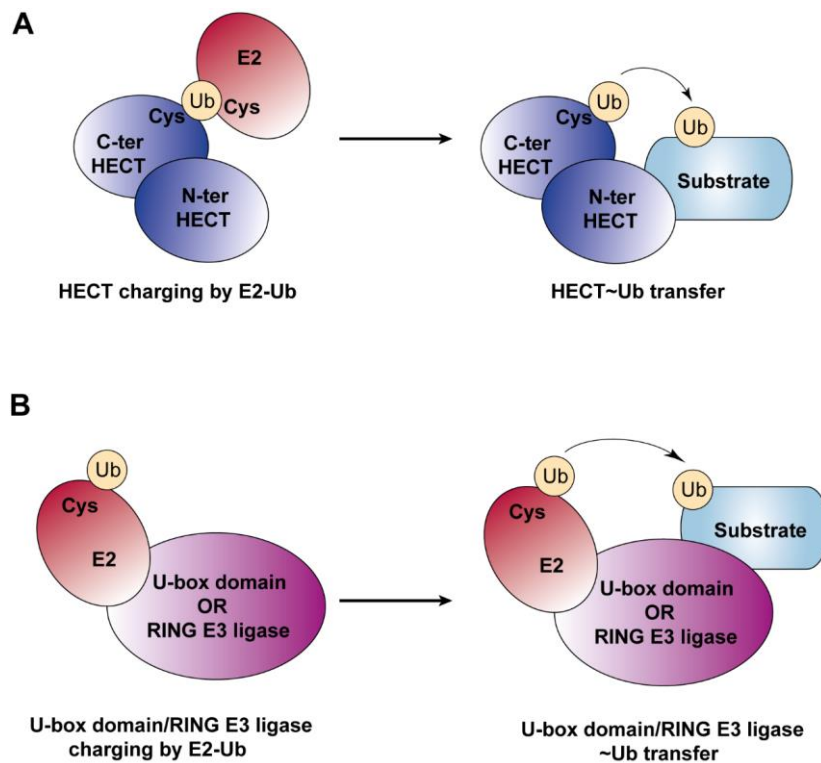
**Figure 1. 6 The different topologies of ubiquitination.** (A) Monoubiquitination. (B) Multimonoubiquitination. (C) Polyubiquitination via the formation of homogeneous ubiquitin chains. (D) Polyubiquitination via the formation of heterogeneous ubiquitin chains. (E) Polyubiquitination via the formation of branched ubiquitin chains. (F) Unanchored ubiquitin chain. (Taken from Komander and Rape, 2012)

endocytosis, whilst, multimonoubiquitination has been shown to regulate membrane receptor endocytosis as well as protein degradation (Hicke, 2001; Haglund *et al.*, 2003; Braten *et al.*, 2016). Lys63-linked chains are primarily implicated with protein scaffolding, cell signalling and autophagy pathways, while, Lys48-linked chains predominantly constitute a proteasomal degron which targets substrates for proteasome-mediated degradation (Jacobson *et al.*, 2009; Hadian *et al.*, 2011). For instance, TNF $\alpha$  (tumour necrosis factor  $\alpha$ ) induces the Lys63 polyubiquitination on Lys377 of RIP1 (receptor interacting protein kinase 1), which has been shown to function as a signalling element to recruit and activate the IKK (I $\kappa$ B kinase) in the NF- $\kappa$ B (nuclear factor  $\kappa$ B) pathway (Ea *et al.*, 2006). Opposing the activation of the NF- $\kappa$ B pathway, A20/TNFAIP3 (tumour necrosis factor alpha-induced protein 3) both deubiquitinates Lys63-linked chains on RIP and mediates the assembly of Lys48-linked polyubiquitin chains on RIP, which target RIP for proteasomal degradation (Wertz *et al.*, 2004). Furthermore, if different linkages alternate at succeeding positions of the chain, then they are described to exhibit mixed topology (**Figure 1. 6 D**) (Komander and Rape, 2012). Two other possible linkages detected in cells are the modification of a single ubiquitin with multiple molecules forming branched chains or unanchored ubiquitin chains (**Figure 1. 6 E-F**) (Komander and Rape, 2012). Lastly, it should be stated that, although ubiquitin is predominantly conjugated to lysine residues of substrate proteins, in rare occasions, isopeptide bonds can be formed between serine, threonine or cysteine residues of substrates and ubiquitin (Kwon and Ciechanover, 2017).

There are three main families of E3 ubiquitin ligases, the HECT (homologous to the E6AP carboxyl terminus) domain E3 ubiquitin ligases (~30

members), the U-box domain E3 ubiquitin ligases (~60 members), and the RING (really interesting new gene) finger domain E3 ubiquitin ligases (more than ~600 members) (Metzger, Hristova and Weissman, 2010; Kwon and Ciechanover, 2017; Ryu *et al.*, 2019; Sharma and Taganna, 2020). HECT domain E3 ubiquitin ligases have been shown to exert important functions in immune responses, protein trafficking and cell signalling pathways regulating cell growth, proliferation and migration (Metzger, Hristova and Weissman, 2010). For instance, SMURF1 (SMAD specific E3 ubiquitin protein ligase 1) has been shown to regulate cell migration by ubiquitinating TRAF4 (tumour necrosis factor receptor associated factor 4) and targeting it for proteasomal degradation, whilst, WWP2 (WW domain containing E3 ubiquitin protein ligase 2) ubiquitinates and targets for degradation TRIF/TICAM1 (TIR-domain containing adapter-inducing interferon- $\beta$ ) which leads to suppression of TLR3 (toll-like receptor 3)-mediated innate immune and inflammatory responses (Wang *et al.*, 2020). Moreover, NEDD4 (neuronal precursor cell-expressed developmentally downregulated 4) and HERC4 (HECT and RLD domain containing E3 ubiquitin protein ligase 4) upregulation has been shown to regulate cell proliferation, whilst SMURF1 has been implicated in the regulation of the TGF- $\beta$  and Wnt signalling pathways by interacting with Smad7 and axin, respectively (Wang *et al.*, 2020). Regarding the structural features of HECT domain E3 ubiquitin ligases, their N-terminal domains are diverse and mediate substrate targeting, whilst, a 350 residues-long HECT domain at their C-terminus contains a key cysteine residue similar to E2 ubiquitin-conjugating enzymes, that facilitates the formation of a thioester bond with a ubiquitin moiety (Metzger, Hristova and Weissman, 2010). Therefore, HECT domain E3 ubiquitin ligases serve as catalytic intermediates by recruiting

ubiquitin from the E2 ubiquitin-conjugating enzymes before transferring it to substrates (Figure 1.7 A).



**Figure 1.7 Mechanism of ubiquitin transfer to a substrate by the three different classes of E3 ubiquitin ligases. (A)** Mechanism of HECT E3 ligase-mediated ubiquitin transfer. A ubiquitin-charged E2 ubiquitin-conjugating enzyme transfers the ubiquitin moiety to a cysteine residue in the C-terminal domain of the HECT. Binding of a substrate to the HECT E3 ubiquitin ligase is followed by transfer of the ubiquitin to the bound substrate protein. **(B)** Mechanism of ubiquitin transfer by U-box domain and RING E3 ubiquitin ligases. A ubiquitin-charged E2 ubiquitin-conjugating enzyme binds the U-box domain or RING E3 ubiquitin ligase and directly transfers the ubiquitin moiety to a bound substrate.



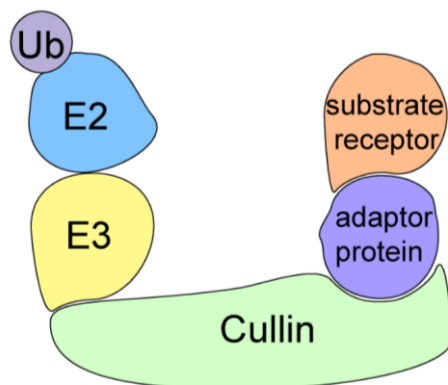
Unlike HECT domain E3 ubiquitin ligases, U-box domain, and RING finger domain E3 ubiquitin ligases do not form a catalytic intermediate with ubiquitin, but instead they serve as scaffolds to recruit ubiquitin-primed E2 conjugating enzymes to ubiquitinate the substrates (**Figure 1. 7 B**) (Cyr, Höhfeld and Patterson, 2002; Metzger, Hristova and Weissman, 2010). U-box domain E3 ubiquitin ligases are characterised by a 75 residues-long U-box domain, first identified in the yeast protein Ufd2, which is critical in the ubiquitination of substrates as it facilitates the recruitment of the ubiquitin-primed E2 conjugating enzyme (Hatakeyama *et al.*, 2001; Cyr, Höhfeld and Patterson, 2002). One of the best studied members of U-box E3 ubiquitin ligases is CHIP/STUB1 which has been identified to promote ubiquitination of CFTR (cystic fibrosis transmembrane conductance regulator), GR (glucocorticoid receptor) and HSP70 (Cyr, Höhfeld and Patterson, 2002). On the contrary, RING finger domain E3 ubiquitin ligases are specified by the RING domain which consists of a short cysteine and histidine rich motif coordinated by two zinc ions that stabilize its globular structure (Cyr, Höhfeld and Patterson, 2002; Pickart and Eddins, 2004). The RING domains of RING finger domain E3 ubiquitin ligases have been shown to promote ubiquitination events by binding E2 enzymes and bringing them into close proximity with a substrate (Pickart and Eddins, 2004).

RING finger domain E3 ubiquitin ligases have been shown to function as either monomers, dimers or within multi-subunit complexes (Cyr, Höhfeld and Patterson, 2002). The CBL E3 ubiquitin ligase family is a characteristic example of single subunit RING finger domain E3 ubiquitin ligase, with members containing both a RING domain and a substrate-binding domain. Multi-subunit RING finger domain E3 ubiquitin ligases function in complexes with scaffolding

proteins of the cullin family which engage both RING finger domain E3 ubiquitin ligases and ancillary proteins that function in recognising and recruiting substrates to the complex (Cyr, Höhfeld and Patterson, 2002; Petroski and Deshaies, 2005). Cullin RING E3 ubiquitin ligases constitute the largest superfamily of E3 ubiquitin ligases with more than 200 members (**Figure 1. 8**) (Sarikas, Hartmann and Pan, 2011). In humans, cullins (CUL) constitute a family of seven structurally related scaffold proteins (CUL1, CUL2, CUL3, CUL4A, CUL4B, CUL5 and CUL7) sharing an evolutionary conserved cullin homology domain (Petroski and Deshaies, 2005). Cullin RING E3 ligases are critical components of the cellular machinery regulating protein degradation and have been implicated in diverse cellular processes such as cell cycle control, DNA replication, apoptosis, angiogenesis and development (Metzger, Hristova and Weissman, 2010; Sarikas, Hartmann and Pan, 2011). For instance, CUL1 assembles RING E3 ubiquitin ligase complexes that regulate cell cycle progression, whilst CUL4 assemblies have been implicated in DNA replication and DNA repair mechanisms (Sarikas, Hartmann and Pan, 2011).

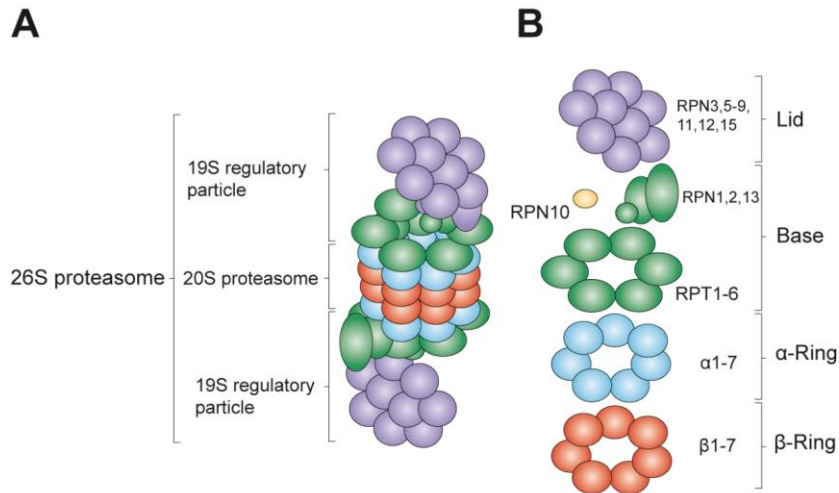
Opposing the role of E3 ubiquitin ligases, DUBs (deubiquitinating enzymes) are proteases that maintain the dynamic state of the cellular ubiquitome by cleaving isopeptide bonds between ubiquitin moieties on substrates (Amerik and Hochstrasser, 2004). To date, seven families of DUBs have been identified, namely, the cysteine protease subfamilies of USPs (Ubiquitin-Specific Proteases), UCHs (Ubiquitin C-terminal Hydrolases), OTU (Ovarian Tumour Proteases), MJDs (Machado-Josephin Domain proteases), MINDYs (motif interacting with Ubiquitin-containing novel DUB family), ZUP1 (Zinc finger-containing ubiquitin peptidase 1) and the zinc-dependent

metalloprotease family of JAMM (Jab1/MPN/MOV34) protease (Clague, Urbe and Komander, 2019). DUBs have several key roles, in between maintaining protein homeostasis by negatively regulating protein degradation and therefore sustaining normal rates of proteolysis (Hanpude *et al.*, 2015).



**Figure 1. 8 Schematic representation of an assembled cullin RING E3 ubiquitin ligase.** The cullin acts a scaffold protein recruiting an E3 ubiquitin ligase enzyme and an adaptor protein which mediates recruitment of a substrate receptor. A ubiquitin-charged E2 ubiquitin-conjugating enzyme binds to the E3 ubiquitin ligase enzyme and mediates transfer of the ubiquitin moiety to a substrate protein bound to the substrate receptor. (Adapted from Nabet *et al.*, 2018)

When not cleaved by the action of deubiquitinating enzymes, certain ubiquitin modifications, for example, Lys11 and Lys48-linked chains, send proteins to the 26S proteasome to be degraded (Bard *et al.*, 2018). The 26S proteasome is an ATP-dependent multiprotein complex consisting of a 20S catalytic subunit and two 19S regulatory cap subunits and acts by mediating the proteolytic degradation of proteins (**Figure 1. 9 A**) (Chen and Madura, 2002; Schweitzer *et al.*, 2016). The 20S catalytic subunit consists of four stacked heptameric rings that form a central pore in which a substrate protein enters to be degraded (**Figure 1. 9 B**) (Schweitzer *et al.*, 2016). The two inner rings consist of seven  $\beta$  subunits that exhibit three catalytic activities at the interior surface of the ring, chymotrypsin-like, trypsin-like, and peptidylglutamyl-peptide hydrolysing (Bard *et al.*, 2018). The outer two rings consist of seven  $\alpha$  subunits which act as molecular gates regulating the entry of proteins within the central pore and additionally recruiting the 19S regulatory cap subunits (Bard *et al.*, 2018). The conserved hexameric ATPase Valosin-containing protein (VCP)/p97 as well as other ubiquitin receptor proteins, such as Rad23, bind to polyubiquitinated proteins and mediate their transfer to the 26S proteasome to be degraded (Chen and Madura, 2002). Subsequently, polyubiquitin chains bind to the 19S subunit and are cleaved off from substrate proteins which are subsequently unfolded and degraded, generating peptides of 3-25 amino acids in length (Bard *et al.*, 2018).



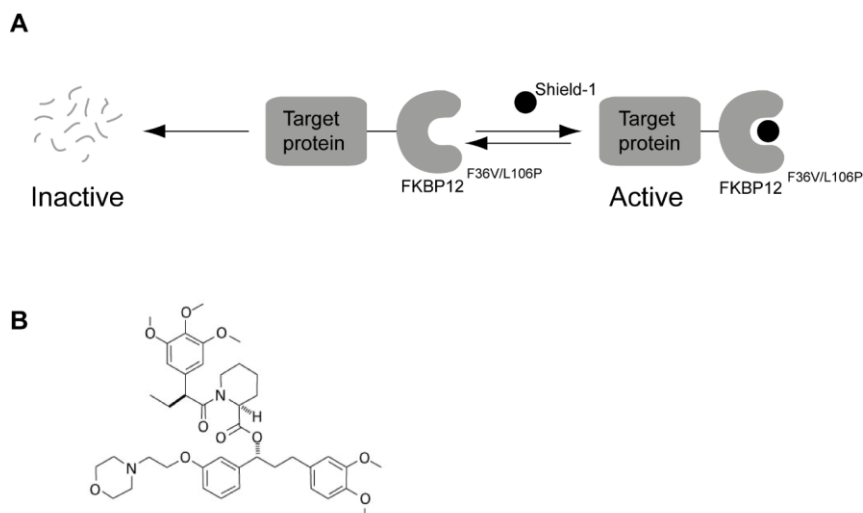
**Figure 1. 9 Schematic diagram of the 26S proteasome.** (A) The 26S proteasome consists of the 19S regulatory particles (RP, also known as PA700) and the catalytic 20S proteasome (two outer  $\alpha$ -rings and two inner  $\beta$ -rings stacked into a barrel formation) (B) Subunit composition of the 26S proteasome. The regulatory particle is further divided into the base and the lid subcomplexes, which are composed of regulatory particle triple-A (RPT) and regulatory particle non-ATPase (RPN) subunits. Yellow coloured RPN10 is located at the base-lid interface. (Taken from Murata, Yashiroda and Tanaka, 2009)

## 1.4.2 Chemical-biology tools exploiting the UPS

### 1.4.2.1 Destabilising Domain system

Thomas Wandless' laboratory developed the DD (Destabilising Domain) system in 2006 (**Figure 1. 10 A**). The DD system is based on an engineered mutant of the human cytosolic prolyl isomerase FKBP12 (FK506- and rapamycin binding protein), namely, FKBP12<sup>F36V/L106P</sup> (Banaszynski *et al.*, 2006). The F36V mutation of FKBP12 was previously identified to form a cavity which allowed binding of synthetic FKBP ligands (Clackson *et al.*, 1998). One such ligand is Shield-1, a chemical analogue of the FKBP ligand SLF\*, in which the carboxylic acid is substituted with a morpholino group, to improve its pharmacokinetic properties and potency (**Figure 1. 10 B**) (Banaszynski *et al.*, 2006). The L106P mutation was identified from a library of mutant FKBP12<sup>F36V</sup> proteins generated by error-prone PCR to induce strong destabilisation when fused to YFP (yellow fluorescent protein) (Clackson *et al.*, 1998; Banaszynski *et al.*, 2006). The DD system incorporates genetic fusion in viral vectors of the FKBP12<sup>F36V/L106P</sup> to a target protein, which when expressed in the absence of Shield-1, the fusion protein rapidly (~4h) and constitutively degrades via the proteasome (Banaszynski *et al.*, 2006). Following addition of Shield-1 (~1 µM) in the system, it binds to the FKBP12<sup>F36V/L106P</sup> stabilising and shielding from degradation the fusion protein (Banaszynski *et al.*, 2006). However, it should be noted that while the DD system achieves rapid degradation of target proteins, it requires

continuous treatment with Shield-1 to maintain protein levels (Banaszynski *et al.*, 2006). In addition to inducing YFP destabilisation, the DD system was also shown to efficiently degrade in NIH3T3 cells a number of exogenously expressed target proteins, such as, GSK-3 $\beta$  (Glycogen synthase kinase 3 $\beta$ ), p21, CDK1 (cyclin-dependent kinase 1), Rac1 (Ras-related C3 botulinum toxin substrate 1), Cdc42 (cell division control protein 42 homolog) and RhoA (Ras homolog family member A) (Banaszynski *et al.*, 2006).



**Figure 1. 10 The Destabilising Domain system.** (A) A target protein genetically fused to FKBP<sup>F36V/L106P</sup> is degraded upon translation. When the chemical compound Shield-1 is added to the system it binds FKBP<sup>F36V/L106P</sup> and protects the target protein from degradation. (B) Chemical structure of Shield-1. (Taken from Banaszynski *et al.*, 2006)

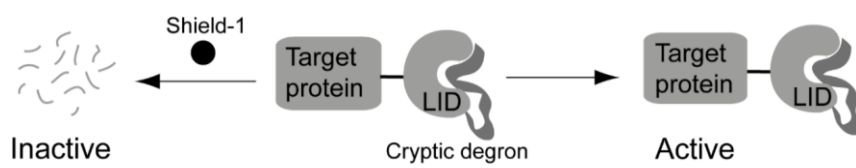
Since its development, the DD system has been successfully applied in several research models. A study in 2007, successfully adapted the DD system to be used in the protozoan malaria parasite (*Plasmodium falciparum*) (Armstrong and Goldberg, 2007). The group demonstrated that FKBP12<sup>F36V/L106P</sup>-tagged YFP and falcipain-2 (a food vacuole cysteine protease) expressed from inducible vectors transfected in the parasites were degraded in the absence of Shield-1 (Armstrong and Goldberg, 2007). Another group in 2009, employed the DD system in two species of *Leishmania* parasites (*Leishmania major* and *Leishmania braziliensis*) for the degradation of cytosolic UDP-galactopyranose mutase, a key enzyme in lipophosphoglycan biogenesis (Madeira da Silva *et al.*, 2009). Lastly, the DD system has been also paired with the Cre-LoxP system to develop a functional system that offers loss of an endogenous target protein with simultaneous exogenous expression of that target protein tagged with FKBP12<sup>F36V/L106P</sup> (An *et al.*, 2015). As a proof of concept, in this study the tumour suppressor protein PTEN (Phosphatase and tensin homolog) was selected and genetically fused to a further mutated FKBP12<sup>F36V/L106P</sup>, namely FKBP12<sup>F36V/L106P/E31S/D32S</sup> so as to remove charged amino acids E31 and D32 that abrogate the function of PTEN. Nucleofection of Cre-LoxP vectors harbouring FKBP12<sup>F36V/L106P/E31S/D32S</sup>-tagged PTEN in PTEN<sup>fllox/fllox</sup> mouse neurons lead to substantial decrease in endogenous PTEN levels and created Shield-1-dependent regulation of exogenous FKBP12<sup>F36V/L106P/E31S/D32S</sup>-tagged PTEN (An *et al.*, 2015).



### 1.4.2.2 Ligand-induced degradation system

In 2011, Thomas Wandless' laboratory developed the LID (ligand-induced degradation) system (**Figure 1. 11**) (Bonger *et al.*, 2011). The LID system involves genetic fusion of a target protein to a LID domain, which is based on the mutant FBP12<sup>F36V</sup> and a 19 amino acid -long degron (TRGVEEVAEGVLLRRRGN) attached to its C-terminus. In the absence of Shield-1, this 19 amino acids-long degron binds to the FKBP12<sup>F36V</sup> active site, which prevents it from inducing degradation of the target protein. Upon addition of the small molecule Shield-1 in the system, this 19 amino acids-long degron is displaced from FKBP12<sup>F36V</sup> and destabilises the target protein, marking it for proteasome-mediated degradation (Bonger *et al.*, 2011). In this study, the LID system was successfully employed for the selective degradation of six transcription factors involved in reprogramming differentiated cells into a pluripotent state, namely Oct-4, Sox2, Myc, Klf4, Lin28 and Nanog (Bonger *et al.*, 2011). NIH3T3 cells were individually transduced with these targets fused to the LID domain and treatment with 2  $\mu$ M of Shield-1 for 24h was demonstrated to degrade all target proteins. Compared to the DD system, the LID system offers the advantage of not requiring continuous Shield-1 treatment to maintain target protein levels. In theory, this could circumvent potential problems relating to unspecific target protein degradation from fluctuating Shield-1 levels and any possible off-target effects stemming from the continuous exposure of the cells to Shield-1. In addition, the LID system has only been shown to degrade C-terminally tagged proteins, with N-terminus fusions resulting in Shield-1-

independent degradation (Bonger *et al.*, 2011). Therefore, this limitation reduces the applicability of the LID system to target proteins that functionally do not tolerate fusions of the LID domain to their C-terminus.



**Figure 1. 11 The ligand-induced degradation (LID) system.** A target protein is expressed genetically fused to the LID domain and is targeted for degradation following addition of chemical compound Shield-1 to the system. (Taken from Bonger *et al.*, 2011)

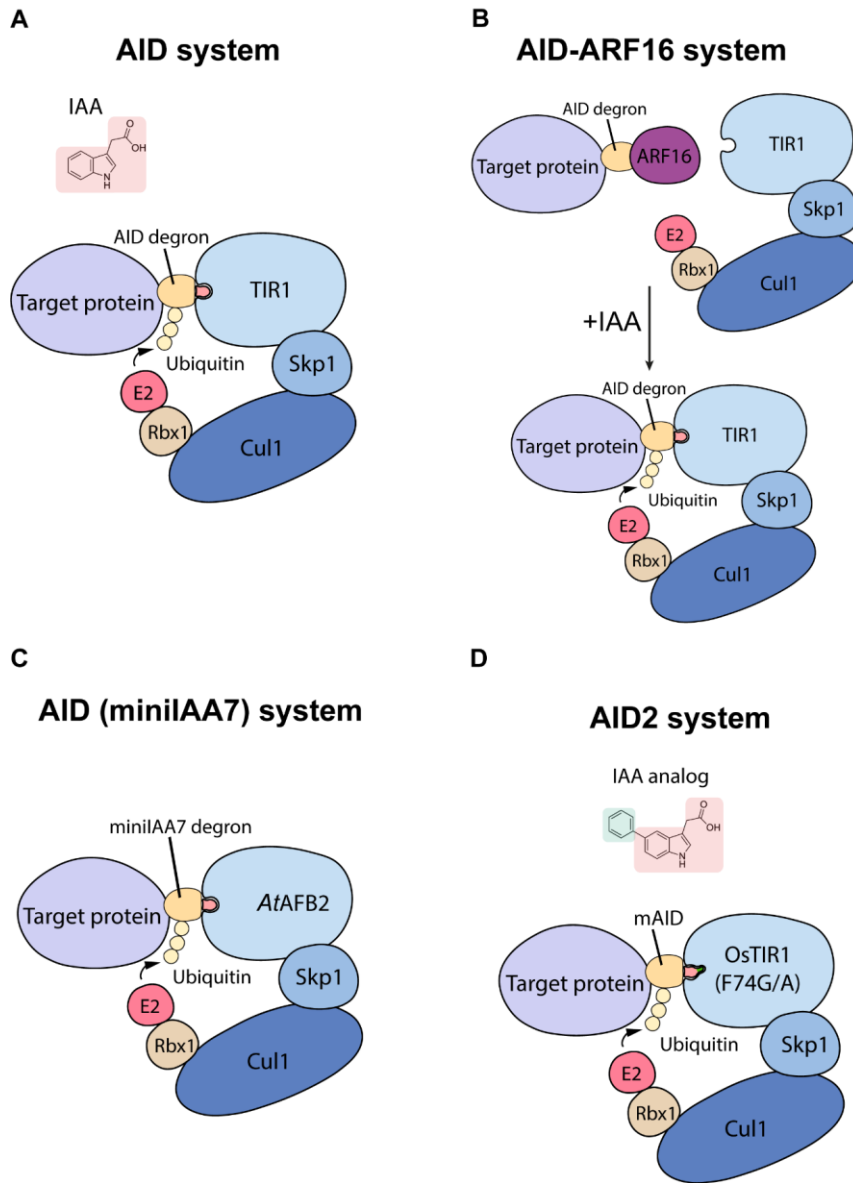
### 1.4.2.3 Auxin inducible degron system

The auxin family is a group of phytohormones implicated in several aspects of plant growth and development by controlling cell division, proliferation, and differentiation (Teale, Paponov and Palme, 2006). These phytohormones can be naturally produced by plants, such as IAA (indole acetic acid), the predominant auxin in plants; or synthetically produced, such as NAA (1-naphthaleneacetic acid) (Teale, Paponov and Palme, 2006). Auxins engage TIR1 (F-box transport inhibitor response 1) protein, a substrate receptor of the SCF (SKp1, Cullin 1 and F-box) E3 ubiquitin ligase complex and allow the recruitment of Aux/IAA transcription repressors (Teale, Paponov and Palme, 2006). Aux/IAA genes along with ARF (auxin response factor) genes regulate expression of auxin-responsive genes. ARFs are a family of transcription factors that bind to AREs (auxin responsive elements) in the promoters of auxin-responsive genes, stimulating their expression. (Teale, Paponov and Palme, 2006). On the contrary, Aux/IAs inhibit this binding of ARFs to AREs, thereby suppressing ARF-mediated gene transcription. Recruitment of Aux/IAs by auxins to the SCF<sup>TIR1</sup> E3 ubiquitin ligase leads to their ubiquitination and subsequent proteasomal degradation, thus effectively re-stimulating expression of auxin-responsive genes via ARFs re-binding to AREs (Teale, Paponov and Palme, 2006).

Based on this auxin-dependent degradation pathway from plants, Masato Kanemaki's laboratory developed in 2009 a chemical-biology tool which they named the AID (Auxin Inducible Degron) system (**Figure 1. 12 A**) (Nishimura *et*

*al.*, 2009). The AID system relies on exogenous expression of TIR1 and an auxin-inducible degron derived from the IAA17 protein of the thale cress (*Arabidopsis thaliana*) fused to a target protein (Nishimura *et al.*, 2009). Auxin addition to the system recruits the target protein fused to AID to the SCF<sup>TIR1</sup> E3 ubiquitin ligase marking it for proteasomal degradation. In this seminal study, NIH3T3 (mouse-derived), CHO-K1 (hamster-derived), and COS1 (monkey-derived) and HEK 293T (human-derived) cell lines were generated to exogenously express the TIR1 and GFP (green fluorescent protein) fused to the AID degron. Treatment with 500  $\mu$ M of either IAA or NAA for 5h was shown to induce rapid and reversible depletion of AID degron-tagged GFP (Nishimura *et al.*, 2009). Furthermore, this study showed that the AID system can be applied in budding yeast (*Saccharomyces cerevisiae*), where endogenous tagging of DNA replication and cell cycle regulators with the AID degron yielded effective protein degradation within minutes following auxin addition (Nishimura *et al.*, 2009).

Over the years, the AID system has been successfully adapted as a chemical-biology tool for the following yeast and animal models: fission yeast (*Schizosaccharomyces pombe*), nematode (*Caenorhabditis elegans*), fruit fly (*Drosophila Melanogaster*) and zebrafish (*Danio Rerio*). In fission yeast, the AID system was modified to develop i-AID (improved auxin-inducible degron) system (Kanke *et al.*, 2011). In the i-AID system, TIR1 is genetically fused to fission yeast Skp1 adaptor protein of the CUL1 E3 ubiquitin ligase complex and this fusion protein allows for enhanced degradation of AID-tagged target proteins following auxin treatment of yeast cells, in between, Mcm4, Orc2, Cdc45 and pol1 (Kanke *et al.*, 2011). Moreover, the i-AID system was combined with transcription



**Figure 1. 12 The AID (Auxin Inducible Degron) system.** (A) The first AID system to be developed was based on exogenous expression of the SCF (Skp1, Cul1 and F-box) E3 ubiquitin ligase substrate receptor TIR1 and the target protein fused to the AID degron. Addition of an

auxin molecule (e.g. IAA) results in recruitment of the target protein fused to the AID degron to the E3 ubiquitin ligase which becomes ubiquitinated and marked for proteasomal degradation. **(B)** The AID-ARF16 system is an adaptation of the first AID system to be developed. It is based on the additional exogenous expression of the PB1 (Phox and Bem1) domain of *Oryza Sativa* ARF16 which binds the AID degron to inhibit its association with TIR1, thus preserving target protein expression levels from auxin-independent degradation. Auxin facilitates the interaction of TIR1 with AID and promotes dissociation of ARF16 and the subsequent ubiquitination and proteasome-mediated degradation of the target protein. **(C)** The AID (miniIAA7) system constitutes another adaptation of the first AID system to be developed. It is based on a novel AID degron (miniIAA7) and the *Arabidopsis thaliana* auxin receptor F-box protein AFB2 instead of TIR1. **(D)** The AID2 system is the second-generation AID system. It is based on the mutant *Oryza Sativa* TIR1<sup>F74G</sup> and an improved mAID degron. An analogue of IAA, 5-phenyl-indole-3-acetic acid (5-Ph-IAA), is used in this system which induces degradation of the target protein even at sub-micromolar levels. (Adapted from Yesbolatova *et al.*, 2020)

repression of target proteins by substituting their promoters with the thiamine-repressible *nmt81* promoter; which upon thiamine administration shuts-off transcription from the genes under its control (Kanke *et al.*, 2011). This system was named *off*-AID and was successfully used in fission yeast for enhanced degradation of the DNA replication proteins Pol1 and Cdc45, compared to the *i*-AID system. However, the *off*-AID system faces a limitation to its widespread applicability; as it requires a double modification of target genes, one to fuse the AID tag sequence and another to replace the promoter with *nmt81* promoter (Kanke *et al.*, 2011). Likewise, in the nematode *Caenorhabditis elegans*, the AID system was adapted to express a mutated *Arabidopsis thaliana* TIR1 protein with enhanced substrate affinity, fused to a red fluorescent protein to track expression. In this adapted version, a 44-amino acid sequence from *Arabidopsis thaliana* IAA17 protein fused to a synthetic GFP gene for visualisation was selected as a minimal AID degron (Zhang *et al.*, 2015). This degron was fused to either the

nuclear protein SMU-2 (suppressor of *mec-8* and *unc-52*) or the cytoplasmic protein DHC1 (dynein heavy chain 1). In the presence of auxin it induced a rapid, conditional and reversible degradation of these targets at all developmental stages of *Caenorhabditis elegans* (Zhang *et al.*, 2015). Moreover, the AID system was initially reported in a proof of concept study to effectively degrade a Cdk inhibitor in the fruit fly *Drosophila Melanogaster* (Trost, Blattner and Lehner, 2016). In a later study, the AID system was also demonstrated to effectively degrade in the nervous system of the fruit fly PERIOD, a protein involved in the regulation of circadian rhythms (Chen, Werdann and Zhang, 2018). Lastly, a recent study demonstrated that fusion of the AID degron to an anti-GFP nanobody stimulates following auxin addition the degradation of GFP-tagged proteins in HeLa cells (Daniel *et al.*, 2018). This hybrid AID-nanobody system was successfully applied in zebrafish embryos, to degrade GFP-tagged proteins expressed in various cellular compartments (Daniel *et al.*, 2018).

Since the initial development of the AID system, Masato Kanemaki's laboratory also established a CRISPR/Cas9-mediated KI (knock-in) method to insert the AID degron sequence in the genomic locus of a target protein, thus allowing the endogenous expression of AID degron-tagged target proteins (Natsume *et al.*, 2016). This method first involves the CRISPR/Cas9-mediated KI of an expression vector encoding TIR1 at the safe harbour AAVS1 (adeno-associated virus integration site 1) locus to generate parental cells. Subsequently, an in-frame AID degron cassette is introduced in these parental cells after the last codon of the gene of interest (Natsume *et al.*, 2016). A donor vector is used to introduce in cells the AID degron which is flanked by short homology arms (~125-220 bp) complementary to the genomic locus of the target

protein (Natsume *et al.*, 2016). Moreover, this method involves the co-transfection of two AID-degron/HA donor vectors, one containing a neomycin and the other a hygromycin resistance marker, thus allowing for selection of parental cells with bi-allelic insertion of the AID degron at the genomic locus of the target protein (Natsume *et al.*, 2016). This method has been successfully applied in HCT116 for inserting the AID degron at the genetic locus of both RAD21 and DHC1 proteins and was also demonstrated to target both proteins for proteasomal degradation following treatment with 500  $\mu$ M IAA for 24h (Natsume *et al.*, 2016). Recently, the same group has also updated their CRISPR/Cas9 method to allow for endogenous tagging of target proteins at their N-terminus (Yesbolatova *et al.*, 2019).

This pairing of the AID system with CRISPR/Cas9-mediated KI of the AID degron is an important advantage of this chemical-biology tool, as it allows to study the functional consequences of acute depletion of a target protein expressed from its genomic locus. Despite this, a drawback in the use of the AID system is the requirement for exogenous expression of the substrate receptor TIR1, which both complicates experimental procedures and could possibly exert unknown biological functions when expressed in non-plant cells. Another main limitation is that leaky degradation of AID-tagged target proteins has been observed in the absence of auxins, thus for some target proteins essential for cell viability requiring conditional expression of TIR1 under a tetracycline-inducible promoter (Natsume *et al.*, 2016). Furthermore, the high auxin concentrations (typically 100-500  $\mu$ M) required to achieve desirable target protein degradation have been shown in some cell lines to negatively affect cell growth, which could be a problem for the application of the AID system in stem cells (Yesbolatova *et*



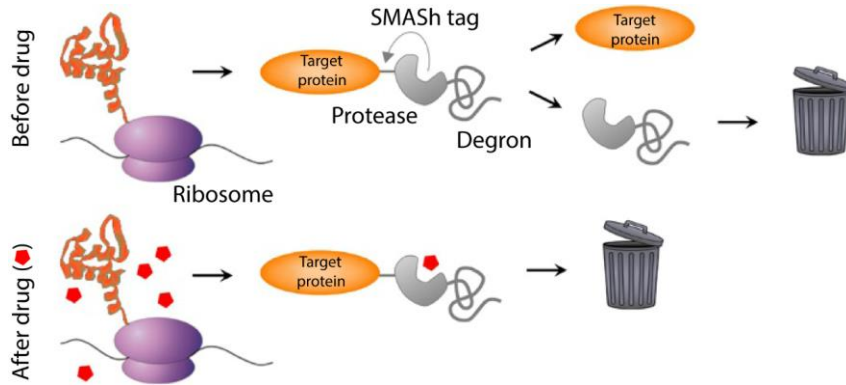
*al.*, 2020). In addition, IAA is known to cause kidney toxicity when it is converted to indoxyl sulphate in the liver. Therefore, when coupled with the already high auxin concentrations used in vitro this possibly limits the AID system from successfully being applied in mice (Yesbolatova *et al.*, 2020).

Over the past few years, several attempts have been made by various research groups to improve the AID system. In one optimised version of the AID system, co-expression of *Oryza sativa* TIR1 with the PB1 (Phox and Bem1) domain of the *Oryza sativa* ARF16 transcription factor was shown to decrease AID-tagged protein degradation in the absence of auxins, whilst, increasing the rate of auxin-induced degradation (**Figure 1. 12 B**) (Sathyan *et al.*, 2019). Another group screened various auxin receptor F-box proteins and degrons in A431 cells and identified the *Arabidopsis thaliana* auxin receptor F-box protein AFB2 and a degron based on the *Arabidopsis thaliana* IAA7 amino acid sequence 37-104 (miniIAA7) as an optimal combination; exhibiting both minimal basal degradation and improving upon the auxin-inducible degradation of target proteins compared to earlier AID system versions (**Figure 1. 12 C**) (Li *et al.*, 2019). However, both adaptations of the original AID system still faced limitations in terms of the high auxin concentrations used to achieve protein degradation. Recently, Masato Kanemaki's laboratory established a second generation of this technology called the AID2 (AID version 2) system (**Figure 1. 12 D**) (Yesbolatova *et al.*, 2020). The AID2 system utilises a point mutant (F74G) of *Oryza sativa* TIR1, which was found to create a hole within the auxin-binding site (Yesbolatova *et al.*, 2020). Several new IAA analogues were tested for efficient binding to this new cavity, from which 5-phenyl-indole-3-acetic acid (5-Ph-IAA) was shown to induce target protein degradation even at sub-micromolar concentrations and

found to cause significantly less off-target gene expression alterations in cells, compared to IAA (Yesbolatova *et al.*, 2020). Furthermore, the use of *Oryza sativa* TIR1<sup>F74G</sup> exhibited no detectable basal degradation of tagged proteins thus overcoming another limitation of this technology (Yesbolatova *et al.*, 2020). In addition, the AID2 system was successfully applied in mammalian cell lines, the budding yeast and both in adult and embryo mice (Yesbolatova *et al.*, 2020). This novel and improved AID2 system is a promising tool for functional studies of target proteins with its application to other model organisms to be further explored.

#### 1.4.2.4 SMASh system

Michael Lin's laboratory developed, in 2015, the SMASh (Small Molecule-Assisted Shutoff) system, which is effectively a chemical-biology tool that utilises a drug-controllable self-removing degron, called SMASh tag (**Figure 1. 13**) (Chung *et al.*, 2015). This SMASh tag is comprised from the HCV (Hepatitis C virus) NS3pro (non-structural protein 3 protease) domain; fused at its amino terminus to an NS3 cleavage site and at its carboxyl terminus to the NS4A (non-structural protein 4A). When the SMASh tag is genetically fused to a target protein, it post-translationally removes itself in the absence of an antiviral drug, by the action of the NS3pro on the N-terminal NS3 cleavage site, thus leaving the target protein untagged (Chung *et al.*, 2015). Exposure of this system to an antiviral drug, such as the selective NS3 protease inhibitor, asunaprevir, blocks degron removal, thus, inducing degradation of the target protein (Chung *et al.*, 2015). Notably, the exact mechanism by which SMASh tag targets a protein for degradation remains elusive, with both the UPS and the autophagy-lysosome pathway involved in the process (Chung *et al.*, 2015). The SMASh system has been successfully employed for the selective degradation of: YFP, mouse PSD95 (postsynaptic density protein 95), mouse CaMKII $\alpha$  (calcium- calmodulin-activated protein kinase II $\alpha$ ), drosophila GluRIIA (glutamate receptor IIA) and CYP21A2 (cytochrome P450 family 21 subfamily A member 2) (Chung *et al.*, 2015). Treatment of HEK 293T or Vero cells expressing these target proteins



**Figure 1. 13 The SMASH tag system.** A SMASH tag degron fused to a target protein is cleaved by its internal protease activity and is degraded due to internal degron activity. Addition of a viral protease inhibitor (e.g. acyclovir) induces degradation of the target protein. (Taken from Chung *et al.*, 2015)

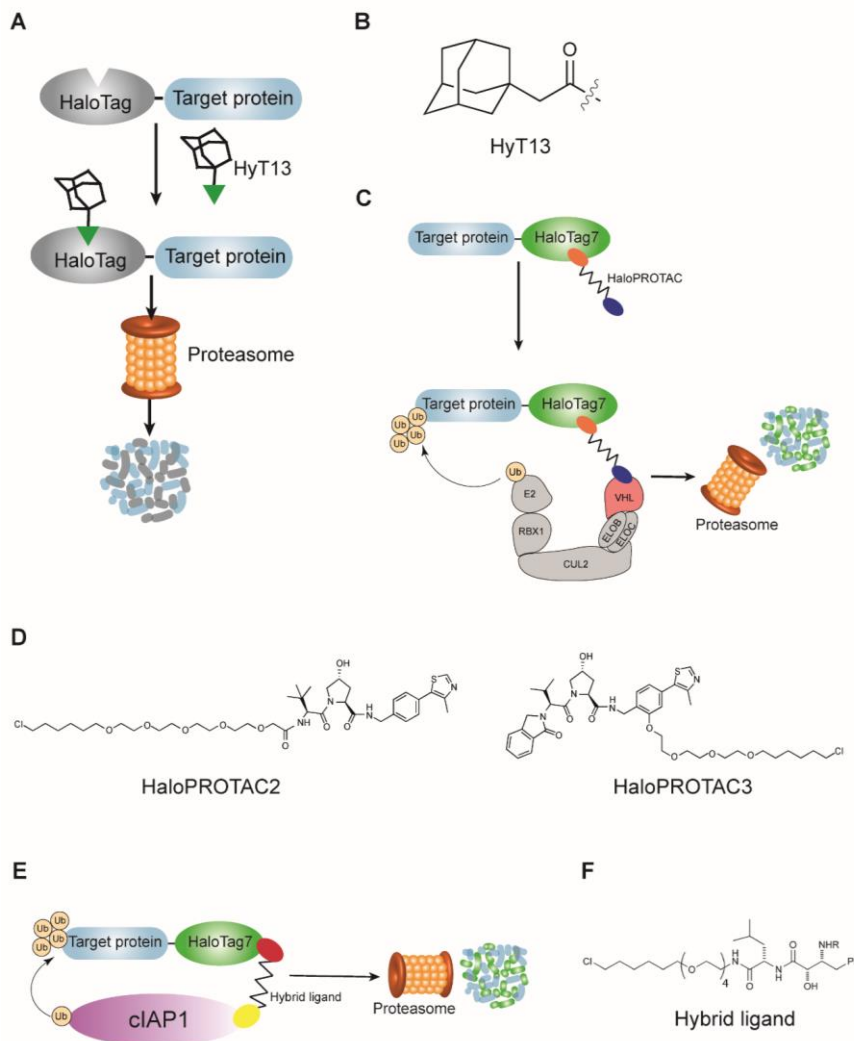
fused to SMASH tag with low  $\mu\text{M}$  concentrations (0.3-3  $\mu\text{M}$ ) of an antiviral drug (asunaprevir or ciluprevir) for 24h induced protein degradation. Moreover, degradation of SMASH tag-YFP was achieved in primary cultures of rat cortico-hippocampal and mouse cortical neurons following an incubation period of seven days with 3  $\mu\text{M}$  asunaprevir (Chung *et al.*, 2015). Furthermore, SMASH tag-tagged yeast endoribonuclease Ysh1 and Sec14 phosphatidylinositol-phosphatidylcholine transfer protein were also shown to be degraded when expressed in yeast cells (*Saccharomyces cerevisiae*) following treatment for 48h with 3  $\mu\text{M}$  or 10  $\mu\text{M}$  asunaprevir, respectively (Chung *et al.*, 2015). Recently, the SMASH tag has been fused to Cas9 protein in an attempt to develop a CRISPR/Cas9 genome editing system of repressible Cas9 expression; with

Cas9-SMASH tag degradation in HEK 293T cells achieved following treatment with 1  $\mu$ M asunaprevir for 24h (Y. Wu *et al.*, 2020). However, despite the fact that the SMASH system has been shown to degrade a number of proteins with different subcellular localisations, a setback in its application is that only newly synthesised target proteins are removed; as SMASH tag is excised from existing target proteins in the absence of the antiviral drug. Therefore, degradation kinetics follow the rate of target protein synthesis and these for proteins with long half-lives tend to be slow. As a consequence, this system might not be suitable for studying rapid dynamics in some biological systems. Lastly, even though the SMASH tag system has been validated in both yeast and mammalian cells it has not yet been tested in animal models.

#### 1.4.2.5 HaloTag-based systems

The HaloTag is a ~34 kDa peptide based on a modified bacterial enzyme, haloalkane dehalogenase, for which several highly specific and interchangeable chloroalkane ligands have been developed (Los *et al.*, 2008; England, Luo and Cai, 2015). Fusion of the HaloTag to a target protein is widely used for protein isolation and purification, capture of interacting partners of a target protein, and in vitro imaging of protein subcellular localisation (Los *et al.*, 2008; England, Luo and Cai, 2015). In 2011, the HaloTag was employed for the development of a chemical-biology tool, the hydrophobic tagging system (**Figure 1. 14 A**) (Neklesa *et al.*, 2011). This system combines genetic fusion of the HaloTag peptide to a target protein with the use of the hydrophobic compound, HyT13 (hydrophobic tag 13), to selectively degrade the target protein fused to HaloTag (**Figure 1. 14 B**). In this seminal study, treatment with 1  $\mu\text{M}$  HyT13 for 24h was demonstrated to degrade the following HaloTag-tagged cytosolic and transmembrane proteins stably expressed in HEK 293T cells: EGFP (enhanced green fluorescent protein), Ror2 receptor tyrosine kinase-like orphan receptor, CDE3 cell surface glycoprotein, CD9 transmembrane protein, G protein-coupled receptor GPR40 and Frizzled-4 transmembrane receptor (Neklesa *et al.*, 2011). Moreover, the hydrophobic tagging system was successfully applied in zebrafish (*Danio Rerio*) embryos and in a mouse tumour model. Treatment with 10  $\mu\text{M}$  HyT13 of zebrafish embryos expressing HaloTag-Smad5 degraded this target protein within 24h and in mice transplanted ectopically with NIH3T3 mouse fibroblast cells expressing HaloTag-Hras<sup>G12V</sup>, treatment with HyT13 (25 mg/kg body

weight) achieved HaloTag-Hras1<sup>G12V</sup> degradation within 24h (Neklesa *et al.*, 2011).



**Figure 1. 14 HaloTag-based systems.** (A) Schematic representation of the hydrophobic tagging system which involves genetic fusion of HaloTag to a target protein and the use of the hydrophobic compound HyT13 (hydrophobic tag 13) to mark the target protein for proteasome-mediated degradation. (B) Chemical structure of HyT13. (C) Schematic representation of a HaloPROTAC HaloTag-based system. Following addition of a HaloPROTAC to the system, a HaloTag-tagged target protein is recruited to CUL2<sup>VHL</sup> E3 ubiquitin ligase where it becomes ubiquitinated and sent to the 26S proteasome to be degraded. (D) Chemical structures of HaloPROTAC2 and HaloPROTAC3. (E) Schematic representation of HaloTag-based system utilising cIAP1 E3 ubiquitin ligase to selectively degrade target proteins. Following addition of a hybrid ligand, a target protein genetically fused to HaloTag7 is recruited to cIAP1, ubiquitinated and subsequently degraded by the 26S proteasome. (F) Chemical structure of the hybrid ligand. (Adapted from Neklesa *et al.*, 2011; Buckley *et al.*, 2015; Nabet *et al.*, 2020)

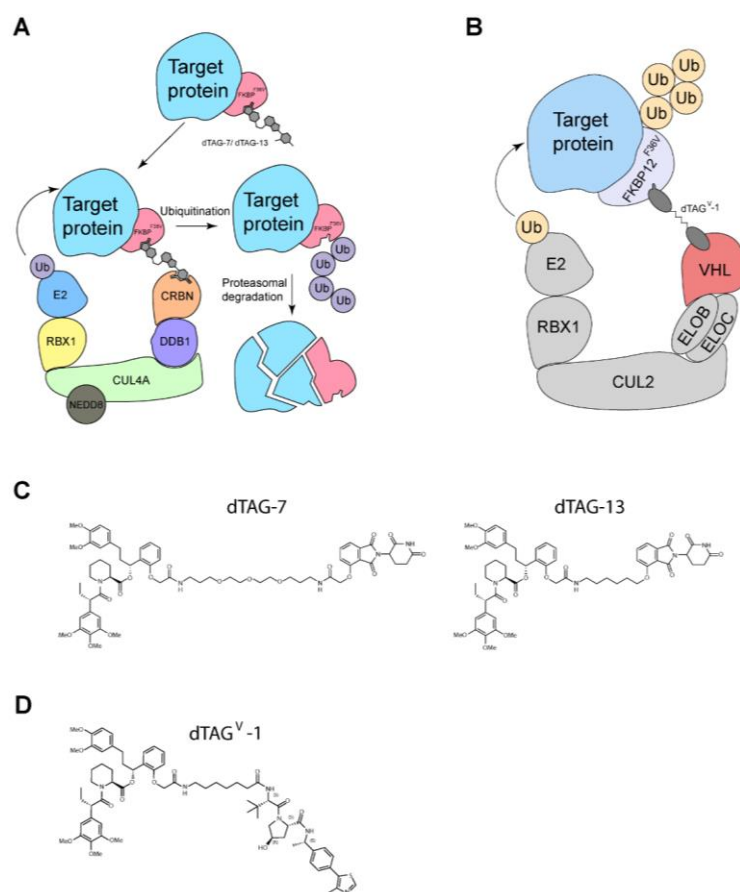
In 2015, Craig Crew's laboratory also developed PROTACs (PROteolysis TArgeting Chimeras), referred to as HaloPROTACs for HaloTag-based systems, against HaloTag (**Figure 1. 14 C-D**) (Buckley *et al.*, 2015). PROTACs/HaloPROTACs are heterobifunctional molecules that consist of a ligand which binds to a target protein, joined by a linker to an E3 ubiquitin ligase recruiter compound (Neklesa, Winkler and Crews, 2017). Recruitment of the target protein to the E3 ubiquitin ligase results in protein ubiquitination and subsequent degradation (Toure and Crews, 2016; Neklesa, Winkler and Crews, 2017). HaloPROTACs were originally designed to engage both HaloTag-tagged proteins and the VHL (von-Hippel-Lindau) protein, the substrate receptor of CUL2<sup>VHL</sup> E3 ubiquitin ligase (Buckley *et al.*, 2015). The group demonstrated that treatment of HEK 293T cells stably expressing HaloTag fused to either GFP, ERK1 (extracellular signal-regulated kinase 1) or MEK1 (mitogen-activated protein kinase kinase) with a HaloPROTAC (500 nM) leads to degradation of these target proteins within 24h. Furthermore, in the same year, Minoru



Ishikawa's laboratory had also successfully developed a hybrid ligand to selectively degrade HaloTag-tagged target proteins by targeting them to the ubiquitin ligase cIAP1 (cellular inhibitor of apoptosis protein 1) (**Figure 1. 14 E-F**) (Tomoshige *et al.*, 2015). The group demonstrated that treatment with 10  $\mu$ M of this hybrid ligand to HEK 293T cells stably expressing HaloTag-CREB1 (cAMP responsive element binding protein 1) and MCF-7 breast cancer cells transiently expressing HaloTag-c-jun degraded the target proteins within 24h, albeit protein levels were not completely reduced (Tomoshige *et al.*, 2015). Lastly, the HaloTag technology has also been successfully combined with the CRISPR/Cas9 genome editing tool to endogenously tag targets with the HaloTag sequence and degrade them, thus, allowing the study of a target protein's function within a physiologically relevant context (Tovell *et al.*, 2019; Caine *et al.*, 2020). Fusion of HaloTag to the genomic locus of SGK3 (serum and glucocorticoid kinase-3) and VPS34 (Class III PI 3-kinase) in HEK 293T cells was demonstrated to successfully degrade these target proteins following HaloPROTAC treatment for 48h (Tovell *et al.*, 2019).

### 1.4.2.6 dTAG system

The dTAG system is another PROTAC-based chemical-biology tool developed by Nathanael Gray's and James Bradner's laboratories in 2018 (**Figure 1. 15 A-B**) (Nabet *et al.*, 2018). This tool involves the fusion of FKBP12<sup>F36V</sup> (~12kDa) to a target protein, and the use of PROTACs dTAG-7, dTAG-13, or dTAG<sup>V</sup>-1 to selectively degrade the target protein (**Figure 1. 15 C-D**) (Nabet *et al.*, 2018, 2020). FKBP12<sup>F36V</sup> was selected as the tag based on previous observations that the F36V mutation confers the formation of a cavity that allows binding of synthetic FKBP ligands (Clackson *et al.*, 1998). Therefore, based on these previously developed compounds, the FKBP12<sup>F36V</sup>-specific ligand groups of dTAG-7, dTAG-13 and dTAG<sup>V</sup>-1 were developed. dTAG-7 and dTAG13 function by engaging both FKBP12<sup>F36V</sup> and CRBN (cereblon), the substrate receptor of CRL4<sup>CRBN</sup> E3 ubiquitin ligase (**Figure 1. 15 A**), whilst dTAG<sup>V</sup>-1 constitutes a second generation dTAG molecule that bridges a FKBP12<sup>F36V</sup>-tagged target protein with CUL2<sup>VHL</sup> E3 ubiquitin ligase (**Figure 1. 15 B**) (Nabet *et al.*, 2018, 2020). The group demonstrated that treatment of MV4-11 leukaemia cells expressing N-terminally or C-terminally FKBP12<sup>F36V</sup>-tagged cMYC, KRAS[G12V], EZH2 (enhancer of zeste 2 polycomb repressive complex 2 subunit), HDAC1 (histone deacetylase 1) or PLK1 (polo-like kinase 1) with 500 nM dTAG-13, induces degradation of these target proteins within 24h (Nabet *et al.*, 2018). The dTAG system was also shown to function in the *in vivo* setting,



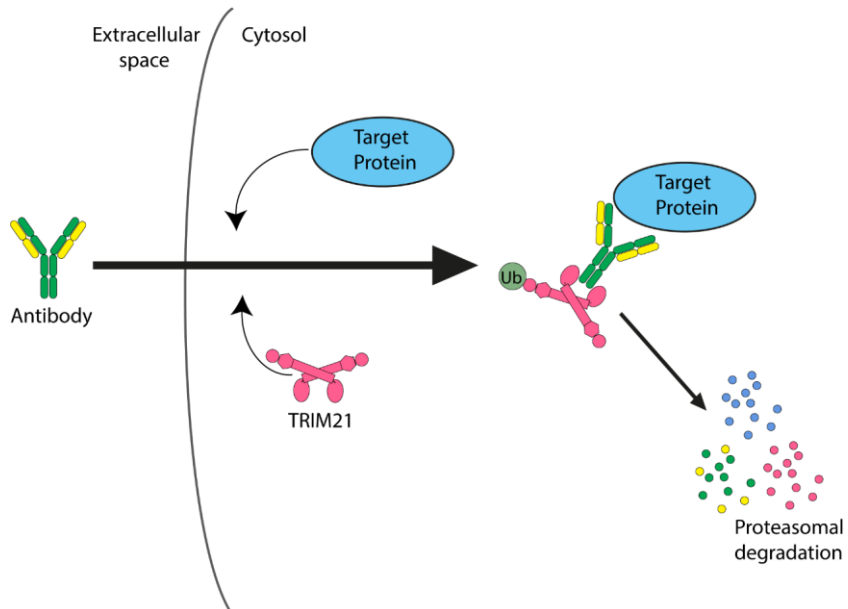
**Figure 1.15 The dTAG system.** (A) Schematic representation of the dTAG system employing the CUL4<sup>CRBN</sup> E3 ubiquitin ligase. Addition of dTAG-7 or dTAG-13 recruit FKBP12<sup>F36V</sup>-tagged target proteins to CUL4<sup>CRBN</sup> E3 ubiquitin ligase, where they are ubiquitinated and degraded by the 26S proteasome. (B) Schematic representation of the dTAG system employing the CUL2<sup>VHL</sup> E3 ubiquitin ligase. Addition of dTAG<sup>V-1</sup> recruits FKBP12<sup>F36V</sup>-tagged target proteins to CUL2<sup>VHL</sup> E3 ubiquitin ligase, thus inducing target protein ubiquitination and degradation. (C) Chemical structures of dTAG-7 and dTAG-13 heterobifunctional degraders. (D) Chemical structure of dTAG<sup>V-1</sup> heterobifunctional degrader. (Adapted from Nabet *et al.*, 2018, 2020)

where following bone marrow engraftments of MV4-11 leukaemia cells expressing luciferase-FKBP12<sup>F36V</sup> in mice, dTAG-13 administration (25 mg/kg) led to effective loss of bioluminescence within 4h (Nabet *et al.*, 2018). dTAG<sup>V</sup>-1 was also tested in mice injected with MV4-11 leukaemia cells expressing luciferase-FKBP12<sup>F36V</sup> and was also found to induce loss of bioluminescence after 4h of administration (35 mg/kg) (Nabet *et al.*, 2020). However, dTAG<sup>V</sup>-1 exhibited an improved duration of degradation when compared to dTAG-13 treated mice (35 mg/kg), with degradation evident 28h after the final administration (Nabet *et al.*, 2020). Interestingly, dTAG<sup>V</sup>-1 as a second generation dTAG molecule has allowed to overcome some of the limitations of the dTAG system, by enabling the degradation of certain target proteins that are recalcitrant to CRBN-recruiting dTAG molecules. In detail, 1  $\mu$ M of dTAG<sup>V</sup>-1 in contrast to dTAG-13 has been demonstrated in EWS502 Ewing sarcoma cells to effectively degrade within 24h FKBP12<sup>F36V</sup>-tagged EWS/FLI; a fusion transcription factor oncoprotein that arises following chromosomal translocation in Ewing sarcoma (Nabet *et al.*, 2020). Similarly, in HCT-116 colorectal carcinoma cells endogenously expressing MED14 (mediator of RNA polymerase II transcription subunit 14) C-terminally tagged with FKBP12<sup>F36V</sup>, treatment with 500 nM of dTAG<sup>V</sup>-1 but not dTAG-7 led to degradation of the target protein within 2h (Jaeger *et al.*, 2020). Furthermore, a CRISPR/Cas9 genome editing system has been successfully applied for locus specific KI of the FKBP12<sup>F36V</sup> degron to a gene of interest (Nabet *et al.*, 2018). This allows for expression of a target protein fused to the FKBP12<sup>F36V</sup> from its genomic locus and the study of the fusion protein's function within a more physiologically relevant context. Coupled with this CRISPR/Cas9-mediated KI method, the dTAG system has already been

successfully applied in studying the functional consequences of target protein loss. In one such study, acute degradation of FKBP12<sup>F36V</sup>-tagged serine/threonine kinase MELK (maternal embryonic leucine zipper kinase) in MDA-MB-468 cells with CRBN-recruiting dTAG molecules was demonstrated to not significantly affect cell growth in these basal-like breast cancer cells, corroborating results obtained from loss of MELK by genetic methods (Huang *et al.*, 2017). Lastly, ENL/MLLT1 protein, a histone acetylation reader and a critical component of SEC (super elongation complex), was demonstrated in MV4-11 leukaemia cells treated with 500 nM of dTAG-7 or dTAG-13 to be degraded within 30min and 1h, respectively (Erb *et al.*, 2017). This acute loss of ENL/MLLT1 protein was shown to suppress genome-wide gene expression of several well-characterized leukemic drivers, with *MYC*, *MYB*, *MEIS1* and *HOXA10* being amongst the earliest and most severely downregulated transcripts (Erb *et al.*, 2017).

### 1.4.2.7 Trim-Away

Trim-Away is a method for selective degradation of target proteins based on the E3 ubiquitin ligase TRIM21. TRIM21 is implicated in the intracellular immune response and functions by binding to antibody-bound pathogens and proteopathic agents, auto-ubiquitinating and subsequently triggering its proteasome-mediated degradation along with its bound substrates (**Figure 1. 16**) (Clift *et al.*, 2018). The Trim-Away method harnesses this characteristic function of TRIM21 for targeted protein degradation by delivering antibodies against target proteins via microinjection of individual cells or electroporation of large cell populations (Clift *et al.*, 2017). Although TRIM21 is widely expressed in diverse cell types and tissues, endogenous levels might not be sufficient for complete target protein degradation in some instances, therefore, requiring exogenous overexpression of TRIM21 to form a functional Trim-Away system. TRIM21 protein can be delivered in cells within a plasmid by transfection protocols or introduced via electroporation together with the antibody (Clift *et al.*, 2017). Despite this added complexity, the Trim-Away system does not require genetic modification of the target. This method has been demonstrated to effectively and rapidly remove target proteins, with degradation occurring within minutes or hours when tested in various cell lines (Clift *et al.*, 2017). In addition, the Trim-Away system has been successfully applied in both zebrafish (*Danio rerio*) and frog (*Xenopus laevis*) embryos for the investigation of target proteins that are critical during the early stages of embryogenesis (Chen *et al.*, 2019; Weir *et al.*, 2021).



**Figure 1. 16 The TRIM-Away system.** Addition of an antibody against a target protein leads to its recruitment to TRIM21 E3 ubiquitin ligase and the subsequent degradation of the whole complex. (Taken from Clift *et al.*, 2017)

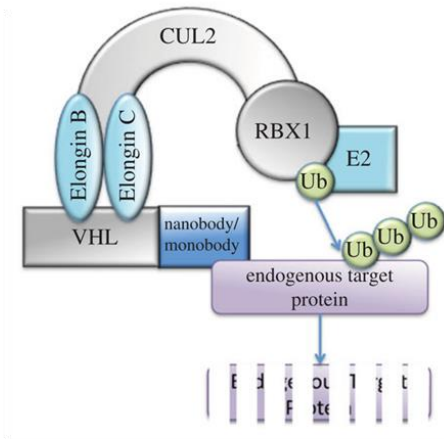
#### 1.4.2.8 AdPROM System

**Commented [SN1]:** New chemical-biology tool section describing the AdPROM system for targeted protein degradation

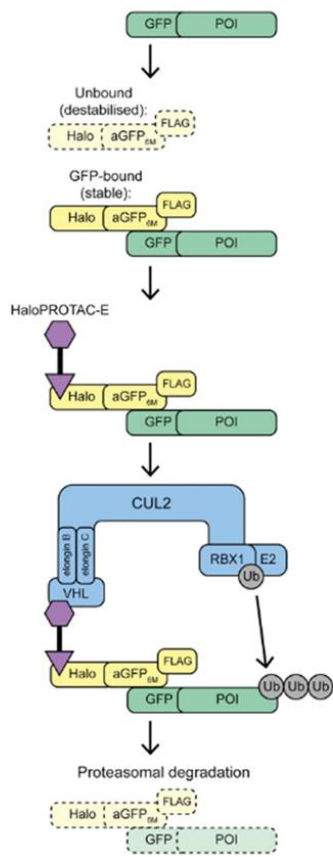
The AdPROM (Affinity-directed Protein Missile) system is a chemical-biology tool developed by Gopal Sapkota's laboratory in 2016 (Fulcher *et al.*, 2016). This tool involves the genetic fusion of VHL protein, the substrate receptor of CUL2<sup>VHL</sup> E3 ubiquitin ligase, to either a camelid-derived VHH domain (nanobody) or a synthetic polypeptide monobody, that selectively bind and recruit target proteins to the CUL2<sup>VHL</sup> E3 ubiquitin ligase complex for ubiquitination and subsequent proteasomal degradation (**Figure 1. 17**) (Fulcher *et al.*, 2016, 2017). The group demonstrated that CRISPR/Cas9-mediated homozygous KI (knock-in) of GFP at the endogenous loci of both VPS34 (vacuolar protein sorting 34) and PAWS1 (protein associated with SMAD1)/ FAM83G in human kidney HEK293 and U2OS osteosarcoma cell lines, respectively, resulted in near-complete proteasomal degradation of these target proteins within 6h following expression of VHL-anti-GFP antibody fusion protein (Fulcher *et al.*, 2016). Moreover, an AdPROM system consisting of VHL conjugated to synthetic monobodies or a nanobody selectively binding the protein tyrosine phosphatase SHP2 and ASC (Apoptosis-associated speck-like protein containing CARD) protein, respectively, were demonstrated to induce a potent degradation of these target proteins in multiple human cancer cell lines (Fulcher *et al.*, 2017). Furthermore, the AdPROM system was recently adapted for the targeted degradation of endogenous GFP-tagged KRAS and untagged, endogenous KRAS and HRAS in A459 human adenocarcinoma cells, HT-29, and SW620 human colorectal adenocarcinoma cells (Röth *et al.*, 2020). Lastly, Gopal



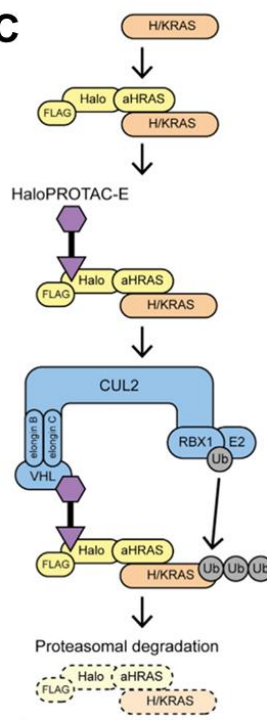
**A**



**B**



**C**



**Figure 1. 17 The AdPROM system. (A)** Schematic diagram depicting AdPROM-mediated target protein degradation using nanobodies or monobodies conjugated to the VHL substrate receptor protein of the CUL2<sup>VHL</sup> E3 ubiquitin ligase, to recruit endogenous GFP-tagged or untagged target proteins. Binding of target proteins to the CUL2<sup>VHL</sup> E3 ubiquitin ligase results in their ubiquitination and subsequent proteasomal degradation. **(B)** Schematic representation of FLAG-aGFP<sub>6M</sub>-Halo HaloPROTAC L-AdPROM system. **(C)** Schematic representation of FLAG-Halo-aHRAS HaloPROTAC L-AdPROM system. (Taken from Fulcher *et al.*, 2017 and Simpson *et al.*, 2020)

Sapkota's and Alessio Ciulli's laboratories, recently established a second generation of this technology termed the ligand-inducible AdPROM (L-AdPROM) system (**Figure 1. 17 B-C**) (Simpson *et al.*, 2020). This modified system, utilises a polypeptide binder consisting of either a destabilizing GFP nanobody (GFP<sub>6M</sub>) or a RAS monobody fused with the HaloTag peptide and a FLAG (5'-DYKDDDDK-3') epitope sequence. This polypeptide binder interacts with a target protein, and subsequently, a HaloPROTAC (i.e. HaloPROTAC-E) binds HaloTag to recruit the target protein to the CUL2<sup>VHL</sup> E3 ubiquitin ligase complex, where it is targeted for proteasome-mediated degradation. The L-AdPROM system has been successfully employed for efficiently and reversibly degrading endogenously GFP-tagged ULK1 (Unc-51 like autophagy activating kinase), FAM83D (Family With Sequence Similarity 83 Member D), and SGK3 (Serum and glucocorticoid kinase-3) within 24h, in ARPE-19 retinal pigment epithelia cells, U2OS human bone osteosarcoma epithelial cells, and HEK293T human embryonic kidney cells, respectively (**Figure 1. 17 B**) (Simpson *et al.*, 2020). Moreover, the L-AdPROM system has been successfully applied for the endogenous degradation of untagged KRAS and HRAS, by using an anti-H-RAS monobody (aHRAS), with dual specificity for both KRAS and HRAS but not NRAS

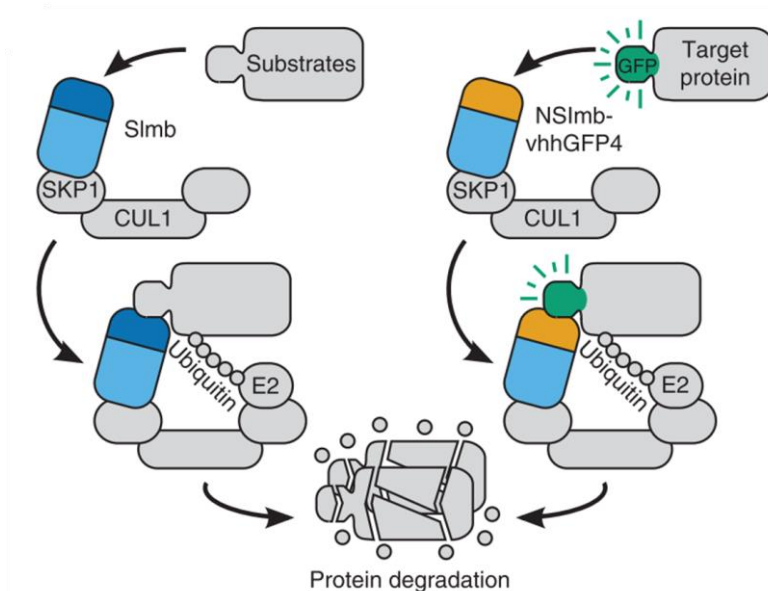
(**Figure 1. 17 C**) (Simpson *et al.*, 2020). This set of data demonstrated that through substitution of the polypeptide binder, the L-AdPROM system could be applied for the inducible degradation of potentially any intracellular target protein, albeit the large molecular weight of the polypeptide binder could negatively interfere with the biological functions of that target protein. Despite this, the development of the L-AdPROM system provides more options for targeted protein degradation strategies and overcomes some of the limitations of previous AdPROM systems, in-between slow kinetics of degradation and leakiness. However it should also be noted that this chemical biology tool has not yet been validated in animal models, thus, currently limiting its use in the in vitro setting.

#### 1.4.2.9 deGradFP system

**Commented [SN2]:** New chemical-biology tool section describing the deGradFP system for targeted protein degradation

In 2011, Markus Affolter's laboratory developed the deGradFP (degrade Green Fluorescent Protein) system, a method for inactivating GFP-tagged proteins based on an engineered F-box protein, named NSlmb-vhhGFP4 (**Figure 1. 18**). NSlmb-vhhGFP4 consists of the F-box domain contained in the N-terminal part of fruit fly (*Drosophila melanogaster*) F-box protein Slmb (supernumerary limbs), fused to a single-domain antibody fragment (VhhGFP4) (Caussinus, Kanca and Affolter, 2012). Under normal physiological conditions, Slmb acts as a substrate receptor for the SCF (SKp1, Cullin 1 and F-box) E3 ubiquitin ligase complex, recruiting substrates to the E3 ligase, which are ubiquitinated and marked for proteasome-mediated degradation (**Figure 1. 18**). NSlmb-vhhGFP4 both binds GFP-tagged target proteins via vhhGFP4, and also acts as a substrate receptor for the SCF E3 ubiquitin ligase complex, therefore recruiting GFP-tagged target proteins to the E3 ligase complex where they become ubiquitinated and subsequently degraded via the proteasome (**Figure 1. 18**). The deGradFP has been successfully applied for the degradation of nuclear, cytoplasmic and transmembrane proteins with experimental evidence showing a loss of EGFP signal at ~30 min after induction, and only a residual 10% or less can be observed after ~3h in most of the cases (Caussinus, Kanca and Affolter, 2012, 2013; Caussinus and Affolter 2016). However it should be noted that, since GFP constitutes a relatively large tag, its fusion to target proteins could affect their functional folding and subsequently their biological functions. Despite this, since several transgenic stocks carrying GFP-tagged proteins for animal model

organisms exist, this constitutes deGradFP as one of the most versatile systems regarding its applicability by the scientific community. Moreover, the deGradFP system has been also successfully adapted by another group to be based on the GFP nanobody VhhGFP4 attached to SPOP E3 ubiquitin ligase (Shin *et al.*, 2015). This was demonstrated to guide the degradation of nuclear proteins, such as GFP-tagged histone H2B more efficiently compared to the first version of this system.



**Figure 1. 18 The deGradFP system.** Slmb F-box protein acts as a substrate receptor for the SCF (SKP1, Cullin 1 and F-box) E3 ubiquitin ligase complex, recruiting substrates to the E3 ligase, which are ubiquitinated and marked for proteasome-mediated degradation. The deGradFP system is based on NSlmb (an F-box domain in the N-terminus of Slmb) fused to a single-domain antibody fragment (VhhGFP4) which specifically binds GFP. When a target protein is fused to GFP, expression of NSlmb-vhhGFP4 results in recruitment of that target protein to the SCF E3 ubiquitin ligase complex, followed by ubiquitination and proteasome-mediated degradation. (Taken from Caussinus, Kanca and Affolter, 2012)

## **1.5 Rationale for developing a novel protein degradation tool**

Genetic methods, such as RNAi technology and the CRISPR/Cas genome editing tool are currently widely adopted by the scientific community for studying the function of target proteins, however, they face several limitations in their application. For instance, regarding the application of the RNAi technology in the study of target function, incomplete target knockdown in certain cases could convolute observed phenotypes and lead to erroneous conclusions regarding the target protein's functions (Sigoillot and King, 2011; Boettcher and Mcmanus, 2015). Moreover, RNAi technologies entail disadvantages in their *in vivo* use, whilst clonal selection steps in the application of CRISPR/Cas methods can lead to cellular adaptation to the loss of the target gene (Kim *et al.*, 2009; Sigoillot and King, 2011; Boettcher and Mcmanus, 2015; Song *et al.*, 2015; Olive *et al.*, 2018). The advent of chemical-biology tools in the late 2000s has addressed some of the limitations of genetic methods; for instance, by allowing for an acute, modulated, and reversible degradation of target proteins, phenotypes arising from molecular compensation and cellular adaptation can be prevented.

Despite progress made in the last few years, current well-established chemical-biology tools exhibit several drawbacks in their use, too. Some of the key issues pertaining their application include toxicity effects of the compounds used to induce loss of target proteins, the absence of appropriate controls, the use of suboptimal degradation motifs with low versatility, multicomponent complexity, and a difficulty to access some of the key reagents. For example, auxin-based systems (AID, AID-ARF16, miniIAA7 and AID2) lack appropriate

controls for any toxicity and off-target effects mediated by the use of auxins (NAA/ IAA/ 5-Ph-IAA) (Nishimura *et al.*, 2009; Li *et al.*, 2019; Sathyan *et al.*, 2019; Yesbolatova *et al.*, 2020). Indeed, IAA has been previously reported to cause kidney toxicity when it is converted to indoxyl sulphate in the liver, therefore potentially limiting the use of these systems for in vivo studies (Yesbolatova *et al.*, 2020). Moreover, the auxin-based systems and the Trim-Away system exhibit multicomponent complexity since the first requires the exogenous overexpression of TIR1 to form a functional E3 ubiquitin ligase for target protein degradation and the latter the exogenous overexpression of TRIM21 to form a functional Trim-Away system (Nishimura *et al.*, 2009; Clift *et al.*, 2017, 2018). Other systems such as the LID, exhibit low versatility as they only mediate degradation of a target when fused at its C-terminus, therefore, excluding the study of target proteins that cannot functionally tolerate genetic fusions at their C-termini (Bonger *et al.*, 2011). Herein, I aimed to develop a novel chemical-biology tool, based on our increasing understanding of the UPS machinery and recent advances in the field of drug discovery, that would allow to overcome some of the current limitations of both genetic and chemical-biology methods. Such a system would be employed as a simple, highly modular, robust, and complementary method to existing tools for studying the functions of target proteins, both in vitro and in vivo.

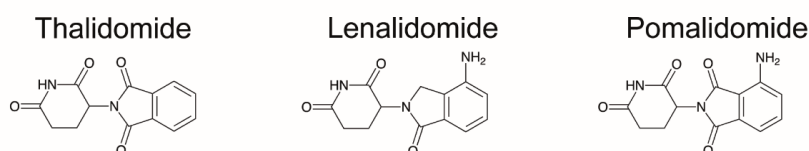
---

### **1.5.1 IMiDs/CELMoDs engage the *CRL4<sup>CRBN</sup>* E3 ubiquitin ligase and induce protein degradation**

One of the best described examples of drug-induced protein degradation stems from the mechanism of action of a class of anti-cancer and anti-inflammatory drugs called the immunomodulatory imide drugs (IMiDs) (Holstein, Hillengass and McCarthy, 2021). Thalidomide constitutes the first in class IMiD developed from a glutamic acid derivative in the 1950s by the German pharmaceutical company Chemie-Grünenthal, and prescribed back then over the counter as a non-barbiturate sedative and anti-emetic for the treatment of morning sickness in pregnant women (Vargesson, 2015; Liu *et al.*, 2017). In the early 1960s, thalidomide was identified to be teratogenic by two independent clinicians, William McBride in Australia and Widukind Lenz in Germany, leading to serious birth defects such as limb malformations (amelia, phocomelia, bone hypoplasia) and ear deformities (anotia, microtia and hearing loss), and was therefore withdrawn from the market (McBride, 1961; Lenz, 1962; Vargesson, 2015). Subsequent research into the mechanisms of thalidomide action revealed that the drug exhibited anti-angiogenic, immunomodulatory and anti-inflammatory effects and was found to be effective for the treatment of a complication of leprosy (erythema nodosum leprosum) and also multiple myeloma (MM) (Sheskin, 1965; D'Amato *et al.*, 1994; Singhal *et al.*, 1999; Sauer, Gu and Wartenberg, 2000). Since then, to increase efficacy and to avoid some of thalidomide's side effects, including constipation, peripheral neuropathy and sedation, more analogues were developed and approved for the treatment of several haematological malignancies (Richardson *et al.*, 2002, 2006, 2014;



Ghobrial and Rajkumar, 2003; Vargesson, 2015). Specifically, lenalidomide and pomalidomide constitute the first IMiDs to be developed and approved for therapeutic use. Their chemical structure closely resembles that of thalidomide, a glutarimide ring fused to a phthalimide ring (**Figure 1. 19**). Both IMiDs are composed of an unmodified glutarimide ring, however, lenalidomide's phthalimide ring contains an extra amino group and a carbonyl group substitution with a methylene group, whilst pomalidomide is characterised by a single amino group modification in its phthalimide ring (**Figure 1. 19**) (Quach *et al.*, 2010). Both lenalidomide and pomalidomide are effective anti-cancer agents for the treatment of relapsed and refractory MM (Richardson *et al.*, 2002, 2006, 2014). In addition, lenalidomide is also currently approved for the treatment of myelodysplastic syndromes with deletion of chromosome 5q, whilst, pomalidomide has been recently approved for the treatment of Kaposi's sarcoma (Alan *et al.*, 2005; Fenaux *et al.*, 2011; Polizzotto *et al.*, 2021).



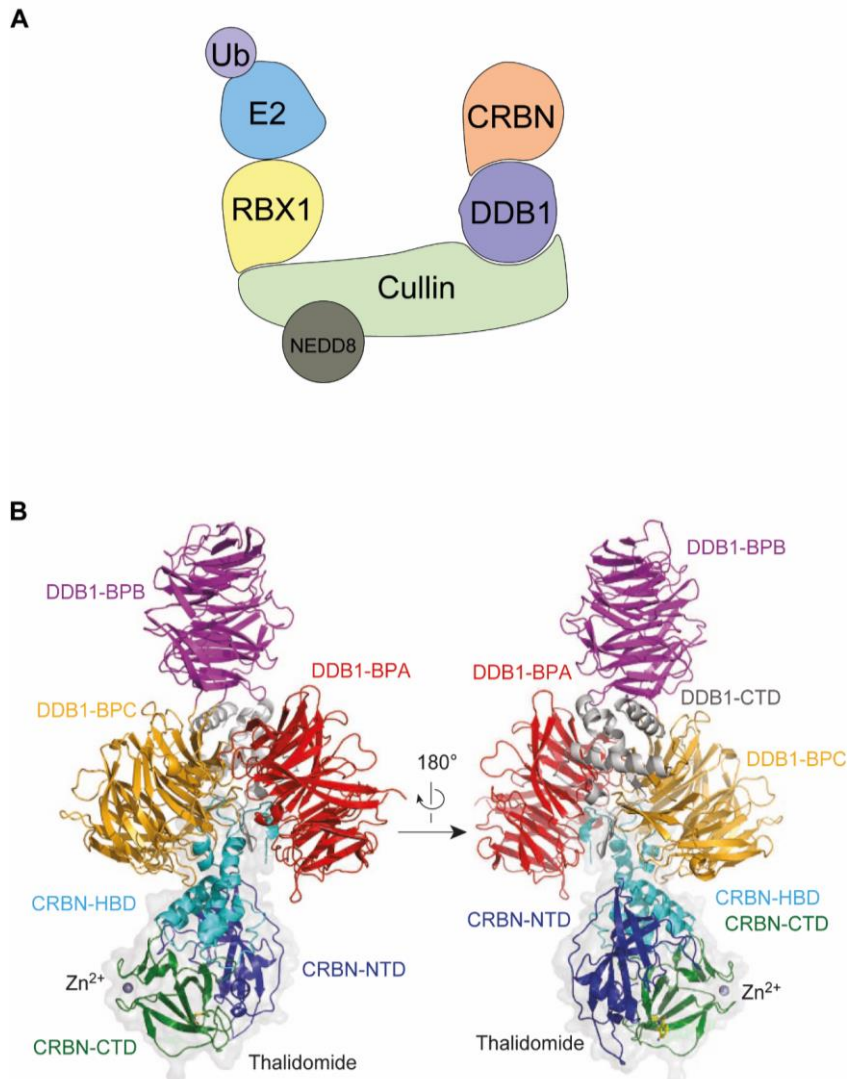
**Figure 1. 19** Chemical structures of IMiDs thalidomide, lenalidomide and pomalidomide.

In a 2010 seminal study by Hiroshi Handa's laboratory, a series of affinity purification assays identified CRBN as the primary binding target of IMiDs/CELMoDs (Ito *et al.*, 2010a). The group successfully identified CRBN thanks to the development and use of a single-step affinity purification bead technology, based on the carboxyl thalidomide derivative conjugated by an alkylation reaction to magnetic ferrite-glycidyl methacrylate (FG)-coated beads (Ito *et al.*, 2010a). Initially, mass spectrometry analysis of eluates from human HeLa cell extracts led to the identification of both CRBN and DDB1 (damaged DNA binding protein 1) as thalidomide-binding proteins (Ito *et al.*, 2010a). Nevertheless, *in vitro* thalidomide was demonstrated to directly interact only with purified recombinant CRBN, whilst DDB1 was shown by co-immunoprecipitation assays to indirectly associate with thalidomide through its interaction with CRBN (Ito *et al.*, 2010a). Moreover, CRBN was found through co-immunoprecipitation assays to also interact with components of an E3 ubiquitin ligase complex, namely Cullin 4A (CUL4A) and RING box-domain 1/ Regulator of Cullins 1 (RBX1/ROC1), thus suggesting a degradation-based function for CRBN (Ito *et al.*, 2010a).

Human *CRBN* gene is located on chromosome 3p26.3 and encodes for a 442 amino acid protein that is highly conserved from plants to mammals, with mutations on its genomic locus linked to autosomal recessive non-syndromic intellectual disability (Xu, Jiang and Jaffrey, 2013; Kim *et al.*, 2016; Sheereen *et al.*, 2017). CRBN acts as a substrate receptor for the cullin-RING CRL4<sup>CRBN</sup> E3 ubiquitin ligase; a multi-subunit protein complex comprised of either cullin 4A (CUL4A) or cullin 4B (CUL4B), RBX1/ROC1 and DDB1 (**Figure 1. 20 A**) (Angers *et al.*, 2006). CRBN is one of estimated ~60 DDB1-CUL4 associated factors

(DCAFs), proteins that bind DDB1 and recruit substrates to be ubiquitinated by CRL4 E3 ubiquitin ligase (Lee and Zhou, 2007). In addition, CRBN has been shown to localise in the nucleus, cytoplasm and plasma membrane and to be ubiquitously expressed in a diverse array of human tissue types including the brain, retina, testis, prostate, placenta, spleen, liver, kidney, pancreas, small intestine and skeletal muscle (J.J *et al.*, 2004; Xin *et al.*, 2008; Schapira *et al.*, 2019).

The first structure of CRBN-DDB1 bound to an IMiD was solved by Nicolas Thomä's laboratory in 2014, where they managed to crystallise a chimeric complex of human DDB1 (residues 1-1140) and chicken (*Gallus gallus*) CRBN (residues 1-445) bound to thalidomide (**Figure 1. 20 B**) (Fischer *et al.*, 2014). The high degree of conservation (92.06%) between chicken CRBN and its human counterpart allowed for structural insights to be directly inferred between the two. CRBN was identified to consist of three main subdomains, an N-terminal seven-stranded  $\beta$ -sheet (residues 1-185) containing a large, highly conserved Lon (ATP- dependent protease) domain, a central helical bundle domain (HBD, residues 186-317) which mediates binding to DDB1, and a C-terminal domain composed of eight  $\beta$ -sheets and harbouring a conserved  $Zn^{2+}$  binding site and a highly conserved across CRBN orthologues tri-tryptophan pocket (Fischer *et al.*, 2014). The phthalimide ring of IMiDs was found to be fully solvent exposed, whilst, their glutarimide ring was shown to bind in the tri-tryptophan pocket of CRBN via the formation of van der Waals interactions and hydrogen bonds with key residues residing within the tri-tryptophan pocket (Fischer *et al.*, 2014). DDB1 was found to consist of three seven-bladed WD40  $\beta$ -propellers (BPA, BPB, and



**Figure 1. 20 The CRL4<sup>CRBN</sup> E3 ubiquitin ligase. (A)** Schematic representation of the CRL4<sup>CRBN</sup> E3 ubiquitin ligase. **(B)** The crystal structure of the DDB1-CRBN complex. Cartoon representation of the structure of human DDB1 bound to *Gallus gallus* cereblon (CRBN) and thalidomide. In DDB1 the following domains are highlighted: BPA (red), BPB (magenta), BPC (orange) and DDB1-CTD (grey). In CRBN the following domains are highlighted: NTD (blue), HBD (cyan) and

CTD (green). The zinc ion is drawn as a grey sphere and thalidomide is depicted by a yellow stick structure. CRBN-NTD : CRBN amino-terminal domain; CRBN-CTD : CRBN carboxy-terminal domain; CRBN-HBD : CRBN helical bundle domain. (Adapted from Fischer *et al.*, 2014; Nabet *et al.*, 2018)

BPC) arranged in a triangular fashion, which corroborated previous structural studies of full-length DDB1 isolated or in complex with SV5 (simian virus 5V) protein of paramyxoviruses or DDB2 (damaged DNA binding protein 2) (Li *et al.*, 2006; Scrima *et al.*, 2008; Fischer *et al.*, 2014). DDB1 acts as an adaptor protein for the recruitment of CRBN to the CRL4 E3 ubiquitin ligase. The HBD domain and a C-terminal part of CRBN bind the cavity formed between the WD40  $\beta$ -propellers BPA and BPC of DDB1, whilst the remaining BPB propeller of DDB1 has been shown to mediate binding to the N-terminal region of CUL4A (Fischer *et al.*, 2014). Since then, these initial findings regarding the structural features of CRBN and DDB1 have also been verified and further described by other structural studies with both human CRBN and DDB1 (Matyskiela *et al.*, 2016, 2020; Petzold, Fischer and Thomä, 2016; Sievers *et al.*, 2018).

In a 2006 study, CUL4A has been shown to predominantly interact with the BPB propeller of DDB1 at its N-terminus through two separate interfaces, the H2 and H5 pair of helices and an N-terminal extension sequence, which is unique among all cullin family members (Angers *et al.*, 2006). Both members of the cullin 4 subfamily, CUL4A and CUL4B, retain a high homology between their amino acid sequence aside from a stretch of extra amino acids in the N-terminal region of CUL4B (Hannah and Zhou, 2015). Moreover, both CUL4A and CUL4B have demonstrated in several studies to exhibit a high degree of compensation

following loss of either paralogue, with CUL4A being highly expressed in small intestine, pancreas, ovaries, testis, prostate, thymus and in multiple types of blood cells, whilst CUL4B has been identified to be highly expressed in pituitary cells, smooth muscle cells, adipocytes, endocrine glands, pancreas, bone marrow, the cerebellum, the lower gastrointestinal tract, and the testes, thus suggesting a tissue specific function in the regulation of certain substrates (Hannah and Zhou, 2015). Both CUL4A and CUL4B act as assembly factors for the CUL4<sup>CRBN</sup> E3 ubiquitin ligase by providing a scaffold for the attachment of both CRBN-DDB1 and RBX1/ROC1; the RING E3 ubiquitin ligase enzyme which recruits E2-ubiquitin enzyme mediating the transfer of ubiquitin molecules to CRBN-bound substrates (Angers *et al.*, 2006).

The development of a new generation of IMiD molecules has occurred in the last few years. These novel compounds, also referred to as CRBN E3 ligase modulation drugs (CELMoDs), are thalidomide analogues constituted of a glutarimide ring fused to an altered phthalimide ring containing multiple chemical modifications and additions of novel chemical groups (**Figure 1. 21**) (Holstein, Hillengass and McCarthy, 2018). CELMoDs include several drugs that are currently under clinical development, such as CC-122 (avadomide), CC-220 (iberdomide), CC-92480 (mezigdomide) and CC-90009 and further experimental compounds such as CC-885 (**Figure 1. 21**). CC-122 has been shown in clinical studies to be an effective agent for the treatment of refractory diffuse large B-cell lymphoma (DLBCL) whilst, CC-220 has been shown to be effective in lenalidomide/pomalidomide-resistant MM (Hagner *et al.*, 2015; Krönke *et al.*, 2015; Cubillos-Zapata *et al.*, 2016; Matyskiela *et al.*, 2017; Bjorklund *et al.*, 2020). CC-92480 is currently in early phase clinical trials for relapsed refractory MM

whilst, CC-90009 is currently in clinical trials for the treatment of relapsed or refractory acute myeloid leukaemia (AML) (Hansen *et al.*, 2020, 2021; Surka *et al.*, 2021).

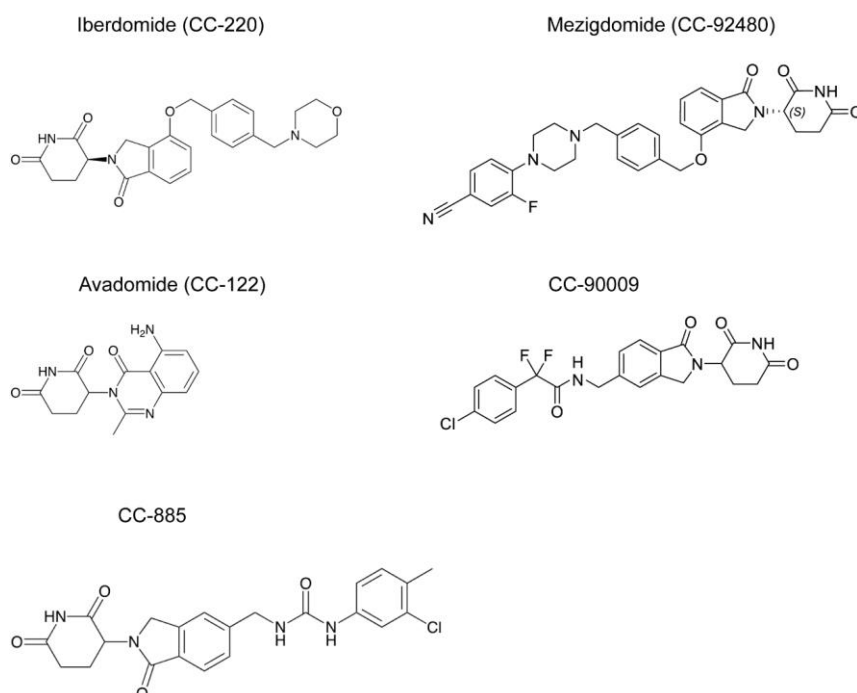
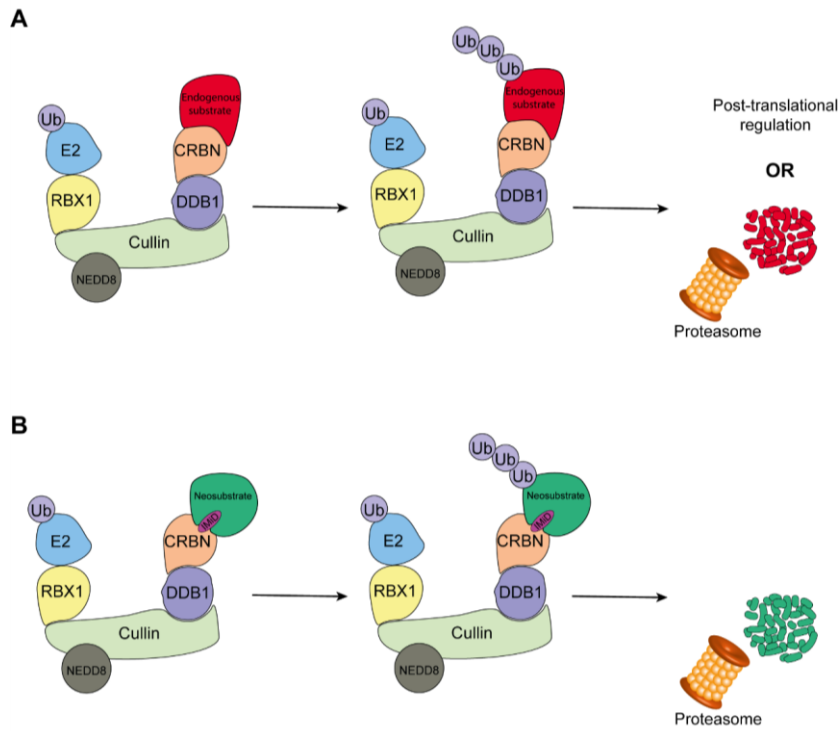


Figure 1. 21 Chemical structures of CELMoDs iberdomide (CC-220), mezigdomide (CC-92480), avadomide (CC-122) and CC-90009 and experimental compound CC-885.

Initial experimental work regarding the functional consequences of IMiD/CELMoD binding to CRL4<sup>CRBN</sup> E3 ubiquitin ligase in MM, suggested a model where IMiDs/CELMoDs alter the abundance, localisation, and activity of CRL4<sup>CRBN</sup> E3 ubiquitin ligase substrates (Zhu *et al.*, 2011; Lopez-Girona *et al.*, 2012). In 2014, three studies identified that IMiD-binding to CRBN leads to the recruitment of two zinc finger (ZF) transcription factors, Ikaros/IKZF1 and Aiolos/IKZF3, which are subsequently ubiquitinated and marked for proteasome-mediated degradation (Gandhi *et al.*, 2013; Krönke *et al.*, 2014; Lu *et al.*, 2014). Therefore, IMiDs/CELMoDs exhibit a drug-induced neomorphic activity, recruiting to the CRL4<sup>CRBN</sup> E3 ubiquitin ligase substrates such as Ikaros and Aiolos, which are termed neosubstrates to differentiate them from endogenous substrates that are physiologically regulated by this complex (**Figure 1. 22 A-B**) (Gandhi *et al.*, 2013; Krönke *et al.*, 2014; Lu *et al.*, 2014). The full repertoire of CRL4<sup>CRBN</sup> E3 ubiquitin ligase endogenous substrates remains largely elusive, with current putative endogenous substrates such as Meis2 (Myeloid ecotropic insertion site 2), GS (glutamine synthetase) and c-jun meriting further investigations into their mechanisms of CRBN-binding and biochemical regulation by the CRL4<sup>CRBN</sup> E3 ubiquitin ligase (Fischer *et al.*, 2014; Nguyen *et al.*, 2016; Yang *et al.*, 2018).

Besides Ikaros and Aiolos, several other CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates have been discovered (**Table 1. 1**). CK1 $\alpha$  (Casein kinase isoform  $\alpha$ ) was identified as a lenalidomide-dependent CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrate using SILAC (stable isotope labelling of amino acids in cell culture)-based quantitative mass spectrometry in myeloid cells (Krönke *et al.*, 2015). This study demonstrated that lenalidomide recruits CK1 $\alpha$  to the CRL4<sup>CRBN</sup> E3 ubiquitin





**Figure 1. 22 CRL4<sup>CRBN</sup> E3 ubiquitin ligase-mediated ubiquitination. (A)** The CRL4<sup>CRBN</sup> E3 ubiquitin ligase mediates ubiquitination of endogenous substrates bound to cereblon (CRBN), which results in their post-translational regulation or targeting to the 26S proteasome to be degraded. **(B)** Binding of an IMiD/CELMoD to CRBN results in the recruitment of neosubstrates to the CRL4<sup>CRBN</sup> E3 ubiquitin ligase, which are subsequently ubiquitinated and marked for proteasomal degradation. (Adapted from Neklesa *et al.*, 2011; Nabet *et al.*, 2018)

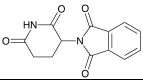
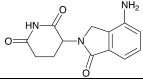
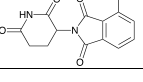
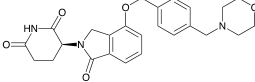
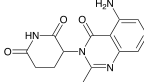
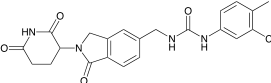
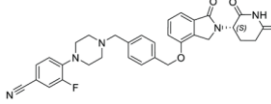
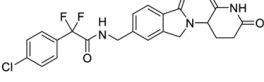
Degradar	2D structure	C2H2 zinc finger containing proteins degraded			Non-zinc finger containing proteins degraded	
		Ikaros (IKZF1)	Aiolos (IKZF3)	ZFP91	GSPT1	Casein Kinase 1 $\alpha$
Thalidomide		X	X			
Lenalidomide		X	X			X
Pomalidomide		X	X	X		
Iberdomide/ CC-220		X	X	X		
Avadomide/ CC-122		X	X	X		
CC-885		X	X		X	
Mezigdomide/ CC-92480		X	X	X		
CC-90009					X	

Table 1. 1 Summary of IMiD/CELMoD neosubstrate specificity.

ligase where it is ubiquitinated and sent for proteasomal degradation (Krönke *et al.*, 2015). Other IMiDs/CELMoDs tested did not affect CK1 $\alpha$  protein levels, with pomalidomide having only weak effects on CK1 $\alpha$  protein levels, thus demonstrating substrate specificity in the degradation of CK1 $\alpha$  by lenalidomide (Krönke *et al.*, 2015). CK1 $\alpha$  is encoded by the *CSNK1A1* gene on the long arm of chromosome 5 and one of its primary functions is the negative regulation of the tumour suppressor protein p53 (Krönke *et al.*, 2015). Since CK1 $\alpha$  gene (*CSNK1A1*) is located in the common deleted region for myelodysplastic syndrome del (5q) and is a haploinsufficient gene, the therapeutic effect of lenalidomide in patients with myelodysplastic syndrome del (5q) can be explained by the degradation of CK1 $\alpha$  restoring p53 activity and growth inhibition in bone marrow myelocyte cancer cells (Schneider *et al.*, 2014; Krönke *et al.*, 2015). This study highlights the fact that even subtle chemical modifications in the structure of IMiDs/CELMoDs, such as one less carbonyl group in the phthalimide ring of lenalidomide when compared to pomalidomide, can result in profound differences in the recruitment of neosubstrates to CRBN (**Table 1.1**) (Krönke *et al.*, 2015).

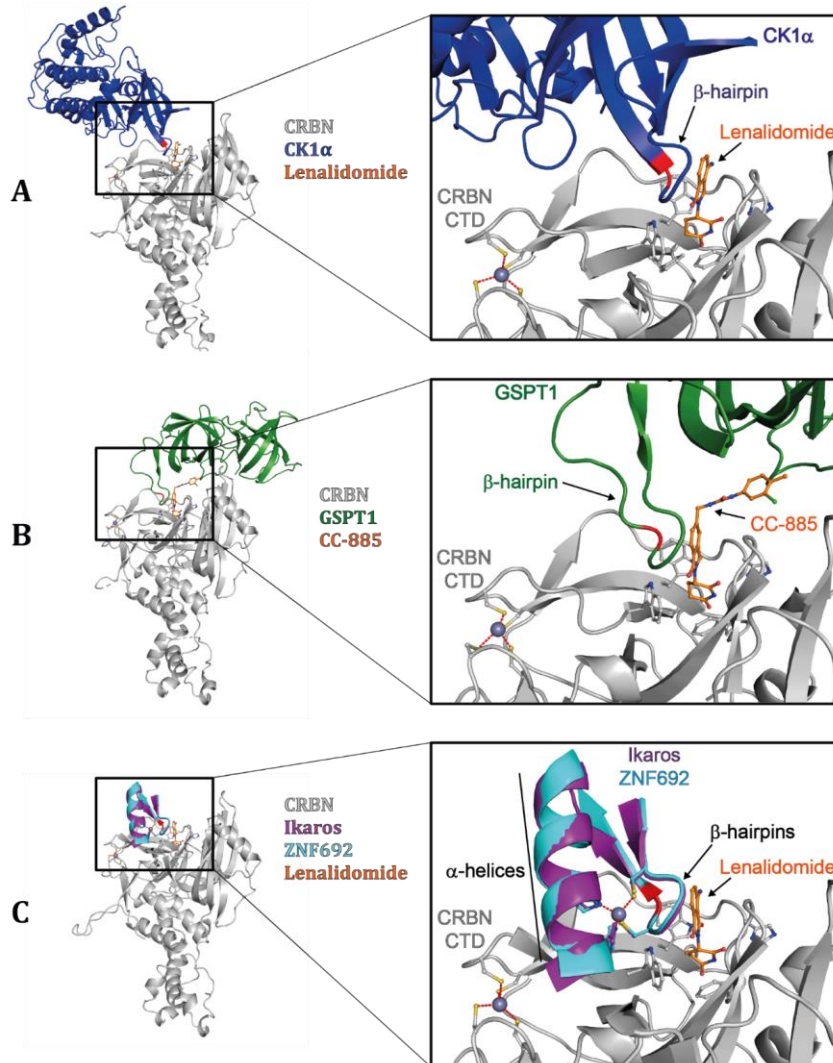
Similarly, in another such study, the thalidomide analogue CC-885 was identified by mass spectrometry analysis of eluates from HEK 293T cell extracts to recruit the translation termination factor GSPT1 (G1 to S Phase Transition 1) to the CRL4<sup>CRBN</sup> E3 ubiquitin ligase (Matyskiela *et al.*, 2016). CC-885 has been found to exert a potent antitumour activity in both haematological and epithelial cancers (Matyskiela *et al.*, 2016). CC-885 is composed of an unmodified glutarimide ring, an isoindolinone moiety and a methylene urea group (**Figure 1.21**). CC-885 has been shown to mediate the ubiquitination and degradation of

GSPT1 resulting in cytotoxicity and is also associated with the degradation of multiple other neosubstrates including CK1 $\alpha$ , Ikaros, and Aiolos (**Table 1. 1**) (Matyskiela *et al.*, 2016). Interestingly, other IMiDs such as lenalidomide and pomalidomide, have been shown to not degrade GSPT1, which has been previously speculated to be due to the lack of CC-885 urea and chloro-methyl-phenyl moieties possibly contributing to the complex affinity through interactions with domain 3 of GSPT1 (**Figure 1. 19; Figure 1. 21**) (Matyskiela *et al.*, 2016). Further to GSPT1, the ZF transcription factor ZFP91 (zinc finger protein 91) has been identified as a lenalidomide-dependent CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrate using a pulse-chase SILAC mass spectrometry-based proteomic approach (An, Charles M Ponthier, *et al.*, 2017). This study also showed that pomalidomide can induce the degradation of ZFP91 similarly to lenalidomide, whilst thalidomide's effect on ZFP91 protein levels was found to be minimal (**Table 1. 1**) (An, Charles M Ponthier, *et al.*, 2017).

### **1.5.2 Exploiting CRL4<sup>CRBN</sup> E3 ubiquitin ligase to develop a novel targeted protein degradation system**

A  $\beta$ -hairpin loop with a key glycine residue at the apex has been previously identified from in vitro mutagenesis assays and structural studies to mediate the binding of neosubstrates to the IMiD/CELMoD-bound CRL4<sup>CRBN</sup> E3 ubiquitin ligase (**Figure 1. 23 A-C**). This  $\beta$ -hairpin loop can be found as part of the tertiary structure in neosubstrates such as GSPT1 and CK1 $\alpha$  or within a C2H2-type ZF domain in ZF transcription factor neosubstrates, such as Ikaros, Aiolos, and ZFP91 (**Figure 1. 23 A-C**) (Krönke *et al.*, 2014; Matyskiela *et al.*, 2016, 2020; Petzold, Fischer and Thomä, 2016; An, Charles M Ponthier, *et al.*, 2017; Sievers *et al.*, 2018). I hypothesised that this minimal degron motif could be further explored and exploited for repurposing the mechanism of IMiD/CELMoD-dependent degradation of neosubstrates and developing a novel chemical-biology tool. During the course of this work, in 2019, William Kaelin's laboratory proved the concept by fusing an Aiolos-based degron to a set of proteins and targeting them for proteasomal degradation following IMiD-addition in the system (Koduri *et al.*, 2019). However, this Aiolos-based degron was found to only degrade C-terminally tagged proteins, therefore highlighting a need for a more systematic evaluation of potential degrons and with better matching to a cognate IMiD/CELMoD agent, in order to achieve optimal target degradation (Koduri *et al.*, 2019). Moreover, further examination is warranted into understanding the relevant importance of residues flanking the  $\beta$ -hairpin loop and other proximal domains in neosubstrates for binding the IMiD/CELMoD-CRBN complex. Furthermore, the fact that CRBN is highly conserved across species and

ubiquitously expressed in different cellular compartments and tissues make it ideal for the development of a novel and widely-applicable chemical-biology tool (J.J *et al.*, 2004; Xin *et al.*, 2008; Schapira *et al.*, 2019). The work presented in this thesis further explores the concept of what constitutes a functional degron motif in CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates and describes the identification of degron-containing domains (DCD) based on the sequences of known neosubstrates; which when fused to a target protein induce a potent, robust, and reversible degradation following IMiD/CELMoD treatment. Twenty-six DCDs were designed and systematically evaluated for inducing target protein degradation and enabling functional characterisation of the target under examination. These DCD motifs were subsequently exploited for the development of iTAG (Inducible and TArgeted protein deGradation), a novel chemical-biology tool based on IMiDs/CELMoDs and a novel inducible degradation TAG comprised of a DCD fused to EGFP for visualisation. The data presented here demonstrate that the iTAG system allows for the degradation of both nuclear and cytoplasmic proteins when the degradation TAG is fused to either terminus. Furthermore, this novel system can be used for *in vivo* target degradation, with effective loss of target as early as 4h. Therefore, the iTAG system constitutes a modular and versatile tool, enabling the acute degradation of specific proteins to study their function in cells and *in vivo*.



**Figure 1.23 Binding mode of various known classes of neosubstrates on CRBN.** (A) Complex formed between human CRBN, Lenalidomide and human CK1 $\alpha$  (PDB 5FQD) (Petzold, Fischer and Thomä, 2016). (B) Complex formed between human CRBN, CC-885 and human GSPT1 (PDB 5HXB) (Matyskiela *et al.*, 2016). (C) Overlay of the complexes formed between human CRBN, Lenalidomide and human Ikaros or ZNF692 (PDBs 6H0F and 6H0G, respectively) (Sievers *et al.*, 2018), highlighting the conserved binding mode of both zinc-finger proteins. The crucial Gly of the G-loop degenon is highlighted in red. CRBN CTD : CRBN carboxy-terminal domain.

## 2. MATERIALS AND METHODS

### *2.1 Cell lines and cell culture*

Cell lines were either sourced from collaborators within the Institute of Cancer Research (ICR), the ICR Cancer Therapeutics Unit (CTU) cell bank or external collaborators. h-TERT (human telomerase reverse transcriptase) immortalised human mammary epithelial cells (HMECs) (kindly provided by Prof Martin Eilers; Julius Maximilian University of Würzburg, Germany) were maintained in Dulbecco Modified Eagle's Medium DMEM/F-12 (1:1) + GlutaMAX™-I (Gibco/Life Technologies), containing 10% (v/v) heat-inactivated foetal calf serum (FCS; Gibco/Life Technologies), supplemented with 15 mM HEPES [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid] buffer (Sigma-Aldrich), 10 µg/ml insulin, human recombinant, zinc solution (Gibco/Life Technologies), 0.5 µg/ml hydrocortisone (Sigma-Aldrich) and 20 ng/ml epidermal growth factor (EGF) recombinant human protein (Life Technologies). HEK 293T (human embryonic kidney 293T) cells and the human triple-negative breast adenocarcinoma MDA-MB-231 cells were maintained in DMEM (Gibco/Life Technologies), containing 25 mM D-glucose (dextrose), 4 mM L-glutamine, 1 mM sodium pyruvate and supplemented with 10% (v/v) FCS and 1% non-essential amino acids (Gibco/Life Technologies). Kelly human



neuroblastoma cells were maintained in Roswell Park Memorial Institute 1640 (RPMI 1640) Medium (Gibco/Life Technologies) containing GlutaMAX™, supplemented with 10% (v/v) heat-inactivated FCS. Cell cultures were incubated at 37°C in a humidified atmosphere with 5% CO<sub>2</sub>. Cells were allowed to grow to a maximum confluency of 70-85% before being passaged by using either TrypLE™ Express Enzyme (no phenol red) or TrypLE™ Select Enzyme (no phenol red) (ThermoFischer Scientific) to detach them from their flasks/plates. Subsequently, cells were transferred to 15 ml falcon tubes (Falcon) and centrifuged for 3min at 1200 rpm in a 5810 or 5417 centrifuge (Eppendorf) and resuspended in fresh medium. Cell counting was performed using a Countess™ II Automated Cell Counter (ThermoFischer Scientific) and following manufacturer's instructions. All cell lines were authenticated by STR (short tandem repeat DNA analysis) profiling at the ICR tumour profiling unit and were regularly controlled for mycoplasma contamination. For long-term storage, cells were resuspended in FCS containing 10% dimethyl sulfoxide (DMSO; Sigma-Aldrich) before being transferred into Nalgene® cryovials (ThermoFischer Scientific) and pre-cooled for at least 24h at -80°C before being transferred to liquid nitrogen storage (-195.8°C to -210°C). Cell recovery from liquid nitrogen was performed by thawing the cells at 37°C in a water bath followed by resuspension in their cognate media and centrifuging for 3min at 1200 rpm in a 5810 or 5417 centrifuge. The cell pellet was then resuspended in fresh cell culture medium before being transferred to a tissue culture flask (Corning; ThermoFischer Scientific).

## 2.2 Plasmids

Plasmids were commercially generated (GeneArt, ThermoFischer Scientific) (**Table 2. 1**), constructed in-house (**Table 2. 2**) or generated by our collaborators (**Table 2. 3**). The following backbone vectors were used for plasmid cloning: pLVX-TetOne-PURO (Takara/Clontech) (**Figure 2. 1**), pET-48B(+) (Merck Biosciences) (**Figure 2. 2**), pHAGE-PGK (Addgene) (**Figure 2. 3**) and pFBDM (Addgene) (**Figure 2. 4**). All plasmids constructed in-house were generated using the In-Fusion® HD Cloning Kit (Takara/Clontech) according to manufacturer's instructions (**Figure 2. 5 A; Figure 2. 6; Figure 2. 7 A**). The pLVX-TetOne-PURO vector was linearized with ECORI (New England Biolabs) and AgeI (New England Biolabs) restriction digestion enzymes in cutsmart buffer (New England Biolabs) (**Figure 2. 5 B; Figure 2. 6 A; Figure 2. 7 B**). The pET-48B(+) vector was linearized with XmaI (New England Biolabs) and XhoI (New England Biolabs) restriction digestion enzymes in cutsmart buffer. Restriction digestions were carried out in a waterbath at 37°C for 30-90min. The following sequences were used for PCR-amplification of each target cDNA: cMYC (UniProtKB - P01106-1, canonical isoform 1), EZH2 (UniProtKB - UniProtKB - Q15910-2, isoform 2), KRAS[G12V] (UniProtKB - P01116-2, isoform 2B, G12V mutant variant), firefly luciferase [Fluc; UniProtKB - Q27758-1 (Q27758\_P HOPY)] and enhanced green fluorescent protein [EGFP; UniProtKB - C5MKY7-1 (C5MKY7\_HCMV)] (**Table 2. 4**). For PCR-amplification the CloneAmp HiFi PCR Premix (Takara/Clontech) was used according to

manufacturer's instructions (**Figure 2. 5 C-D; Figure 2. 6 B; Figure 2.7 B-D**).

For each PCR-amplification reaction, 100-200 ng of donor vector containing target protein cDNA and 0.25  $\mu$ M of primer mix were used. Following optimization, the following thermocycler conditions were used: 35 cycles of denaturation at 98°C for 10 seconds, primer annealing at 63°C for 15 seconds and elongation at 72°C for 30 seconds to 1min. Both plasmid digestion and PCR-amplification were verified by agarose gel electrophoresis. Agarose gels were prepared at final concentrations of 1% (w/v) by dissolving 0.4 g of UltraPure™ agarose (Invitrogen) in 40 ml TAE buffer (40 mM Tris pH7.6, 20 mM acetic acid and 1mM EDTA), using brief heating in a microwave oven. GelGreen™ nucleic acid stain (Biotium) was added at a ratio of 1: 10000 in the agarose gel mixture prior to solidification to allow visualization of DNA bands. Purple Gel Loading Dye (New England Biolabs) was added to the samples prior to gel loading. The Quick-Load® 50bp, 100bp or 1kb DNA Ladders (New England Biolabs) were used as markers. Gel electrophoresis was done at 120V for 40min. DNA band visualisation was performed using the gel imaging U:Genius3 UV Transilluminator (AlphaMetrix Biotech). DNA extraction was done using the Nucleospin Gel and PCR Clean-up kit (MACHEREY-NAGEL), according to manufacturer's protocol. Ligation of the PCR-amplicons was done using the In-Fusion® HD Cloning Kit (Clontech) according to manufacturer's instructions. The plasmids generated by the In-Fusion® HD ligation reaction were subsequently used to transform either DH5 $\alpha$  competent cells (ThermoFischer Scientific) or Stellar™ chemically competent cells (Takara/Clontech). Bacterial cells were thawed on ice for 20-30min before being incubated for 30min on ice with 2.5  $\mu$ l or 5  $\mu$ l of In-Fusion® HD ligation

reaction. Then, bacterial cells underwent heat shock transformation in a waterbath at 42°C for 45 seconds, followed by a brief incubation period on ice for 2min. Pre-warmed stable outgrowth medium (SOC; New England Biolabs, Takara/Clontech) was then added to the cells which were then incubated for 1h in a shaking incubator at 220 rpm and 37°C. Bacterial cell cultures were then plated at different ratios on sterile LB (Lysogeny Broth) agar plates containing either a 100 µg/ml ampicillin or 50 µg/ml kanamycin and placed overnight in a shaking incubator at 220 rpm and 37°C. Subsequently, single colonies were selected and inoculated in 5 ml of LB medium to be incubated overnight in a shaking incubator at 220 rpm and 37°C. Cultures were allowed to grow and plasmids were subsequently eluted using either the QIAprep® Plasmid Miniprep kit (Qiagen) or the QIAprep® Plasmid Maxi kit (Qiagen), according to manufacturer's instructions. The NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies) or NanoPhotometer™ N60 Micro-Volume UV-VIS spectrophotometer (IMPLEN) were used to measure the concentration and purity of the extracted plasmid DNA. For long-term storage, overnight bacterial cultures were mixed with glycerol at a ratio of 1:1 and placed at -80°C. To verify plasmids, DNA extracts were digested with either ECORI and AgeI or XmaI and XhoI restriction enzymes in a waterbath for 30-90min at 37°C and run on an agarose gel to determine DNA band length as described above. Verified plasmids were subsequently submitted for DNA sequencing to Source Bioscience according to company specifications and sequencing data were aligned to reference sequences using the SerialCloner 2.6.1 software.

	Backbone Vector	Fusion target protein encoded
1	pLVX-TetOne-PURO	cMYC-EGFP-25mer
2	pLVX-TetOne-PURO	cMYC-EGFP(P2A)25mer
3	pLVX-TetOne-PURO	cMYC-EGFP-DCD1
4	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD1
5	pLVX-TetOne-PURO	cMYC-EGFP-DCD2
6	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD2
7	pLVX-TetOne-PURO	cMYC-EGFP-DCD3
8	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD3
9	pLVX-TetOne-PURO	cMYC-EGFP-DCD4
10	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD4
11	pLVX-TetOne-PURO	cMYC-EGFP-DCD5
12	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD5
13	pLVX-TetOne-PURO	cMYC-EGFP-DCD6
14	pLVX-TetOne-PURO	cMYC(P2A)EGFP-DCD6
15	pLVX-TetOne-PURO	cMYC-EGFP-DCD7
16	pLVX-TetOne-PURO	cMYC(P2A)EGFP-DCD7
17	pLVX-TetOne-PURO	cMYC-EGFP-DCD8
18	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD8
19	pLVX-TetOne-PURO	cMYC-EGFP-DCD9
20	pLVX-TetOne-PURO	cMYC(P2A)EGFP-DCD9
21	pLVX-TetOne-PURO	cMYC-EGFP-DCD10
22	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD10
23	pLVX-TetOne-PURO	cMYC-EGFP-DCD11
24	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD11
25	pLVX-TetOne-PURO	cMYC-EGFP-DCD12
26	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD12
27	pLVX-TetOne-PURO	cMYC-EGFP-DCD13
28	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD13
29	pLVX-TetOne-PURO	cMYC-EGFP-DCD14
30	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD14
31	pLVX-TetOne-PURO	cMYC-EGFP-DCD15
32	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD15
33	pLVX-TetOne-PURO	cMYC-EGFP-DCD16

	Backbone Vector	Fusion target protein encoded
34	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD16
35	pLVX-TetOne-PURO	cMYC-EGFP-DCD17
36	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD17
37	pLVX-TetOne-PURO	cMYC-EGFP-DCD18
38	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD18
39	pLVX-TetOne-PURO	cMYC-EGFP-DCD19
40	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD19
41	pLVX-TetOne-PURO	cMYC-EGFP-DCD20
42	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD20
43	pLVX-TetOne-PURO	cMYC-EGFP-DCD21
44	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD21
45	pLVX-TetOne-PURO	cMYC-EGFP-DCD22
46	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD22
47	pLVX-TetOne-PURO	cMYC-EGFP-DCD23
48	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD23
49	pLVX-TetOne-PURO	cMYC-EGFP-DCD24
50	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD24
51	pLVX-TetOne-PURO	cMYC-EGFP-DCD25
52	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD25
53	pLVX-TetOne-PURO	cMYC-EGFP-DCD26
54	pLVX-TetOne-PURO	cMYC-EGFP(P2A)DCD26
55	pLVX-TetOne-PURO	DCD19-EGFP-EZH2
56	pLVX-TetOne-PURO	DCD19(P2A)EGFP-EZH2
57	pLVX-TetOne-PURO	EZH2-EGFP-DCD19
58	pLVX-TetOne-PURO	EZH2-EGFP(P2A)DCD19
59	pLVX-TetOne-PURO	DCD19-EGFP-KRAS[G12V]
60	pLVX-TetOne-PURO	DCD19(P2A)EGFP-KRAS[G12V]
61	pLVX-TetOne-PURO	KRAS[G12V]-EGFP-DCD19
62	pLVX-TetOne-PURO	KRAS[G12V]-EGFP(P2A)DCD19
63	pLVX-TetOne-PURO	DCD19-EGFP-cMYC
64	pLVX-TetOne-PURO	DCD19(P2A)EGFP-cMYC
65	pLVX-TetOne-PURO	DCD23-EGFP-cMYC
66	pLVX-TetOne-PURO	DCD23(P2A)EGFP-cMYC

Table 2. 1 Vectors commercially generated.

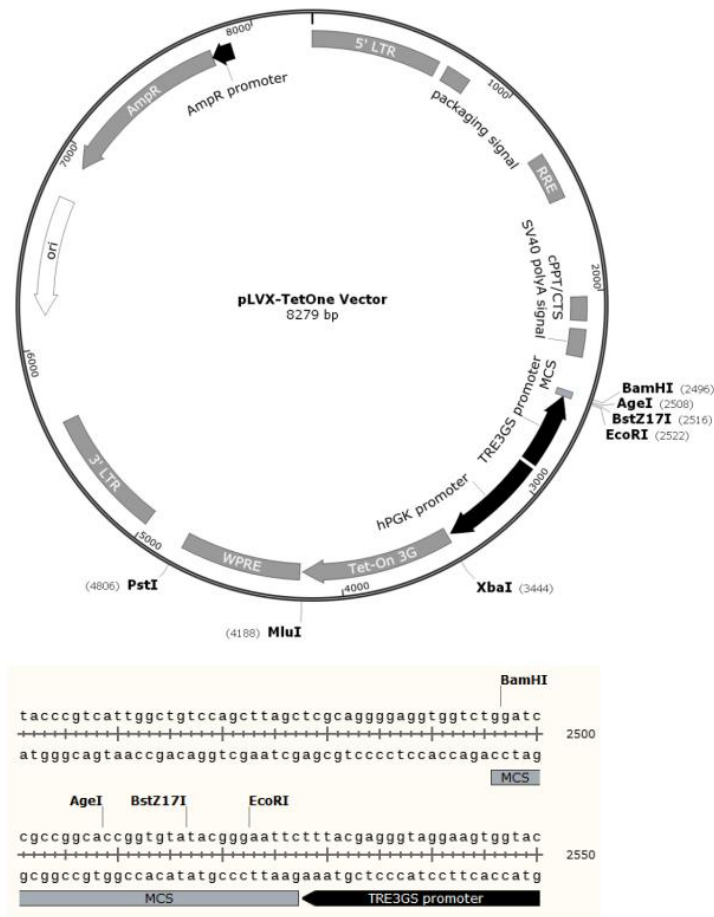
	Backbone Vector	Fusion target protein encoded
1	pET48b	DCD19
2	pLVX-TetOne-PURO	DCD23-EGFP-EZH2
3	pLVX-TetOne-PURO	DCD23(P2A)EGFP-EZH2
4	pLVX-TetOne-PURO	EZH2-EGFP-DCD23
5	pLVX-TetOne-PURO	EZH2-EGFP(P2A)DCD23
6	pLVX-TetOne-PURO	DCD23-EGFP-KRAS[G12V]
7	pLVX-TetOne-PURO	DCD23(P2A)EGFP-KRAS[G12V]
8	pLVX-TetOne-PURO	KRAS[G12V]-EGFP-DCD23
9	pLVX-TetOne-PURO	KRAS[G12V]-EGFP(P2A)DCD23
10	pLVX-TetOne-PURO	EZH2
11	pLVX-TetOne-PURO	Fluc-EGFP-DCD23
12	pLVX-TetOne-PURO	Fluc-EGFP(P2A)DCD23
13	pLVX-TetOne-PURO	EGFP
14	pLVX-TetOne-PURO	DCD23-EGFP
15	pLVX-TetOne-PURO	DCD23(P2A)EGFP
16	pLVX-TetOne-PURO	EGFP-DCD23
17	pLVX-TetOne-PURO	EGFP(P2A)DCD23
18	pLVX-TetOne-PURO	DCD23mut1-EGFP
19	pLVX-TetOne-PURO	DCD23mut1(P2A)EGFP
20	pLVX-TetOne-PURO	EGFP-DCD23mut1
21	pLVX-TetOne-PURO	EGFP(P2A)DCD23mut1
22	pLVX-TetOne-PURO	DCD23mut2-EGFP
23	pLVX-TetOne-PURO	DCD23mut2(P2A)EGFP
24	pLVX-TetOne-PURO	EGFP-DCD23mut2
25	pLVX-TetOne-PURO	EGFP(P2A)DCD23mut2
26	pLVX-TetOne-PURO	DCD23mut3-EGFP
27	pLVX-TetOne-PURO	DCD23mut3(P2A)EGFP
28	pLVX-TetOne-PURO	EGFP-DCD23mut3
29	pLVX-TetOne-PURO	EGFP(P2A)DCD23mut3

Table 2. 2 Vectors generated in-house.

	<b>Backbone Vector</b>	<b>Fusion protein encoded</b>
<b>1</b>	pET48b	DCD1
<b>2</b>	pET48b	DCD23
<b>3</b>	pFBDM	6His-ZZ-CRBN
<b>4</b>	pFBDM	DDB1-Strep
<b>5</b>	pHAGE-PGK	FLAG-HA-DCD23-RBM39
<b>6</b>	pHAGE-PGK	FLAG-HA-DCD23-CDK9
<b>7</b>	pHAGE-PGK	FLAG-HA-DCD23-MAPK9
<b>8</b>	pHAGE-PGK	FLAG-HA-DCD23-HPRT1
<b>9</b>	pHAGE-PGK	FLAG-HA-DCD23-SMS
<b>10</b>	pHAGE-PGK	FLAG-HA-DCD23-MAVS

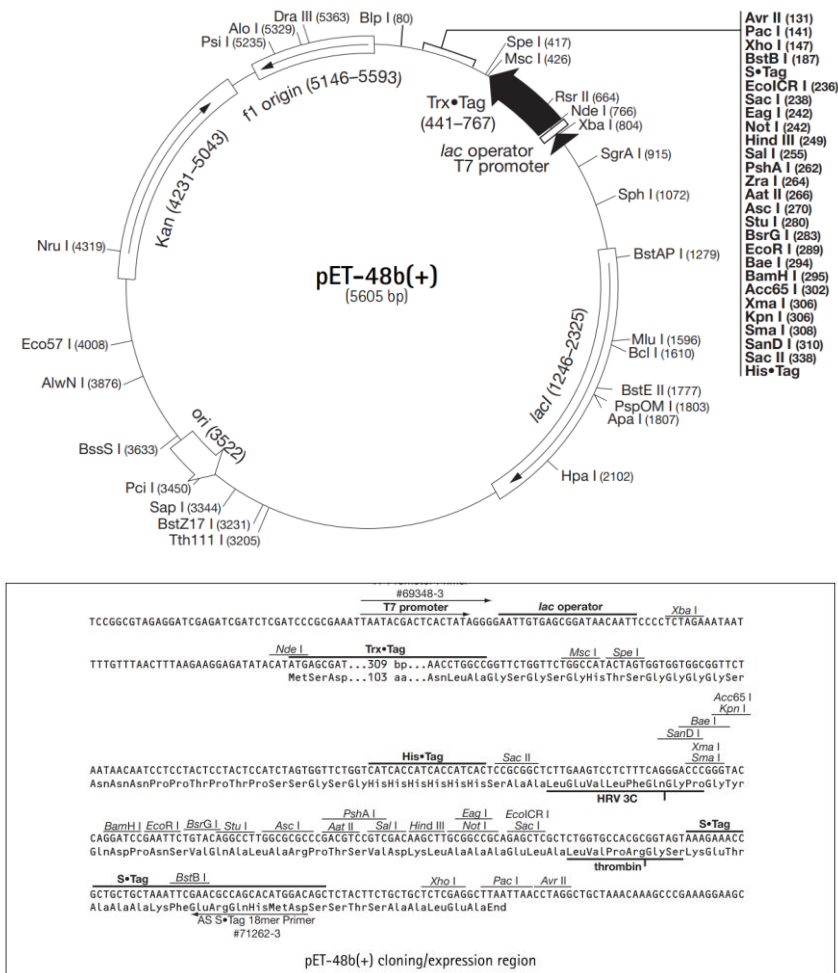
**Table 2. 3 Vectors generated by collaborators.**





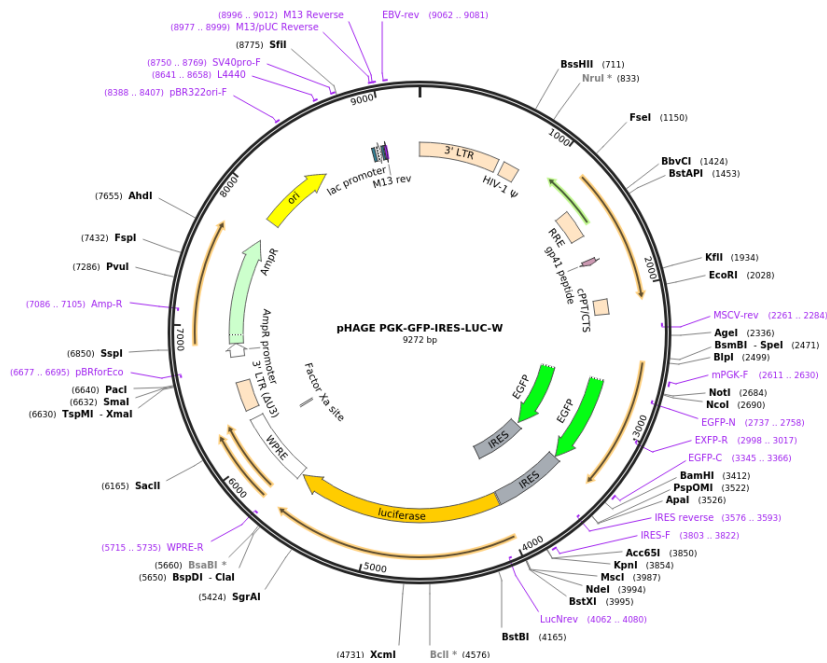
**Figure 2. 1 pLVX\_TetOne\_PURO lenti-vector map and multiple cloning site (MCS).** A plasmid of 8279bp designed to express inserted cDNA sequences from the *TRE3GS* promoter. *hPGK* promoter controls the expression of Tet-On 3G, a modified, improved Tet-On advanced transcriptional activator protein (transactivator). In the presence of DOX (doxycycline), Tet-On 3G binds to *TRE3GS* promoter and stimulates expression of the transgene inserted in the MCS. Plasmid map also shows key genes and selected restriction endonuclease sites. Indicated on the map are Ori (origin of DNA replication), AmpR (ampicillin resistance gene), 5' LTR/3' LTR (5' long terminal repeat/ 3' long terminal repeat), RRE (Rev response element), Cppt/ CTS (central polypurine tract/central termination sequence) and WPRE (Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element). Numbers indicate the nucleotide location of the restriction endonuclease site on the map of pLVX\_TetOne\_PURO: PstI (4806), MluI (4188), XbaI (3444), BamHI (2496), AgeI (2508), BstZ171 (2516) and EcoRI (2522).

Commented [SN3]: Expanded figure legend description



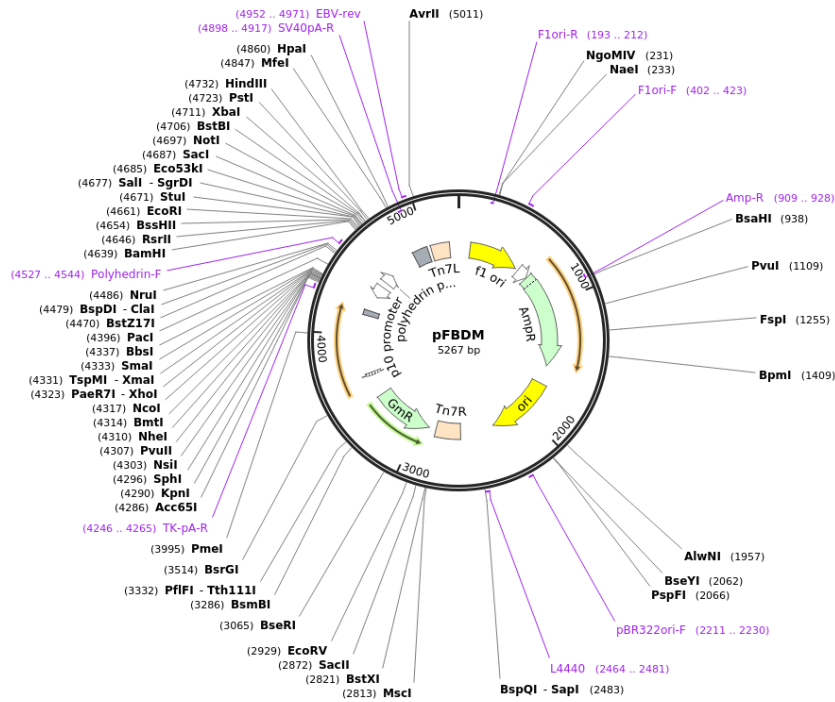
**Figure 2. 2 pET\_48b (+) vector map and MCS.** The pET\_48b (+) plasmid (5605bp) is designed for cloning and high-level expression of target proteins fused with the 109aa Trx•Tag™ thioredoxin protein that is cleavable with the human rhinovirus (HRV) 3C protease. Cloning sites are available for producing fusion proteins also containing a cleavable N-terminal His•Tag® sequence for detection and purification. Vector also contains an optimized RBS (ribosome binding site), a T7lac promoter, and the coding sequence for the HRV 3C protease cleavage site for N-terminal fusion tag removal. The MCS and vector backbone contain several restriction enzyme sites, with the numbers next to them indicating their respective nucleotide location on the pET\_48b (+) vector map.

**Commented [SN4]:** Expanded figure legend description



**Figure 2. 3 pHAGE-PGK vector map.** A bicistronic lentiviral plasmid of 9272bp designed to express EGFP and luciferase under the control of a *hPGK* promoter. Plasmid map also shows key genes and selected restriction endonuclease sites. Indicated on the map are Ori (origin of DNA replication), AmpR (ampicillin resistance gene), 5' LTR/3' LTR (5' long terminal repeat/ 3' long terminal repeat), RRE (Rev response element), CcpI/ CTS (central polypurine tract/central termination sequence), IRES (internal ribosome entry site) and WPRE (Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element). pHAGE-PGK plasmid contains several restriction enzyme sites, with the numbers next to them indicating their respective nucleotide location on the vector map.

Commented [SN5]: Expanded figure legend description



**Figure 2. 4 pFBDM vector map.** A plasmid of 5267bp designed to express inserted cDNA sequences in insect cells from the p10 baculovirus promoter. Plasmid map also shows key genes and selected restriction endonuclease sites. Indicated on the map are Ori (origin of DNA replication), AmpR (ampicillin resistance gene), GmR (gentamicin resistance gene), Tn7R/Tn7L (mini-Tn7 elements), f1 Ori (phage-derived origin of DNA replication) and polyhedrin promoter. pFBDM plasmid contains several restriction enzyme sites, with the numbers next to them indicating their respective nucleotide location on the vector map.

Commented [SN6]: Expanded figure legend description

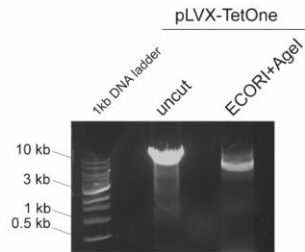
	Target Protein PCR-amplified	Target protein cDNA amplified from donor vectors
1	DCD19	pLVX-TetOne-PURO(DCD19-EGFP-cMYC)
2	DCD23-EGFP-EZH2	pLVX-TetOne-PURO(cMYC-EGFP-DCD23) and pLVX-TetOne-PURO(DCD19-EGFP-EZH2)
3	DCD23(P2A)EGFP-EZH2	pLVX-TetOne-PURO(cMYC-EGFP-DCD23) and pLVX-TetOne-PURO(DCD19(P2A)EGFP-EZH2)
4	EZH2-EGFP-DCD23	pLVX-TetOne-PURO(EZH2-EGFP-DCD19) and pLVX-TetOne-PURO(cMYC-EGFP-DCD23)
5	EZH2-EGFP(P2A)DCD23	pLVX-TetOne-PURO(EZH2-EGFP(P2A)DCD19) and pLVX-TetOne-PURO(cMYC-EGFP-DCD23)
6	DCD23-EGFP-KRAS[G12V]	pLVX-TetOne-PURO(cMYC-EGFP-DCD23) and pLVX-TetOne-PURO(DCD19-EGFP-KRAS[G12V])
7	DCD23(P2A)EGFP-KRAS[G12V]	pLVX-TetOne-PURO(cMYC-EGFP-DCD23) and pLVX-TetOne-PURO(DCD19(P2A)EGFP-KRAS[G12V])
8	KRAS[G12V]-EGFP-DCD23	pLVX-TetOne-PURO(KRAS[G12V]-EGFP-DCD19) and pLVX-TetOne-PURO(cMYC-EGFP-DCD23)
9	KRAS[G12V]-EGFP(P2A)DCD23	pLVX-TetOne-PURO(KRAS[G12V]-EGFP(P2A)DCD19) and pLVX-TetOne-PURO(cMYC-EGFP-DCD23)
10	EZH2	pLVX-TetOne-PURO(EZH2-EGFP-DCD19)
11	Fluc-EGFP-DCD23	pGL4.10[luc2] (Promega) and pLVX-TetOne-PURO(cMYC-EGFP-DCD23)
12	Fluc-EGFP(P2A)DCD23	pGL4.10[luc2] (Promega) and pLVX-TetOne-PURO(cMYC-EGFP(P2A)DCD23)
13	EGFP	pLVX-TetOne-PURO(EZH2-EGFP-DCD23)
14	DCD23-EGFP	pLVX-TetOne-PURO(DCD23-EGFP-KRAS[G12V])
15	DCD23(P2A)EGFP	pLVX-TetOne-PURO(DCD23(P2A)EGFP-KRAS[G12V])
16	EGFP-DCD23	pLVX-TetOne-PURO(EZH2-EGFP-DCD23)
17	EGFP(P2A)DCD23	pLVX-TetOne-PURO(EZH2-EGFP(P2A)DCD23)
18	DCD23mut1-EGFP	pLVX-TetOne-PURO(EGFP-DCD23mut1) and pLVX-TetOne-PURO(DCD23-EGFP-KRAS[G12V])
19	DCD23mut1(P2A)EGFP	pLVX-TetOne-PURO(EGFP-DCD23mut1) and pLVX-TetOne-PURO(DCD23(P2A)EGFP-KRAS[G12V])
20	EGFP-DCD23mut1	pMK-DCD23mut1 (GeneArt, Invitrogen) and pLVX-TetOne-PURO(EZH2-EGFP-DCD23)
21	EGFP(P2A)DCD23mut1	pLVX-TetOne-PURO(EGFP-DCD23mut1) and pLVX-TetOne-PURO(EZH2-EGFP(P2A)DCD23)
22	DCD23mut2-EGFP	pLVX-TetOne-PURO(EGFP-DCD23mut2) and pLVX-TetOne-PURO(DCD23-EGFP-KRAS[G12V])
23	DCD23mut2(P2A)EGFP	pLVX-TetOne-PURO(EGFP-DCD23mut2) and pLVX-TetOne-PURO(DCD23(P2A)EGFP-KRAS[G12V])
24	EGFP-DCD23mut2	pMK-DCD23mut2 (GeneArt, Invitrogen) and pLVX-TetOne-PURO(EZH2-EGFP-DCD23)
25	EGFP(P2A)DCD23mut2	pLVX-TetOne-PURO(EGFP-DCD23mut2) and pLVX-TetOne-PURO(EZH2-EGFP(P2A)DCD23)
26	DCD23mut3-EGFP	pLVX-TetOne-PURO(EGFP-DCD23mut3) and pLVX-TetOne-PURO(DCD23-EGFP-KRAS[G12V])
27	DCD23mut3(P2A)EGFP	pLVX-TetOne-PURO(EGFP-DCD23mut3) and pLVX-TetOne-PURO(DCD23(P2A)EGFP-KRAS[G12V])
28	EGFP-DCD23mut3	pMK-DCD23mut3 (GeneArt, Invitrogen) and pLVX-TetOne-PURO(EZH2-EGFP-DCD23)
29	EGFP(P2A)DCD23mut3	pLVX-TetOne-PURO(EGFP-DCD23mut3) and pLVX-TetOne-PURO(EZH2-EGFP(P2A)DCD23)

Table 2. 4 Donor vectors used for PCR-amplification of target proteins' cDNA.

**A**

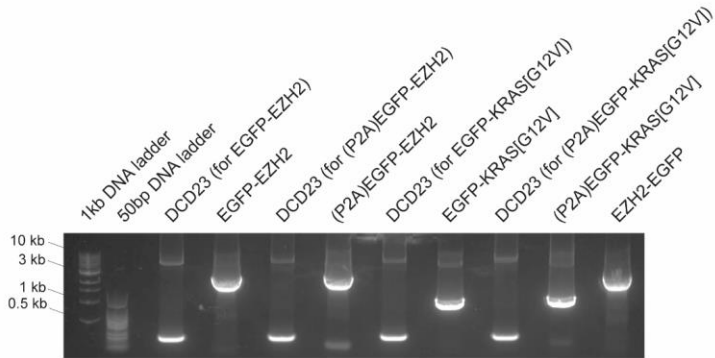
Plasmids Generated
pLVX-DCD23-EGFP-EZH2
pLVX-DCD23(P2A)EGFP-EZH2
pLVX-DCD23-EGFP-KRAS[G12V]
pLVX-DCD23(P2A)EGFP-KRAS[G12V]
pLVX-EZH2-EGFP-DCD23
pLVX-EZH2-EGFP(P2A)DCD23
pLVX-KRAS[G12V]-EGFP-DCD23
pLVX-KRAS[G12V]-EGFP(P2A)DCD23

**B**



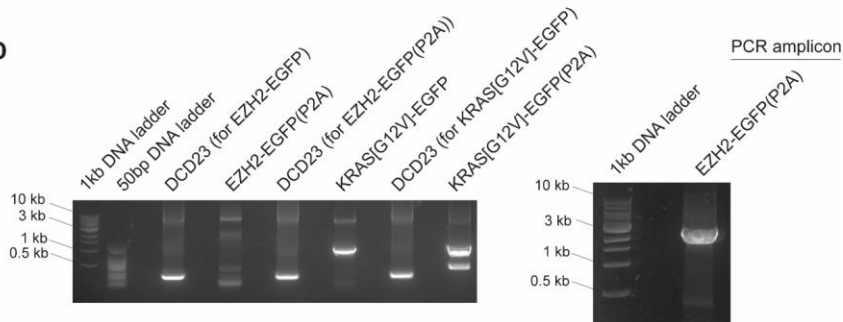
PCR amplicons

**C**



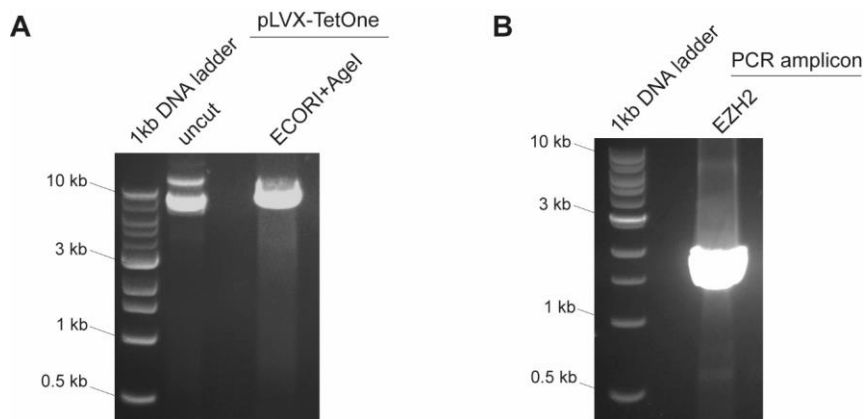
PCR amplicons

**D**



PCR amplicon

**Figure 2. 5 Generation of lentiviral backbone vectors encoding DCD23-tagged KRAS[G12V] and EZH2.** (A) Table summarising the constructs generated. (B) Agarose gel electrophoresis image of linearised pLVX-TetOne lentiviral backbone vector with ECORI and AgeI restriction digest enzymes. (C-D) Agarose gel electrophoresis images of PCR-amplicons cloned in linearised pLVX-TetOne lentiviral backbone vector to generate the final fusion constructs.

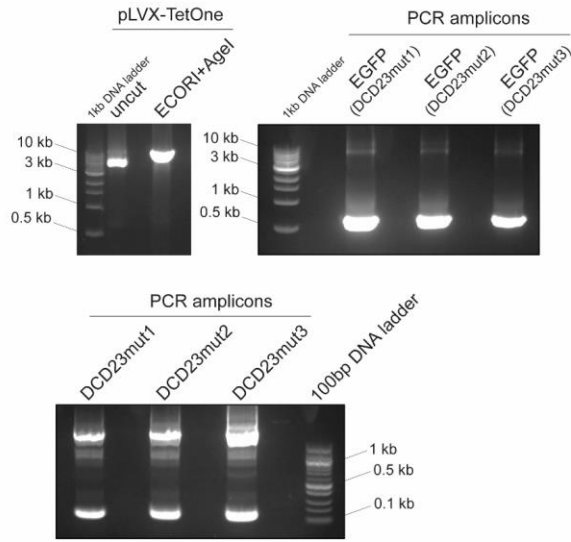


**Figure 2. 6 Generation of lentiviral backbone vector encoding untagged EZH2.** (A) Agarose gel electrophoresis images of linearised pLVX-TetOne lentiviral backbone vector with ECORI and AgeI restriction digest enzymes and (B) PCR-amplicon cloned in linearised pLVX-TetOne lentiviral backbone vector to generate the final fusion construct.

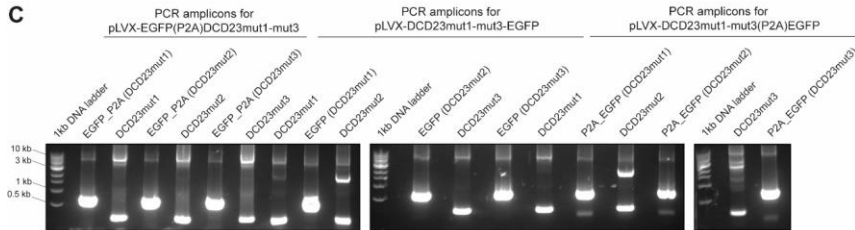
**A**

Plasmids Generated
pLVX-EGFP
pLVX-DCD23WT-EGFP
pLVX-DCD23WT(P2A)EGFP
pLVX-EGFP-DCD23WT
pLVX-EGFP(P2A)DCD23WT
pLVX-DCD23mut1-EGFP
pLVX-DCD23mut1(P2A)EGFP
pLVX-EGFP-DCD23mut1
pLVX-EGFP(P2A)DCD23mut1
pLVX-DCD23mut2-EGFP
pLVX-DCD23mut2(P2A)EGFP
pLVX-EGFP-DCD23mut2
pLVX-EGFP(P2A)DCD23mut2
pLVX-DCD23mut3-EGFP
pLVX-DCD23mut3(P2A)EGFP
pLVX-EGFP-DCD23mut3
pLVX-EGFP(P2A)DCD23mut3

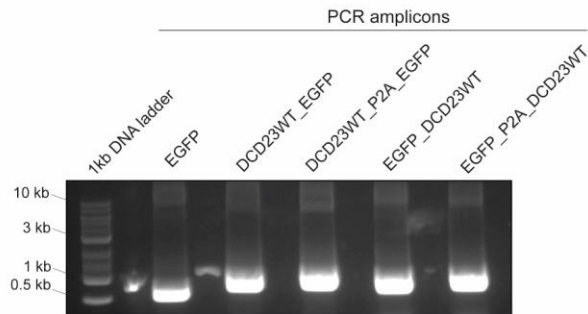
**B**



**C**



**D**





**Figure 2. 7 Generation of lentiviral backbone vectors encoding EGFP tagged at either terminus with DCD23 and DCD23mut1-mut3.** (A) Table summarising the constructs generated. (B-D) Agarose gel electrophoresis images of linearised pLVX-TetOne lentiviral backbone vector with ECOR1 and AgeI restriction digest enzymes and PCR-amplicons cloned in linearised pLVX-TetOne lentiviral backbone vector to generate the final fusion constructs.

### ***2.3 Transient cell transfection***

Cells were grown to a confluence of 70-80 % in 6-well plates (Falcon). For each well, 1-2.5 µg of DNA and 5 µL P3000 reagent (Invitrogen) were diluted in 125 µl of Opti-MeM™ I reduced serum media (Gibco/Life Technologies) and then mixed with 3.75 µl Lipofectamine™ 3000 Transfection Reagent (Invitrogen) diluted in 125 µl Opti-MeM™ bringing the final volume to 250 µl. This mix was incubated for 10-15min at RT (room temperature) and added dropwise to the well containing 2 ml of growth medium. Cells were grown for additional 24-48h and then harvested for flow cytometry or lysed for immunoblotting.

## ***2.4 Virus production and cell line transduction***

For viral particle production,  $9 \times 10^6$  HEK 293T cells were plated in each 100 mm  $\times$  20 mm Corning® tissue-culture treated culture dishes (Corning) 24h prior to transfection. Cell seeding density was previously optimised by testing different numbers of HEK 293T cells. For each dish, 10  $\mu$ g pLVX-TetOne-PURO lentiviral vector expressing target protein fused to iTAG degron, 4.4  $\mu$ g psPAX2 (packaging plasmid), 1.8  $\mu$ g pMD.2G (enveloping plasmid) and 32.4  $\mu$ L P3000 reagent (2x plasmid DNA amount) were diluted in 1.5 ml of Opti-MeM™ and then mixed with 41  $\mu$ l Lipofectamine™ 3000 Transfection Reagent diluted in 1.5 ml Opti-MeM™ bringing the final volume to 3 ml. This mix was incubated for 10-15min at RT and added dropwise to the dish containing 9 ml of growth medium. Growth medium was replaced 6h post-transfection and virus-containing supernatants were harvested 48h post-transfection and filtered through a 0.45  $\mu$ m filter (Sartorius Stedim) to be stored in cryovials (ThermoFischer Scientific) at -80°C. Optimal time-point for cell harvest was deduced by collecting cells at 24h, 48h and 72h and counting the number of viral particles produced. Viral particle titre was determined using the Lenti-X™ GoStix™ kit (Clontech), according to manufacturer's instructions. For cell transduction,  $1.5 \times 10^5$  HMEC/ $1.5 \times 10^5$  MDA-MB-231 cells were plated per well in a 6-well plate. Cell seeding density was previously optimized by plating and transducing different numbers of HMEC cells. Virus-containing supernatants were thawed on ice for 2h and to transduce the cells a multiplicity of infection (MOI) of 3 was selected as previously optimised by testing different MOIs to transduce HMEC cells. The final virus-

containing supernatants were supplemented with polybrene at a final concentration of 4-8  $\mu\text{g/ml}$ . Plates were centrifuged at 600-900 g, 32°C for 30-90min and cultured at 37°C in humidified atmosphere with 5% CO<sub>2</sub> for 24h. Cells were cultured for at least two passages before inducing construct expression with 1  $\mu\text{g/ml}$  of DOX (doxycycline) for 24h. Epifluorescence microscopy was used to determine infection efficiency which was estimated to be between 50-70%. Fluorescence-activated cell sorting (FACS) was used to select for the top 30% of the cell population with the highest EGFP expression. Alternatively, cell lines were treated with puromycin (Sigma Aldrich) for 3-5 days in order to select for cells having incorporated in their genome the lenti-vectors. Concentration of puromycin used for selection was previously experimentally determined by assessing cell viability at various timepoints and by using different puromycin concentrations. Expression of fused proteins was additionally confirmed by immunoblotting post-FACS or puromycin selection.

## ***2.5 Drug storage and treatment***

Thalidomide, lenalidomide, pomalidomide, CC-885, CC-220, CC-122 and CC-92480 were synthesised by ICR CTU chemistry department. Stocks of these drugs were generated by dilutions in DMSO to a final concentration of 10 mg/ml, which were stored at -20°C. DOX was diluted in sterile water to a concentration of 2 mg/ml and stored in -20°C. Prior to drug treatment, cells were plated in either 24-well, or 12-well or 6-well plates (Falcon) or in T25 flasks (Corning) at appropriate seeding densities. Expression of target proteins was stimulated following addition of 1 µg/ml DOX 24h post-seeding. Treatment with either an IMiD or a CELMoD was performed 24h post-DOX addition at indicated concentrations in each experiment. All DMSO treatments of control groups were also performed 24h post-DOX addition at the highest concentration of IMiD/CELMoD used in that particular assay. Cells were incubated for the indicated amount of time before being harvested for flow cytometry analysis or lysed for immunoblotting.

## **2.6 Cell Lysis**

Cells were grown in either 12-well plates (Falcon) or 6-well plates (Falcon) or T25 flasks (Corning) under different treatment conditions. Cells were washed once with PBS (Phosphate-buffered saline) solution, then lysed with SDS (Sodium Dodecyl Sulfate) lysis buffer (2% SDS (Sigma Aldrich), 10% Glycerol, 50mM Tris pH 6.8) and boiled at 96°C for 30min. The Pierce™ bicinchoninic acid (BCA) Protein Assay kit (ThermoFischer Scientific) was used to quantify protein concentration according to manufacturer's instructions and albumin (2 mg/ml; Invitrogen) was used as a standard. Absorbance reading at 562 nm was performed using a SpectraMax® M5 Multi-Mode Microplate Reader (Molecular Devices). Protein concentrations for each sample was determined using a 6-point standard curve generated from titrating albumin. Following protein quantification, SB denaturing buffer [62.5 mM Tris-HCl pH 7.5, 2% SDS, 10% Glycerol, 0.04% Bromophenol blue and 1 mM Dithiothreitol (DTT; Sigma-Aldrich)] was added to the samples which were subsequently heated at 96°C for 10-20min to enable complete protein denaturation. Samples were then placed at -20°C or -80°C for either short- or long-term storage.

## ***2.7 Electrophoresis and Western Blotting***

NuPAGE™ MES SDS running buffer (ThermoFischer Scientific) and NuPAGE™ 10% Bis-Tris Midi gels (Invitrogen) were used for electrophoresis. For each sample, 20-100 µg of proteins were loaded on the gel and run at 150 V for approximately 60-90min. Proteins analysed in NuPAGE™ gels were subsequently transferred to an Immobilon-P PVDF (Polyvinylidene Difluoride) membrane (Merck Millipore) (previously soaked in methanol) at 1.0 A for 2h using the wet-transfer method (Bio-Rad). The transfer buffer consisted of 20% (v/v) methanol and 10 % protein electrophoresis buffer (24.77 mM Tris and 0.192M glycine). Ponceau S solution (Sigma-Aldrich) was used to stain proteins after transfer and subsequently, the PVDF membranes were blocked in a solution of 5% (w/v) bovine serum albumin (BSA; Sigma-Aldrich) dissolved in Tris- buffered saline containing 0.1% Tween20 (Sigma-Aldrich) [Tween 20 (TBS-T)] for 30min on a shaker at RT. The primary antibodies were diluted in 5% (w/v) BSA – 0.1% TBS-T (**Table 2. 5**). Membranes were probed with the primary antibodies at 4°C overnight on a shaker. Anti-mouse and anti-rabbit secondary antibodies (LI-COR Biosciences) were diluted 1:5000 in 5% (w/v) BSA/ 0.1% TBS-T. Three 10min washes were performed between primary and secondary antibodies incubation. Incubation with the secondary antibodies was performed at RT for 2h on a shaker, followed by an additional three 10min washes prior to membrane imaging. The signal of the PVDF membranes was imaged using the LI-COR Odyssey (LI-COR Biosciences) and the images were analysed using Image Studio Lite software (LI-COR Biosciences).

Antibody	Supplier	Catalog Number	Species	Working Dilution
Recombinant Anti-c-Myc antibody [Y69]	Abcam	ab32072	Rabbit	1 in 1000
Living Colors® Full-Length GFP Polyclonal Ab	Takara/Clontech	632592	Rabbit	1 in 1000
Anti-CRBN	Sigma Aldrich	HPA045910	Rabbit	1 in 1000
β-Actin (13E5) mAb	CST	4970	Rabbit	1 in 2000
Anti-beta Actin	Abcam	ab8227	Mouse	1 in 2000
Vinculin Monoclonal Antibody (VLN01)	ThermoFischer	MA5-11690	Mouse	1 in 2000
Ras (G12V Mutant Specific) (D2H12) mAb	CST	14412	Rabbit	1 in 250/ 1 in 500
Purified Mouse Anti-EZH2 Clone 11/EZH2	BD	612667	Mouse	1 in 1000
Histone H3 [Trimethyl Lys27] Antibody	Active Motif	61017	Mouse	1 in 1000
Histone H3 antibody	Active Motif	39763	Mouse	1 in 1000
Anti-Histone H3 antibody	Abcam	ab1791	Rabbit	1 in 1000
MEK1 (D2R1O) mAb	CST	12671	Rabbit	1 in 1000
Anti-Phospho MEK-1 (T292) mAb	Abcam	ab76314	Rabbit	1 in 1000
GFP Polyclonal Antibody	Invitrogen	A-11122	Rabbit	1 in 1000

**Table 2. 5 Primary antibodies used for immunoblotting.**

## 2.8 Flow Cytometry

Cells stably were seeded in either 24-well plates, or 12-well plates or 6-well plates, or T25 flasks overnight. The next day the cells were treated with 1 µg/ml DOX for 24h to induce expression of the target proteins. Subsequently the cells were treated with IMiDs/CELMoDs at concentrations and time-points indicated for each experimental setting. Cells were collected in sterile 5 ml polystyrene round-bottom tubes. DAPI (4',6-diamidino-2-phenylindole; ThermoFisher Scientific) was added to the cell suspension at a ratio of 1: 1000, as a marker for cell viability and therefore exclusion of dead cells. Cells were kept

on ice until the analysis. The EGFP signal was measured on the BD LSR II flow cytometer (BD Biosciences). Further processing of the results was performed using FlowJo 10 Software (FlowJo).

## **2.9 Cell Viability Assay**

To assess cell viability, the Cell Titer Blue® viability assay (Promega) was performed according to manufacturer's instructions. Briefly, HMEC, HEK 293T, Kelly and MDA-MB-231 cells were plated in 96-well plates in triplicate and treated with CC-885, thalidomide, lenalidomide, pomalidomide, CC-220 or CC-122 for five consecutive days. At the end of the assay, 30 µL of the CellTiter-Blue® reagent (Promega) was directly added to each well and cells were incubated at 37°C for 4h. The absorbance was measured with an a EnVision 2105 Multimode Plate Reader (PerkinElmer) at 570 nm. As control, average background absorbance value was subtracted from all readings when plotting the viability curve.



## **2.10 Recombinant protein production and purification**

### **2.10.1 CRBN/ DDB1**

The recombinant protein production for CRBN/DDDB1 was carried out by Dr Craig McAndrews (Department of Cancer Therapeutics, Institute of Cancer research). Full length human 6His-ZZ-CRBN and DDB1-Strep proteins were co-expressed in sf9 insect cells and purified as previously described (Chessum *et al.*, 2018). The protein complex was stored at -80°C in a buffer containing 50 mM HEPES pH 8.0, 200 mM NaCl, 0.5 mM TCEP [tris(2-carboxyethyl)phosphine] and 5% glycerol.

### **2.10.2 DCD1**

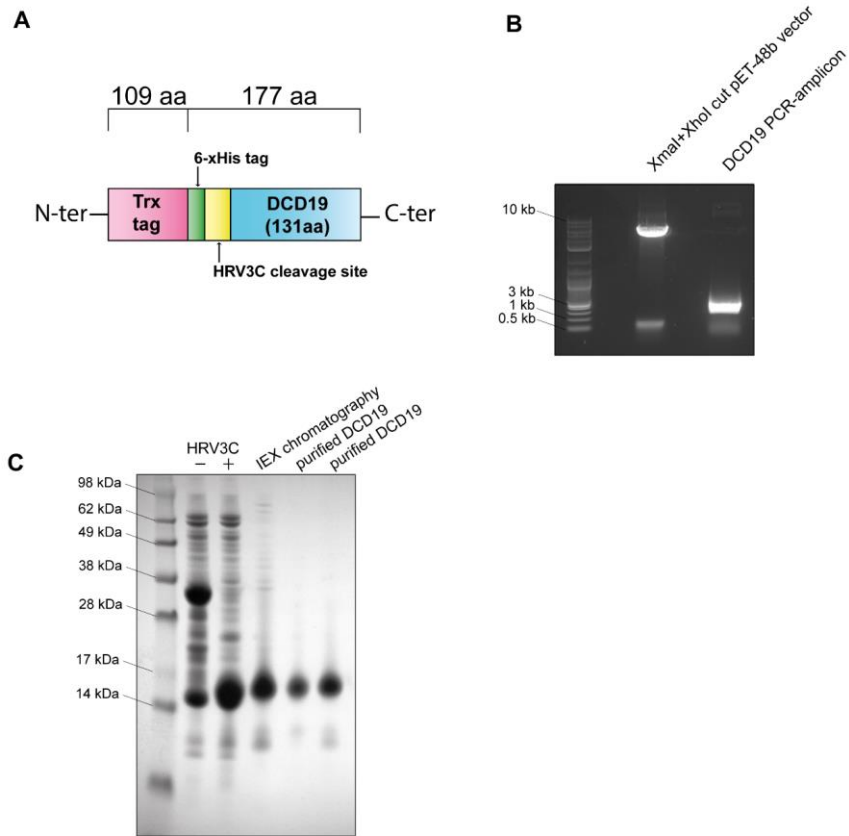
The recombinant protein production for DCD1 was carried out by Dr Craig McAndrews (Department of Cancer Therapeutics, Institute of Cancer research). The sequence coding for DCD1, consisting of residues 438 to 634 of GSPT1 (UniProtKB - P15170-3), was codon optimized for production in *E. coli* and synthesized by Eurofins. It was subcloned into pET-48b(+) plasmid between

XmaI and XhoI restriction sites to generate a plasmid coding for a Trx-6His-HRV3C-DCD1 protein. Transformed BL21-AI *E. coli* cells were grown in TB (Terrific Broth) media supplemented with 50 mg/L kanamycin at 37°C until an OD<sub>600 nm</sub> of 0.6 was reached. Protein expression was then induced by addition of 0.2 mM IPTG (isopropylthio-β-galactoside) and 0.2% Arabinose. Expression was carried out at 18°C for 18h. Cells were harvested by centrifugation at 5500 g for 30min at 4°C and stored at -80°C. Cells were re-suspended in a buffer composed of 50 mM HEPES pH 7.4, 500 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.5 mM TCEP, 1x cOmplete ULTRA protease inhibitors, 1mg/mL Lysozyme, and 12.5 U/ml Benzonaze. Cells were lysed by sonication followed by centrifugation at 56,000 g for 20min at 4°C. The supernatant was loaded onto a 5 ml Hitrap TALON crude column and eluted with 150 mM Imidazole. The Trx-6His-HRV3C tag was cleaved with HRV-3C protease, and DCD1 further purified with Superdex75 26/60 followed by reverse-TALON steps. Finally, the resulting DCD1 protein was ran through a PD10 column to exchange buffer to 10 mM HEPES pH 7.4, 240 mM NaCl, 3 mM TCEP, and stored at -80°C.

### **2.10.3 DCD19**

I carried out the recombinant protein production for DCD19. The sequence coding for DCD19, consisting of residues 83 to 197 and 239 to 255 of Ikaros (UniProtKB - Q13422), was codon optimized for production in *E. coli* and synthesized by GeneArt. It was subcloned into pET-48b(+) plasmid between

XmaI and XhoI restriction sites to generate a plasmid coding for a Trx-6His-HRV3C-DCD19 protein (**Figure 2. 8 A-B**). Transformed BL21-AI *E. coli* cells were grown in TB media supplemented with 50 mg/L kanamycin at 37°C until an OD<sub>600 nm</sub> of 0.6 was reached. Protein expression was then induced by addition of 0.25 mM IPTG and 0.2% Arabinose, and culture was supplemented with 50 µM ZnCl<sub>2</sub>. Expression was carried out at 18°C for 18h. Cells were harvested by centrifugation at 5500 g for 30min at 4°C and stored at -80°C. Cells were re-suspended in a buffer composed of 50 mM HEPES pH 7.4, 200 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.25 mM TCEP, 1x cOmplete ULTRA protease inhibitors, 1mg/mL Lysozyme, and 12.5 U/ml Benzonaze. Cells were lysed by sonication followed by centrifugation at 56,000 g for 20min at 4°C. The supernatant was loaded onto a HisTrap FF column, washed at 1M NaCl to remove DNA contaminants, and eluted with 160 mM Imidazole. The Trx-6His-HRV3C tag was cleaved with HRV-3C protease, and DCD19 further purified with MonoS HR5/5 and Superdex75 16/60 steps (**Figure 2. 8 C**). Finally, the resulting DCD19 protein was stored at -80°C in a buffer containing 20 mM HEPES pH 7.4, 200 mM NaCl and 3 mM TCEP.



**Figure 2. 8 In vitro production and purification of DCD19.** (A) DCD19 was cloned in a pET-48b vector as a recombinant protein fused at the N-terminus to a thioredoxin (Trx) tag, a 6xHis tag and an HRV3C cleavage site. (B) Agarose gel depicting pET-48b vector linearised by the restriction digest enzymes XmaI and PCR-amplified DCD19 to be cloned in pET-48b linearised vector. (C) SDS-PAGE gel of DCD19 lysate at various steps of the purification procedure.

#### 2.10.4 DCD23

The recombinant protein production for DCD23 was carried out by Dr Craig McAndrews (Department of Cancer Therapeutics, Institute of Cancer research). The sequence coding for a slightly shorter version of DCD23 than used in cells, consisting of residues 141 to 144 of Ikaros (UniProtKB - Q13422) followed by residues 400-410 of ZFP91 (UniProtKB - Q96JP5) and residues 156-188 of IKAROS, was codon optimized for production in *E. coli* and synthesized by Eurofins. It was subcloned into pET-48b(+) plasmid between XmaI and XhoI restriction sites to generate a plasmid coding for a Trx-6His-HRV3C-DCD23 protein. Transformed BL21-AI *E. coli* cells were grown in TB media supplemented with 50 mg/L kanamycin at 37°C until an OD<sub>600 nm</sub> of 0.6 was reached. Protein expression was then induced by addition of 0.2 mM IPTG and 0.2% Arabinose. Expression was carried out at 18°C for 18h. Cells were harvested by centrifugation at 5500 g for 30min at 4°C and stored at -80°C. Cells were re-suspended in a buffer composed of 50 mM HEPES pH 7.4, 200 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.25 mM TCEP, 1x cOmplete ULTRA protease inhibitors, 1mg/mL Lysozyme, and 12.5 U/ml Benzonaze. Cells were lysed by sonication followed by centrifugation at 56,000 g for 20min at 4°C. The supernatant was loaded onto a HisTrap FF column and eluted with 250 mM Imidazole. The Trx-6His-HRV3C tag was cleaved with HRV-3C protease, and DCD23 further purified with a Superdex75 16/60 followed by reverse-HisTrap FF and MonoS HR5/5 steps. Finally, the resulting DCD23 was ran through a PD10 column to exchange buffer to 50 mM Tris pH 8, 200 mM NaCl, 0.25 mM TCEP, and stored at -80°C.

## **2.11 IMiD-based Fluorescence Polarisation assay**

The IMiD-based FP (Fluorescence Polarisation) assays were carried out by Dr Mark Stubbs (Department of Cancer Therapeutics, Institute of Cancer research). All FP assays were run in Black 384-well ProxiPlate Plus (PerkinElmer), in a buffer containing 20 mM HEPES pH 8.0, 150 mM NaCl, 0.5 mM TCEP and 0.05% Tween 20, with a final assay volume of 10  $\mu$ L. An Echo E5XX (Beckman Coulter) acoustic liquid dispenser was used to create final concentration ranges of compound from 300  $\mu$ M to 0.9375 nM, with a final DMSO concentration of 3%. WT full length CRBN/DDB1 complex was added with a Tempest liquid handler (Formulatrix) to all wells to a final concentration of 80 nM, except for low controls where WT CRBN was substituted with mutant CRBN<sup>Y384A/W386A</sup> (mutant unable to bind the probe) (Fischer *et al.*, 2014). The Echo was also used to add 5 nM final concentration of fluorescent IMiD-based probe to each well. Plates were sealed, centrifuged at 200 g for 1min and stored overnight at 4°C. Fluorescence Polarization and total fluorescence were read using a PHERAstar FSX plate reader (BMG Labtech).

## 2.12 Aiolos peptide-based TR-FRET assay

The Aiolos peptide-based TR-FRET assay was carried out in collaboration with Dr Mark Stubbs (Department of Cancer Therapeutics, Institute of Cancer research). All TR-FRET assays were ran in Black 384-well ProxiPlate Plus (PerkinElmer), in a buffer containing 20 mM HEPES pH 8.0, 150 mM NaCl, 0.5 mM TCEP and 0.05% Tween 20, with a final assay volume of 10  $\mu$ L. An Echo E5XX (Beckman Coulter) acoustic liquid dispenser was used to create final concentration ranges from 90  $\mu$ M down to 2.25 nM for DCD1 and DCD23, and 30  $\mu$ M to 0.75 nM for DCD19. Final concentrations to 5 nM WT full length CRBN/DDB1 complex, 750 nM fluorescent Aiolos-based peptidic probe (Cambridge Research Biochemicals) and 0.5 nM mAb Anti-6HIS-Terbium cryptate Gold (Cisbio) were added with a Tempest liquid handler (Formulatrix). Finally, IMiDs/CELMoDs were added to a final concentration of 10  $\mu$ M (or 1  $\mu$ M for CC-885 due to solubility issues) with the Echo. Plates were sealed, centrifuged at 200 g for 1min and stored overnight at 4°C. The TR-FRET signals were read using a PHERAstar FSX plate reader (BMG Labtech). Final signal was measured as the ratio:

$$\text{TR - FRET signal} = \frac{\text{chanel1 665nm}}{\text{chanel2 620nm}}$$

Chanel1 665 nm representing a positive FRET energy transfer and chanel2 620 nm representing the emission of the terbium tag.

### ***2.13 In vivo degradation assay***

MDA-MB-231 cells ( $5 \times 10^6$  cells) engineered to express DCD23-EGFP-Fluc or DCD23(P2A)EGFP-Fluc were transplanted subcutaneously into the right flank of 4–5-week-old Athymic Nude-Foxn1<sup>nu</sup> (Envigo) mice. Subcutaneous tumour volume was measured using callipers. Animals were administered sucrose water with DOX (2 mg/mL) for 3-5 days when tumours reached 200-350 mm<sup>3</sup>. After DOX administration, animals were randomized for vehicle or CC-220 [30 mg/kg in 0.5% carboxymethyl cellulose (Na salt) with 0.25% Tween80 in water] treatment. Tumour luminescence was quantified using non-invasive bioluminescence imaging, with luminescence measured 15 seconds post retro-orbital luciferin injection (150 mg/kg) using an IVIS imager. Tumour luminescence was measured before and at the indicated timepoint after vehicle or CC-220 administration. Tumours were then harvested and flash frozen for further analysis. All in vivo experiments were performed by members of Prof Trey Westbrook's laboratory at the Baylor College of Medicine, Houston, Texas, USA.



## **2.14 Immunofluorescence**

HMEC cells were seeded in a 96-well plate at a 1000 cells/well in a final volume of 100  $\mu$ l growth media and incubated for 24h at 37°C. The next day the cells were treated with DOX at indicated concentration and incubated for a further 24h. Growth media was removed and the cells were washed three times with PBS solution, then fixed by incubation with 4% paraformaldehyde in PBS solution for 15min at 37°C and subsequently washed three times with ice-cold PBS solution. Permeabilization of the cells was performed by incubating the cells with 0.05% Triton-X-100 detergent (Sigma Aldrich) for 10min at RT and then washing the cells three times with PBS solution, 5min each wash. For immunostaining, fixed and permeabilized cells were incubated with 1% BSA, 22.52 mg/ml glycine in PBS solution containing 0.1% Tween 20 for 30min to block unspecific binding of the antibody. To stain the cell cytoplasm, the MS  $\beta$ -actin antibody (Abcam, ab8227) was diluted at 5  $\mu$ g/ml in 1% BSA in PBS solution containing 0.1% Tween 20 and was added to the cells for 1h on a plate rocker. The solution was then decanted and the cells were washed three times with PBS solution containing 0.1% Tween 20, for 5min each wash. The secondary antibody Alexa fluor 546 MS (Invitrogen, A-11030) was diluted at 4  $\mu$ g/ml in 1% BSA in PBS solution containing 0.1% Tween 20 and was added to the cells for 1h at RT on a plate rocker in the dark. Then, the solution was decanted and the cells were washed three times with PBS solution containing 0.1% Tween 20, for 5min each wash. Prior to proceeding to cell imaging, the cells were incubated for 1min with

1 µg/ml DAPI (ThermoFisher Scientific) to stain the cell nuclei and rinsed with PBS solution. Cells were visualised on the Zeiss Axio Vert.A1 FL-LED inverted epifluorescence microscope (Zeiss) which has a four-position mount for LED modules for reflected light fluorescence and is connected to a HXP 12V metal halide fluorescence light source (Zeiss). The objective LD A-Plan 20x/0.35 Ph1 M27 (Zeiss) was used to image the cells at a magnification of 20x. The LED module at a wavelength of 470 nm was used to excite the EGFP protein (emission at 509 nm), the LED module at a wavelength of 530 nm was used to excite the Alexa fluor 546 conjugated to the β-actin antibody (emission at 573 nm) and the LED module at a wavelength of 365 nm was used to excite DAPI (emission at 461 nm). Images were captured using the ZEN Blue software (Zeiss) and image processing was performed using Fiji (Fiji Is Just ImageJ) open-source image processing software.

## ***2.15 Statistical analysis***

Statistical analysis was performed using the GraphPad Prism software. To calculate the statistical significance between the nuclear/cytoplasmic fluorescence signal ratio of untagged EGFP and EGFP fused to either WT DCD23/ DCD23mut1/ DCD23mut2/ or DCD23mut3, unpaired Student's t-tests with a two-tailed distribution were performed. Statistically not significant results are denoted with ns ( $\geq 0.05$ ). Statistically significant results are denoted with an

asterisk (\*  $p = 0.01 - 0.05$ , \*\*  $p = 0.001 - 0.01$ , \*\*\*  $p = 0.0001-0.001$ , \*\*\*\*  $p < 0.0001$ ) (Table 2. 6).

Symbol	Meaning
ns	$P > 0.05$
*	$P \leq 0.05$
**	$P \leq 0.01$
***	$P \leq 0.001$
****	$P \leq 0.0001$

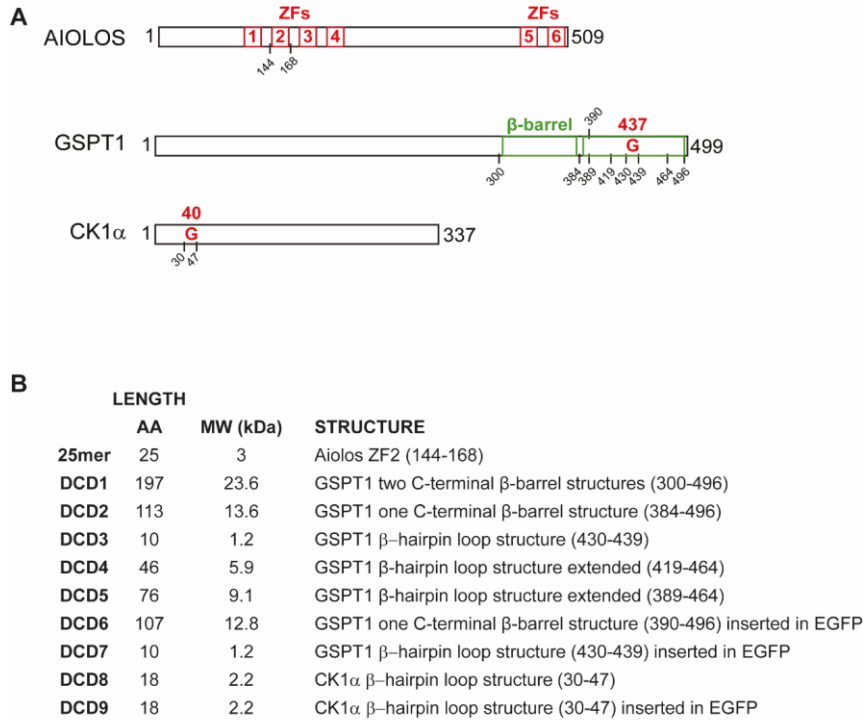
**Table 2. 6 Symbol meaning.**

## **3. RESULTS**

***Subchapter 3.1 : Identification of novel DCD motifs mediating acute and potent degradation of target proteins***

### 3.1.1 Design and generation of DCD1 to DCD9

In order to define an optimal degron-containing domain (DCD) for target protein degradation with IMiD/CELMoD compounds, initially, structural and sequence analysis of the known CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates, GSPT1 and CK1 $\alpha$ , was undertaken. Both GSPT1 and CK1 $\alpha$  have been shown previously in two independent structural studies to engage the CC-885 or lenalidomide-bound CRL4<sup>CRBN</sup> E3 ubiquitin ligase, respectively, via a  $\beta$ -hairpin loop with a glycine at the apex (Matyskiela *et al.*, 2016; Petzold, Fischer and Thomä, 2016). This  $\beta$ -hairpin loop is located in a C-terminal  $\beta$ -barrel domain in GSPT1 (residues 430-439, isoform 1), whilst, in CK1 $\alpha$  is found in its kinase N-lobe (residues 35-41, isoform 1). The apex glycine residue has been demonstrated by site-directed mutagenesis to be critical for both GSPT1 and CK1 $\alpha$  binding to CRBN, by forming key interactions both with residues on the surface of CRBN and solvent-exposed moieties of the bound CC-885 and lenalidomide, respectively (Matyskiela *et al.*, 2016; Petzold, Fischer and Thomä, 2016). In the GSPT1  $\beta$ -hairpin loop, the carbonyl oxygen atoms of three amino acids after the turn have been shown to form hydrogen bonds with residues on the surface of CRBN (Matyskiela *et al.*, 2016). Furthermore, residues in GSPT1  $\beta$ -hairpin loop have been also shown to form hydrophobic and Van der Waals interactions with the isoindolinone ring of CC-885 (Matyskiela *et al.*, 2016). Similarly, CK1 $\alpha$   $\beta$ -hairpin loop has been demonstrated to interact with the solvent exposed phthalimide ring of lenalidomide and CRBN residues that surround the



**Figure 3.1. 1 Structure and sequence-based design of DCD1 to DCD9 from CUL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates GSPT1 and CK1 $\alpha$ .** (A) Schematic representation of Aiolos, GSPT1 and CK1 $\alpha$  protein sequences. The different ZF domains of Aiolos (ZF1-ZF6) and the apical Glycine (G) of the  $\beta$ -hairpin loop of GSPT1 and CK1 $\alpha$  implicated in the binding to CRBN are highlighted in red. (B) Summary of 25mer Aiolos ZF2-based degron (Koduri *et al.*, 2019) and DCD1 to DCD9 sequences evaluated for this thesis.

lenalidomide binding pocket (Petzold, Fischer and Thomä, 2016). Therefore, this  $\beta$ -hairpin loop in GSPT1 and CK1 $\alpha$  could constitute a conserved minimal degron motif, with its geometric arrangement rather than the peptide sequence, mediating the interactions with both key residues on the surface of CRBN and the bound IMiD/CELMoD (Figure 1. 21 A-B). Based on these previous structural

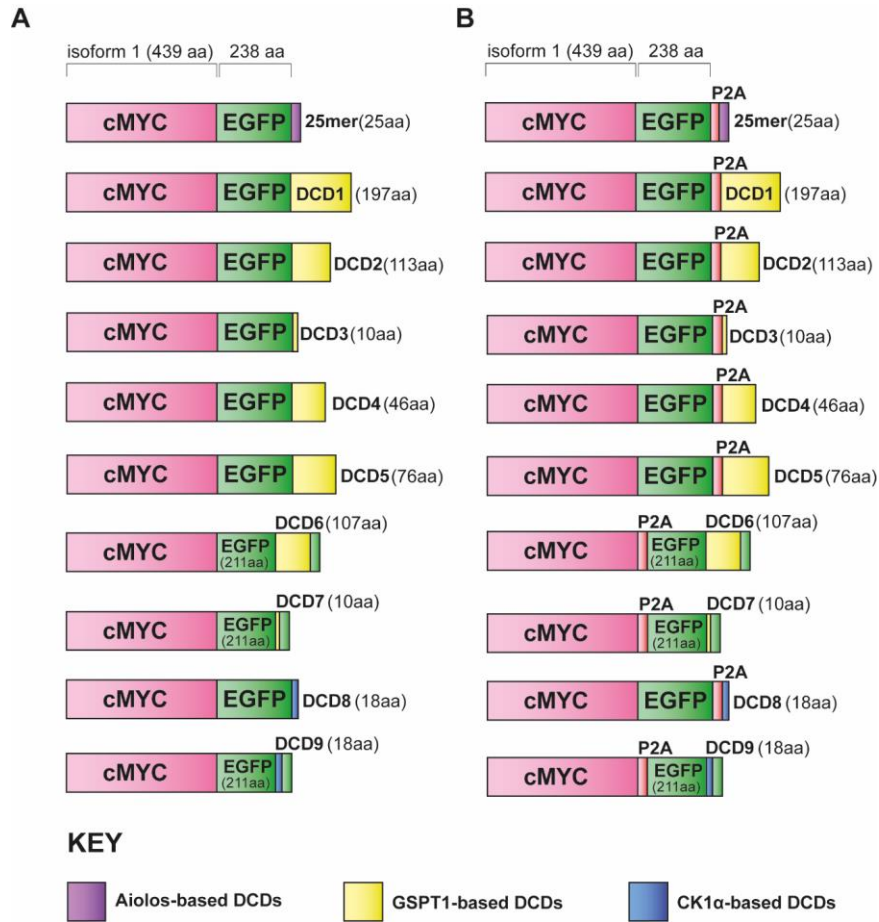
observations, I hypothesised that, this  $\beta$ -hairpin loop from GSPT1 and CK1 $\alpha$  could represent a functional degron when fused to a target protein, with flanking residues both in GSPT1 and CK1 $\alpha$ , potentially being critical for engaging the IMiD/CELMoD-bound CRBN. To test this hypothesis, we designed a set of nine peptide sequences, named DCD1-DCD9, based on either the GSPT1 or CK1 $\alpha$   $\beta$ -hairpin loop alone, or flanked with variable-length amino acid sequences from these neosubstrates (**Figure 3.1. 1 A-B**). DCD3 and DCD7 were designed based the GSPT1 isoform 1  $\beta$ -hairpin loop with Gly437 at the apex, whilst DCD8 and DCD9 were based on the CK1 $\alpha$  isoform 1  $\beta$ -hairpin loop (Gly40 at the apex) with additional amino acids stabilizing the loop (**Figure 3.1. 1 A-B**). Since the exact amino acid sequence length flanking the  $\beta$ -hairpin loop and potentially required for binding to IMiD/CELMoD-CRBN complex was unknown, it had to be empirically defined. Therefore, GSPT1  $\beta$ -hairpin loop-based DCD1, DCD2, DCD4, DCD5 and DCD6 degron motifs were designed to also incorporate variable length amino acid sequences from the GSPT1 protein flanking this  $\beta$ -hairpin loop (**Figure 3.1. 1 A-B**). A concern for the sequence design of these DCDs was that to maintain the correct fold of the peptide, the whole C-terminal  $\beta$ -barrel domain of GSPT1 containing the  $\beta$ -hairpin loop had to be included. Therefore, DCD1, DCD2 and DCD6 were designed to incorporate this, with DCD1 containing an additional GSPT1 C-terminal  $\beta$ -barrel domain. Despite this concern, shorter variants of the GSPT1  $\beta$ -barrel domain containing the  $\beta$ -hairpin loop, namely DCD4 and DCD5, were also designed for empirical testing. In order to track target protein levels in cells, these DCDs were also fused to either the C-terminus of EGFP or replaced a solvent exposed loop (residues 210-216) of EGFP. The rationale behind inserting the DCDs at two different sites in EGFP

was to define an optimal fusion site which would not affect the functionality of the DCDs due to either steric hindrances or incorrect folding. Specifically, DCD1 to DCD5 and DCD8 were fused to the C-terminus of EGFP whilst, DCD6, DCD7 and DCD9 in the solvent exposed loop (residues 210-216) of EGFP (**Figure 3.1. 1 B; Figure 3.1. 2 A**).



### **3.1.2 Fusion of EGFP-DCD sequences to cMYC and stable expression in HMEC cells**

To test the functionality of the DCDs when fused to a target protein, the transcription factor and proto-oncogene cMYC was selected. cMYC constitutes a well-characterised and very important protein in the field of cancer research, making it an ideal target protein for the development of a novel chemical-biology tool (Dang, 2013; Amati and Sabo, 2014; McMahon, 2014; Nabet *et al.*, 2018). The generated EGFP-DCD1 to EGFP-DCD9 sequences were fused at the C-terminus of cMYC (**Figure 3.1. 2 A**). A set of fusion constructs, harbouring the 22 residues-long ribosomal skipping sequence P2A (2.17 kDa) from the porcine teschovirus-1, were also designed as controls for any potential IMiD/CELMoD off-target effects (**Figure 3.1. 2 B**) (Kim *et al.*, 2011). This P2A peptide functions by self-cleaving the peptide bond between its C-terminal proline and glycine residues during ribosome-mediated translation, thus allowing for the translation of two separate proteins, with the upstream protein containing the P2A peptide (21 amino acids) at its C-terminus and the downstream protein having a single extra proline fused at its N-terminus (Kim *et al.*, 2011). For the DCDs that were fused to the C-terminus of EGFP the respective P2A sequences were placed between the EGFP and the DCD. In contrast, for the DCDs replacing the solvent exposed loop of EGFP (residues 210-216) the P2A sequences interspaced the target protein cMYC and EGFP-DCD, since these DCDs were embedded in the sequence of EGFP and could not be otherwise separated from the target protein (**Figure 3.1. 2 A-B**). Furthermore, as a reference point for the degradation efficiency of these DCDs, the 25mer Aiolos ZF2-based degra-

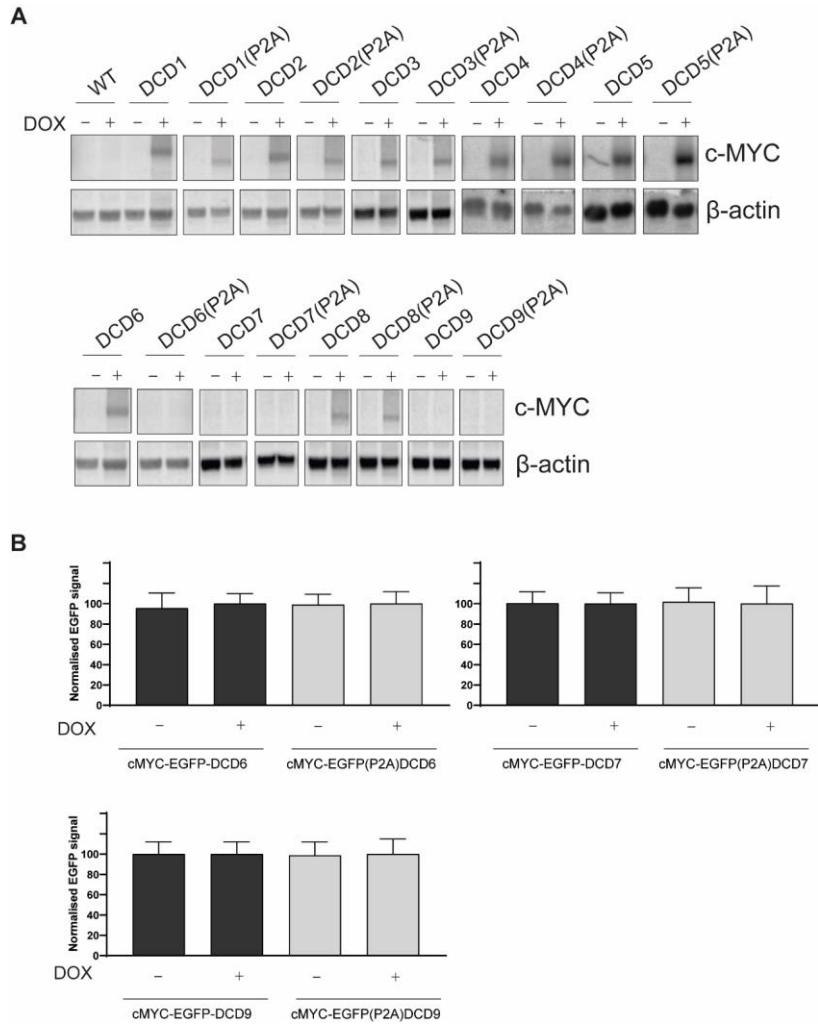


**Figure 3.1. 2 DCD1 to DCD9 fusion to the C-terminus of cMYC-EGFP for functional evaluation of the various DCDs. (A)** Schematic representation of 25mer Aiolos ZF2-based degnon (Koduri *et al.*, 2019) and DCD1 to DCD9 fused along with EGFP to the C-terminus of cMYC. **(B)** Insertion of the P2A sequence between EGFP and DCD or cMYC and EGFP leads to the translation of two separate peptides.

developed by William Kaelin's group was also selected to be fused at the C-terminus of both cMYC-EGFP and cMYC-EGFP(P2A) (**Figure 3.1. 2 A-B**) (Koduri *et al.*, 2019). These fusion constructs were commercially synthesised and cloned into the pLVX-TetOne DOX (doxycycline)-inducible lentiviral plasmid. Subsequently, lentiviral particles were generated with these fusion construct-encoding vectors, which in turn, were used to generate hTERT (human telomerase reverse transcriptase) immortalised HMECs (human mammary epithelial cells) stably expressing the fusion proteins following DOX treatment. HMEC cells were selected on the basis of their relative ease in culturing and their high efficiency of lentiviral transduction, as tested experimentally.

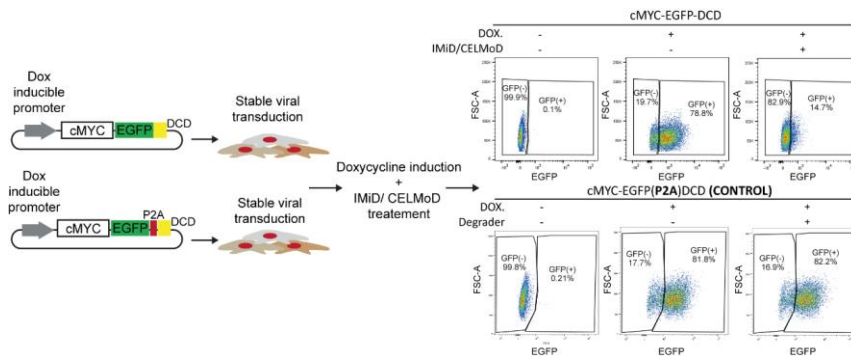
### ***3.1.3 DCD1 and DCD2 induced a potent decrease of cMYC-EGFP levels following CC-885 treatment***

Prior to examining the functionality of DCD1 to DCD9, expression of the fusion constructs was determined. HMEC cells stably transduced with lentiviral vectors encoding cMYC-EGFP-DCD1 to DCD9 and their respective P2A controls were first treated with 1 µg/ml DOX for 24h to induce expression and subsequently lysed and tested by immunoblotting, probing for cMYC protein levels. From the Western blot analysis, expression of cMYC-EGFP fused to DCD1 to DCD6, DCD8 and cMYC-EGFP(P2A) fused to DCD1 to DCD5, DCD8 was verified (**Figure 3.1. 3 A**). On the contrary, cMYC fusions to either EGFP-DCD7, DCD9 or EGFP(P2A)DCD6, DCD7 and DCD9 were not expressed in the HMEC cells (**Figure 3.1. 3 A**). All three non-detected DCDs were designed as fused peptides replacing a solvent exposed loop (residues 210-216) of EGFP, therefore demonstrating that this fusion site cannot be adapted for inserting DCDs as it hinders fusion protein expression. An exception to this was cMYC-EGFP-DCD6, which was confirmed to be expressed by immunoblotting. However, when EGFP fluorescence signal was measured by flow cytometry along with DCD7 and DCD9 cMYC-EGFP fusions, no EGFP signal was detected, suggesting that the fusion of all three DCDs within the EGFP protein sequence abrogates EGFP fluorescence (**Figure 3.1. 3 A-B**). Based on these data, HMEC cells expressing cMYC-EGFP fused to DCD1 to DCD5, DCD8 and their respective P2A controls were further selected for evaluating the potency, efficiency, and selectivity of IMiD/CELMoD-induced cMYC degradation.



**Figure 3.1.3 Examining the expression of cMYC-EGFP fused to DCD1 to DCD9 and their respective P2A controls. (A)** Immunoblots of HMEC cells treated with DOX for 24h to induce expression of the cMYC-EGFP-DCD1 to DCD9 and P2A control constructs. A cMYC antibody was used to probe for the levels of the fusion constructs. Images are representative of 3 independent biological experiments (n=3). **(B)** HMEC cells transduced with cMYC-EGFP-DCD6, DCD7 and DCD9 and cMYC-EGFP(P2A)DCD6, DCD7 and DCD9 were treated with DOX for 24h to stimulate expression of the fusion proteins. Median EGFP intensity was measured in each condition using flow cytometry. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

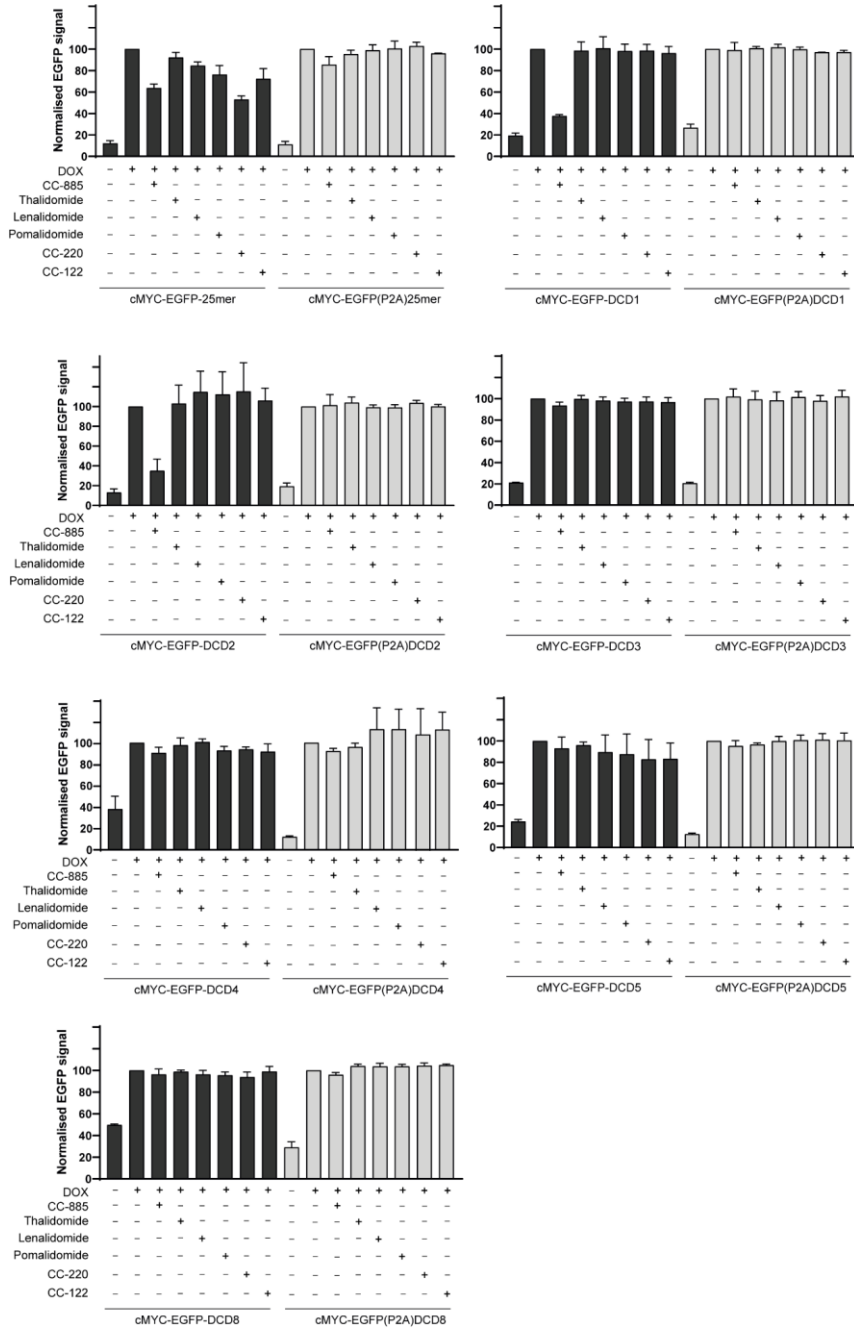
For testing the functionality of the various DCDs, an experimental pipeline was developed (**Figure 3.1. 4**). In this workflow, treatment with 1  $\mu\text{g/ml}$  DOX for 24h was performed to induce expression of fusion constructs in lentiviral-transduced HMEC cells. Then, HMEC cells were treated with either a vehicle (DMSO) or a panel of six IMiDs/CELMoDs, namely, CC-885, thalidomide, lenalidomide, pomalidomide, CC-220 and CC-122 at 10  $\mu\text{M}$  for 4h to capture acute degradation events. Subsequently, cells were harvested and the EGFP fluorescence signal was measured by flow cytometry as a surrogate measurement of cMYC-EGFP and cMYC-EGFP(P2A) protein levels. This workflow analysis aimed at identifying an optimal DCD candidate to develop a novel chemical-biology tool. An optimal DCD would constitute a minimal peptide sequence that would induce an acute, robust, and reversible degradation of cMYC-EGFP following treatment with non-toxic concentrations of IMiDs/CELMoDs. Furthermore, this optimal DCD sequence would also induce a more potent and efficient degradation of the target protein compared to the previously described 25mer Aiolos ZF2-based degron (Koduri *et al.*, 2019). Flow cytometry analysis showed that the control 25mer Aiolos ZF2-based degron led to a decrease of EGFP fluorescence signal with various IMiDs/CELMoDs, with CC-220 inducing the strongest decrease (residual signal: 52.99%) at 4h (**Figure 3.1. 5**). As expected, no significant decrease in EGFP fluorescence levels of the respective cMYC-EGFP(P2A)25mer control was observed. Among the tested IMiDs/CELMoDs and DCDs only CC-885-treated HMEC cells expressing cMYC-EGFP-DCD1 (residual signal: 37.63%) and cMYC-EGFP-DCD2 (35.11%) showed a significant decrease of EGFP fluorescence signal, compared to



**Figure 3.1. 4 Workflow diagram of the experimental pipeline used to evaluate the various DCD constructs.** Cells were stably transduced with plasmids expressing DCDs fused to the C-terminus of cMYC-EGFP or cMYC-EGFP(P2A). EGFP positive cells were FACS-sorted and IMiD/CELMoD-dependent degradation of target protein was assessed by flow cytometry.

vehicle-treated control cells (**Figure 3.1. 5**). Based on these data, the degradation efficiency between the longer GSPT1-based DCD1 (197 residues) and its shorter DCD2 variant (113 residues) was similar. DCD3, DCD4, DCD5 and DCD8 which contained either only the  $\beta$ -hairpin loop or an extended  $\beta$ -hairpin loop structure based on GSPT1 or CK1 $\alpha$  residues flanking this degron motif, failed to induce a decrease in EGFP signal. This suggested that the induction of degradation requires at least the whole C-terminal  $\beta$ -barrel domain of GSPT1 containing this  $\beta$ -hairpin loop, as in DCD2 (**Figure 3.1. 5**). As expected, cells expressing the corresponding cMYC-EGFP(P2A)DCD controls had EGFP levels similar to vehicle-treated control cells for all tested DCDs, due to the uncoupling of the DCD from the cMYC-EGFP (**Figure 3.1. 5**). These results indicated that CC-885-mediated reduction of EGFP fluorescence signal in cMYC-EGFP-DCD1 and cMYC-EGFP-DCD2 transduced HMEC cells is dependent on

## Results





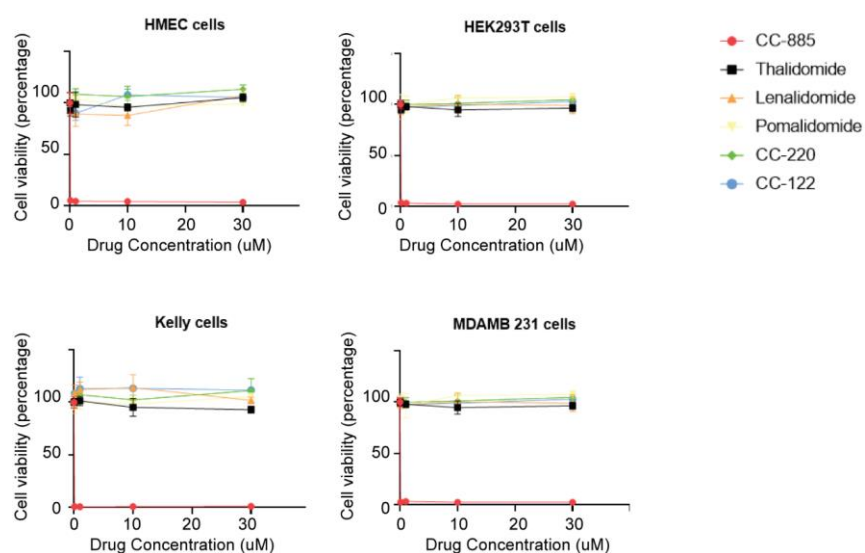
**Figure 3.1. 5 Functional evaluation of DCDs fused to the C-terminus of cMYC-EGFP.** HMEC cells transduced with cMYC-EGFP fused to 25mer Aiolos ZF2-based degron (Koduri *et al.*, 2019), DCD1 to DCD5, DCD8 and their respective P2A controls were treated with 1 µg/ml DOX for 24h to stimulate expression of the fusion constructs, followed by treatment with 10 µM of various IMiDs/CELMoDs for 4h. Median EGFP intensity was measured in each condition using flow cytometry. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

the presence of DCD1 (197 residues) or DCD2 (113 residues) degron motifs. From this cohort of tested DCDs, DCD2 was selected as the most optimal candidate as it induced an acute and potent degradation of cMYC-EGFP and had a shorter amino acid sequence compared to DCD1.

### **3.1.4 CC-885 treatment induces cellular toxicity**

IMiDs/CELMoDs are currently used in the treatment of several haematological malignancies such as multiple myeloma (MM) and certain lymphomas (Singhal *et al.*, 1999; Alan *et al.*, 2005; Fenaux *et al.*, 2011; Richardson *et al.*, 2014; Polizzotto *et al.*, 2021). To evaluate whether IMiDs/CELMoDs exhibit cytotoxicity in non-haematological cells, a panel of cell lines including HMEC (mammary epithelial-derived cells), HEK 293T (embryonic kidney-derived cells), Kelly (neuroblastoma-derived cells) and MDA-MB-231 (breast adenocarcinoma-derived cells) were treated with increasing concentrations of IMiDs/CELMoDs for five consecutive days. Cell viability was analysed after five days and amongst all tested IMiDs/CELMoDs, only CC-885

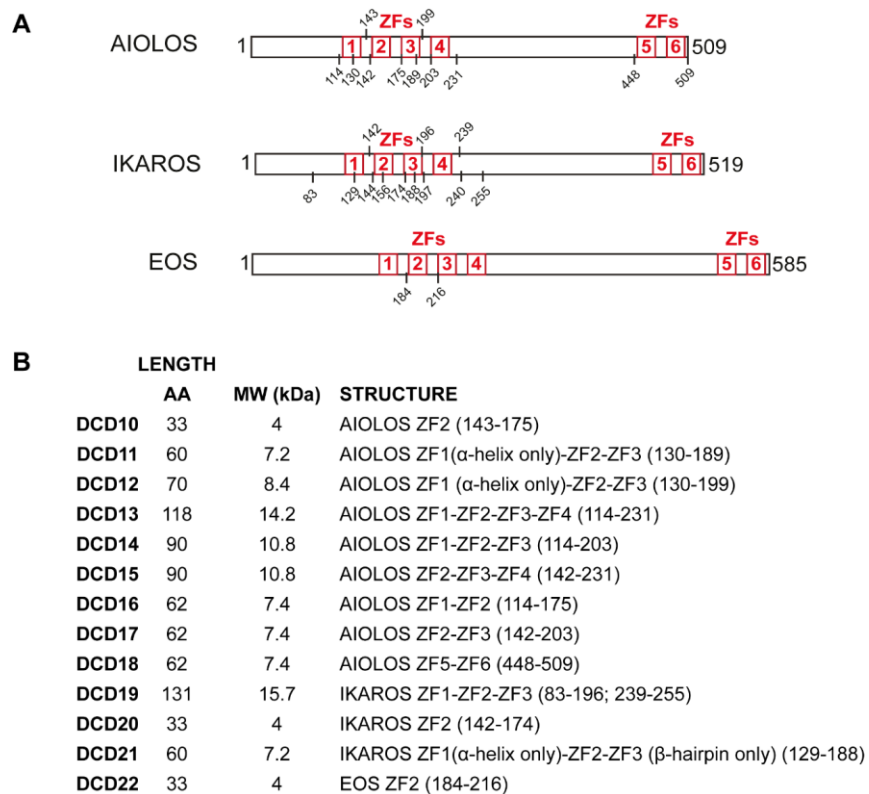
significantly decreased cell viability, starting at sub- $\mu\text{M}$  concentrations (**Figure 3.1. 6**). Collectively, these data highlighting the cytotoxicity of CC-885 prompted us to explore additional degron motifs from other CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates that would induce degradation of a target protein following treatment with an IMiD/CELMoD non-cytotoxic to non-haematological cell lines.



**Figure 3.1. 6 Examining cell viability of various cell lines following IMiD/CELMoD treatment.** The Promega CellTiter-Blue® cell viability assay was used to measure the cellular toxicity effects of the indicated IMiDs/CELMoDs at four different concentrations (0.1, 1, 10 and 30  $\mu\text{M}$ ) for five consecutive days. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

### 3.1.5 Design and generation of DCD10 to DCD22

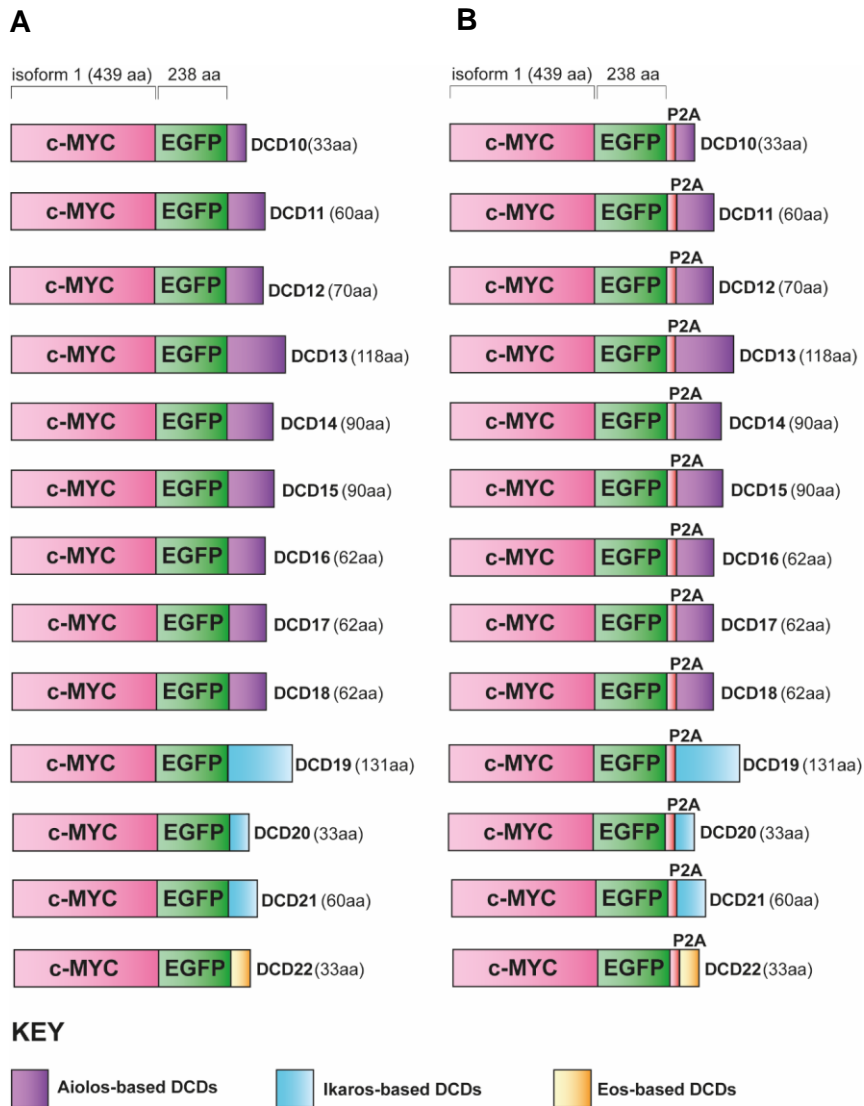
Since the DCD2/CC-885 pair did not represent a suitable combination for further evaluation, DCDs were designed based on the CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates Ikaros/IKZF1 and Aiolos/IKZF3. These proteins are degraded in the presence of IMiDs/CELMoDs that were shown to be non-cytotoxic in non-haematological cell lines, such as, thalidomide, lenalidomide, pomalidomide, CC-220 and CC-122 (**Figure 3.1. 6**) (Gandhi *et al.*, 2013; Krönke *et al.*, 2014; Lu *et al.*, 2014; Hagner *et al.*, 2015; Cubillos-Zapata *et al.*, 2016; Matyskiela *et al.*, 2017; Bjorklund *et al.*, 2020). The few published crystal structures of Ikaros and Aiolos suggest that a  $\beta$ -hairpin loop with a critical Gly at the apex, found as part of their C2H2 zinc finger 2 (ZF2) domain, mediates binding to the IMiD/CELMoD-bound CRL4<sup>CRBN</sup> E3 ubiquitin ligase (Krönke *et al.*, 2014; Petzold, Fischer and Thomä, 2016; Sievers *et al.*, 2018). These observations are further corroborated by in vitro studies and cellular degradation assays demonstrating that the ZF2 domains of Ikaros (residues 145-167) and Aiolos (residues 146-168) containing this  $\beta$ -hairpin loop, are critical for engaging the IMiD/CELMoD-CRBN complex (Krönke *et al.*, 2014; Sievers *et al.*, 2018). Generally, C2H2 ZF domains are found in tandem and function by binding DNA sequences 20-40 base pairs long (Fedotova *et al.*, 2017). The C2H2 ZF structure consists of two anti-parallel  $\beta$ -strands that form a  $\beta$ -hairpin loop at the turn, followed by an  $\alpha$ -helix (**Figure 1. 21 C**). Two  $\beta$ -sheet cysteine residues (C2) and two  $\alpha$ -helix histidine residues (H2) stabilize the tertiary structure of the ZF by coordinating a zinc atom (Fedotova *et al.*, 2017).



**Figure 3.1. 7 Structure and sequence-based design of DCD10 to DCD22 from Aiolos, Ikaros and Eos.** (A) Schematic representation of Aiolos, Ikaros and Eos protein sequences. The different ZF domains of Aiolos, Ikaros and Eos (ZF1-ZF6) are highlighted in red. (B) Summary of DCD10 to DCD22 sequences evaluated for this thesis.

A set of nine DCDs, named DCD10 to DCD18, were designed based on the full-length C2H2 ZF2 domain of Aiolos (**Figure 3.1. 7 A-B**). DCD10 to DCD17 contained the full Aiolos ZF2 domain and in the case of DCD11 to DCD17, this was extended to include multiple different combinations of ZF1, ZF3 and ZF4

domains, in order to explore the contribution of these ZF domains in mediating protein degradation. In contrast, DCD18 was based on the two C-terminal ZF domains of Aiolos, namely ZF5 and ZF6 (residues 448-509), aiming to test the relevance of these ZF domains in inducing degradation. Furthermore, a set of three DCDs, named DCD19 to DCD21, were designed based on Ikaros C2H2 ZF2 domain (**Figure 3.1. 7 A-B**). DCD20 was designed solely based on Ikaros ZF2 domain, whilst DCD19 and DCD21 contained additional residues to incorporate ZF1 and ZF3 domains. Lastly, a DCD, named DCD22, was designed based on the C2H2 ZF2 domain of the ZF transcription protein Eos (residues 187-209), which compared to both Ikaros and Aiolos, Eos ZF2 had a single amino acid difference, a Gln to His at position 188 (**Figure 3.1. 7 A-B**). Even though Eos has not been previously described as a CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrate, the rationale behind DCD22 design was to explore and compare the effects of single amino acid substitutions within the ZF2  $\beta$ -hairpin loop with respect to target protein degradation by the IMiD/CELMoD-CRL4<sup>CRBN</sup> E3 ubiquitin ligase complex. Similarly, to DCD1-DCD5 and DCD8, DCD10-DCD22 were fused at the C-terminus of cMYC-EGFP fusion construct (**Figure 3.1. 8 A**). P2A controls were also specifically designed for each DCD and were placed between the EGFP and the DCD motif (**Figure 3.1. 8 B**). The fusion constructs were commercially synthesised and cloned in the pLVX-TetOne lentiviral backbone vector under the control of a DOX-inducible promoter, prior to generating HMEC cells for stable expression.

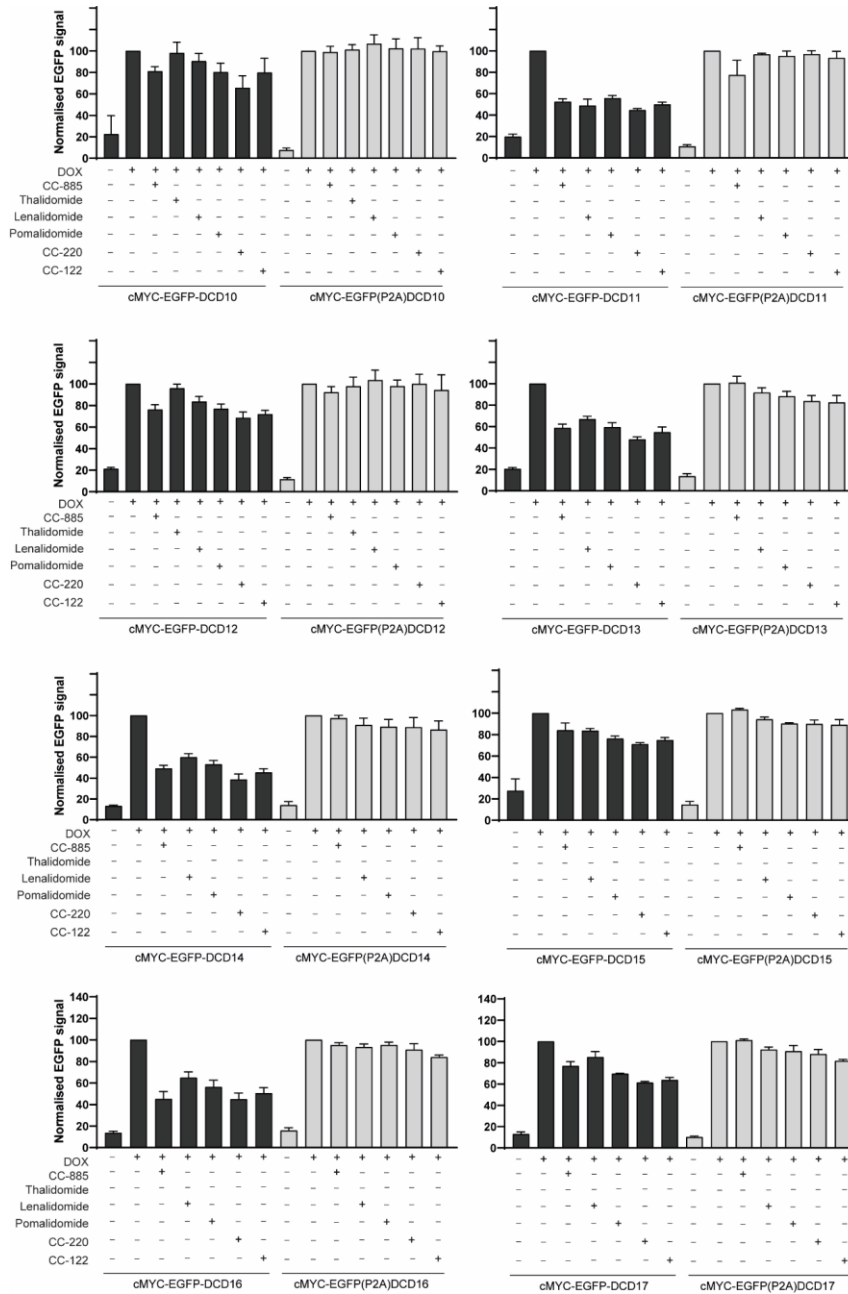


**Figure 3.1. 8 DCD10 to DCD22 fusion to the C-terminus of cMYC-EGFP for functional evaluation of the various DCDs. (A) Schematic representation of DCD10 to DCD22 fused to the C-terminus of cMYC-EGFP and (B) the respective P2A controls.**

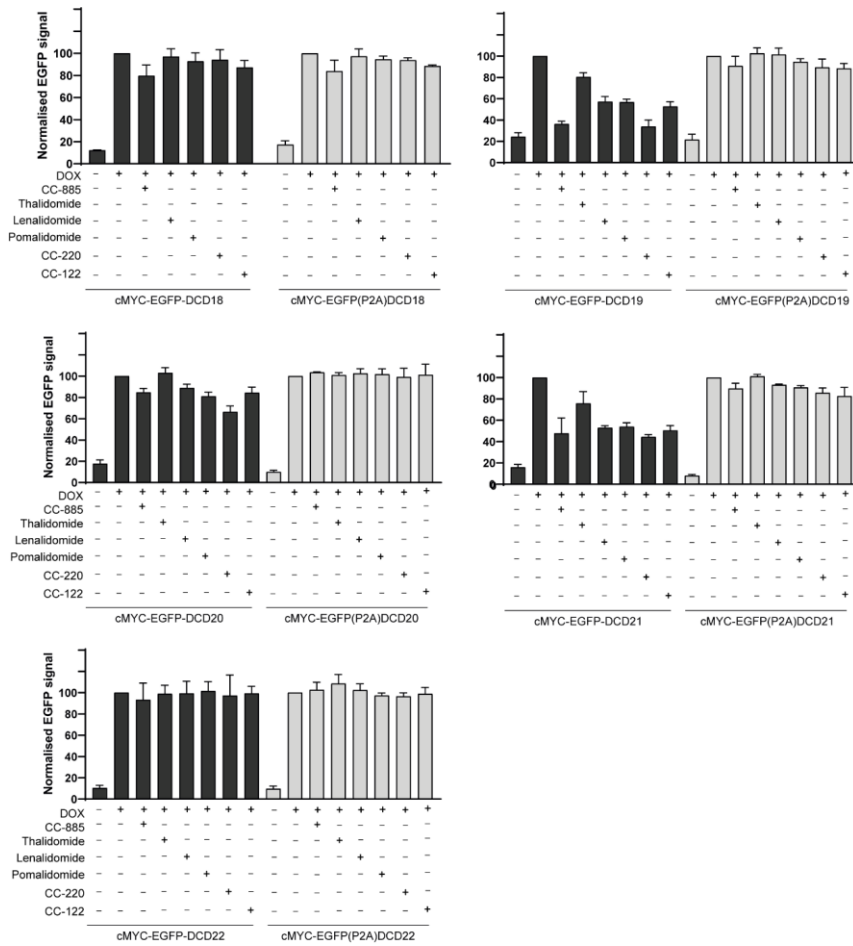
### **3.1.6 Aiolos and Ikaros-based DCDs induced a potent decrease of cMYC-EGFP levels following IMiD/CELMoD treatment**

HMEC cells stably expressing cMYC-EGFP fused to DCD10 to DCD22 were treated with either a vehicle (DMSO) or 10  $\mu$ M of a set of six IMiDs/CELMoDs (CC-885, thalidomide, lenalidomide, pomalidomide, CC-220 or CC-122) for 4h. Subsequently, cells were collected and analysed by flow cytometry to assess the loss of EGFP signal following treatment with the various IMiDs/CELMoDs. A robust loss of EGFP signal was observed with all IMiDs/CELMoDs for the cMYC-EGFP fusions with the Aiolos-based DCDs DCD10 to DCD17 and the Ikaros-based DCDs DCD19 to DCD21 (**Figure 3.1. 9**). CC-220 was identified as the most potent protein degrader inducing a variable reduction of EGFP signal in cells expressing cMYC-EGFP fused to DCD10 (residual signal: 65.62%), DCD11 (44.74%), DCD12 (68.59%), DCD13 (48.02%), DCD14 (41.68%), DCD15 (71.03%), DCD16 (45%), DCD17 (61.47%), DCD19 (38.85%), DCD20 (66.41%) and DCD21 (44.37%) (**Figure 3.1. 9**). In their respective P2A controls, the EGFP fluorescence levels remained unchanged, highlighting the specificity of the effect. In contrast, no loss of EGFP signal under any IMiD/CELMoD treatment was observed in HMEC cells expressing cMYC-EGFP fused to either DCD18 (based on the two C-terminal ZF domains of Aiolos, ZF5 and ZF6) or DCD22 (based on Eos ZF2 domain). This suggested both that ZF2 is essential for mediating the interaction with the IMiD/CELMoD-bound CRBN and that single amino acid differences in ZF2 sequence can dictate this

## Results

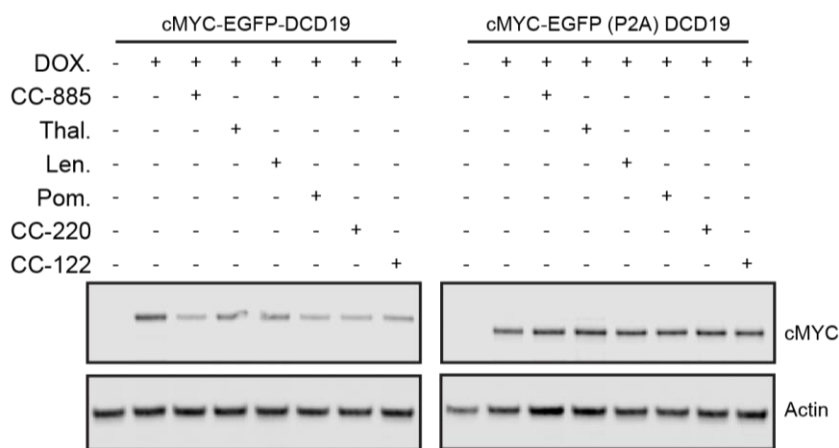






**Figure 3.1. 9 Functional evaluation of DCD10 to DCD22 fused to the C-terminus of cMYC-EGFP.** HMEC cells transduced with cMYC-EGFP fused to DCD10 to DCD22 and their respective P2A controls were treated with 1  $\mu\text{g/ml}$  DOX for 24h to stimulate expression of the fusion constructs, followed by treatment with 10  $\mu\text{M}$  of various IMiDs/CELMoDs for 4h. Median EGFP intensity was measured in each condition using flow cytometry. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

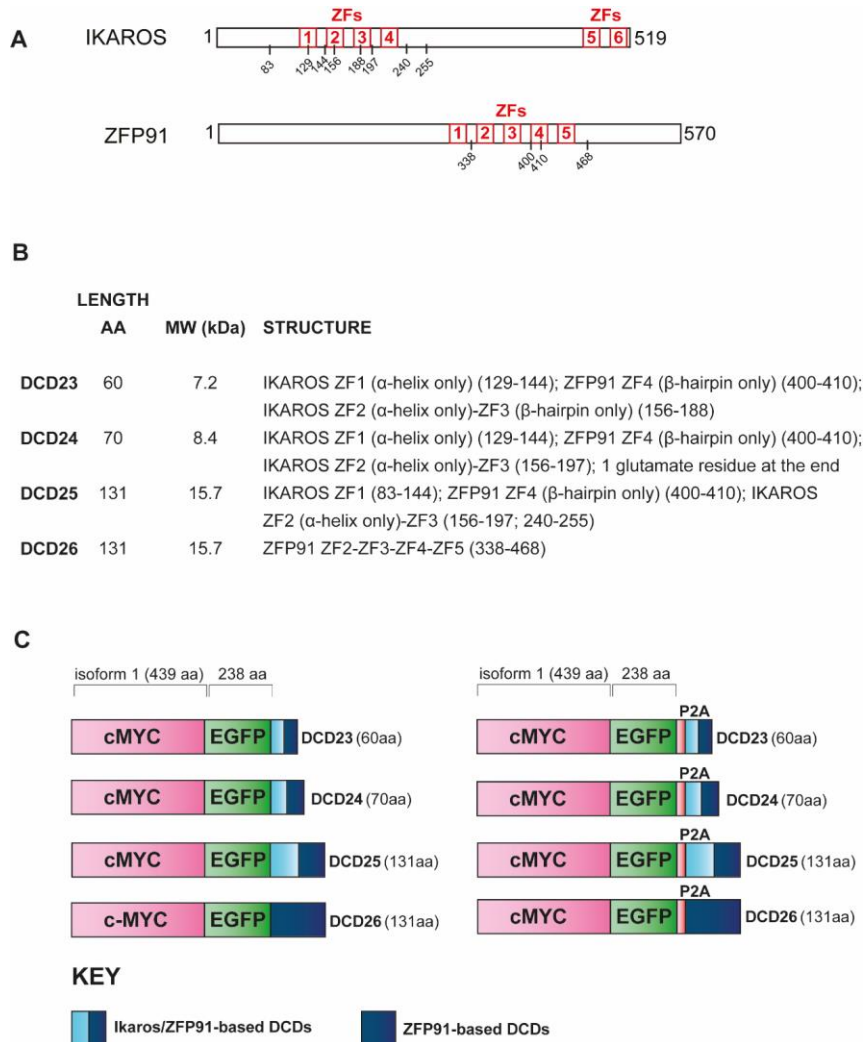
binding (**Figure 3.1. 9**). From all tested DCDs, DCD19 (131 residues) induced the most potent loss of EGFP signal (residual signal: 33.85%) in combination with CC-220 (**Figure 3.1. 9**). In addition, loss of cMYC-EGFP-DCD19 following 4h treatment with the IMiD/CELMoD panel was also validated by immunoblotting, with DCD19 being selected as the best DCD candidate for the development of a novel chemical-biology tool at this stage (**Figure 3.1. 10**).



**Figure 3.1. 10 Loss of cMYC-EGFP-DCD19 following IMiD/CELMoD treatment is verified by immunoblotting.** HMEC cells transduced with cMYC-EGFP-DCD19 or cMYC-EGFP(P2A)DCD19 were induced with 1  $\mu\text{g/ml}$  DOX for 24h, followed by treatment with 10  $\mu\text{M}$  of indicated IMiDs/CELMoDs for 4h. A cMYC antibody was used to probe the levels of the fusion constructs by Western blotting. Images are representative of 3 independent biological experiments (n=3).

### 3.1.7 Design and generation of DCD23 to DCD26

A third group of DCDs was also designed with the aim to identify an optimal DCD with a shorter peptide sequence compared to DCD19 (131 residues) and a more potent degradation of target proteins. This final group of DCDs was based on the previous work of Nicolas Thomä's group, which demonstrated in an in vitro binding assay that a chimeric ZF based on the  $\beta$ -hairpin loop of ZFP91 ZF4 fused to the  $\alpha$ -helix of Aiolos ZF2 had a higher binding affinity to CRBN compared to Aiolos ZF2-based or ZFP91 ZF4-based degrons (Sievers *et al.*, 2018). Based on these data and the sequence of previously characterised DCD19, a set of ZFP91-Ikaros chimeric DCDs was designed and named DCD23 to DCD25 (**Figure 3.1. 11 A-B**). These DCDs had in common the ZFP91 ZF4  $\beta$ -hairpin loop (residues 400-410) fused in tandem to the Ikaros ZF2  $\alpha$ -helix (residues 156-167). In order to define the optimal sequence length for target protein degradation, in each of the three chimeric DCDs this core sequence was extended to include a part or the whole of ZF1 and ZF3 domains. In DCD23, the chimeric ZF2 domain was extended to include the  $\alpha$ -helix of Ikaros ZF1 and the  $\beta$ -hairpin loop of Ikaros ZF3 domain, whilst DCD24 encompassed the  $\alpha$ -helix of Ikaros ZF1 and the whole Ikaros ZF3 while DCD25 sequence included the whole Ikaros ZF1 and ZF3 domains (**Figure 3.1. 11 A-B**). Furthermore, a DCD, named DCD26, was designed based on an extended ZFP91 ZF4 domain to include ZFP91 ZF2, ZF3 and ZF5, to be used for comparing the degradation efficiency of the chimeric DCDs, DCD23 to DCD25

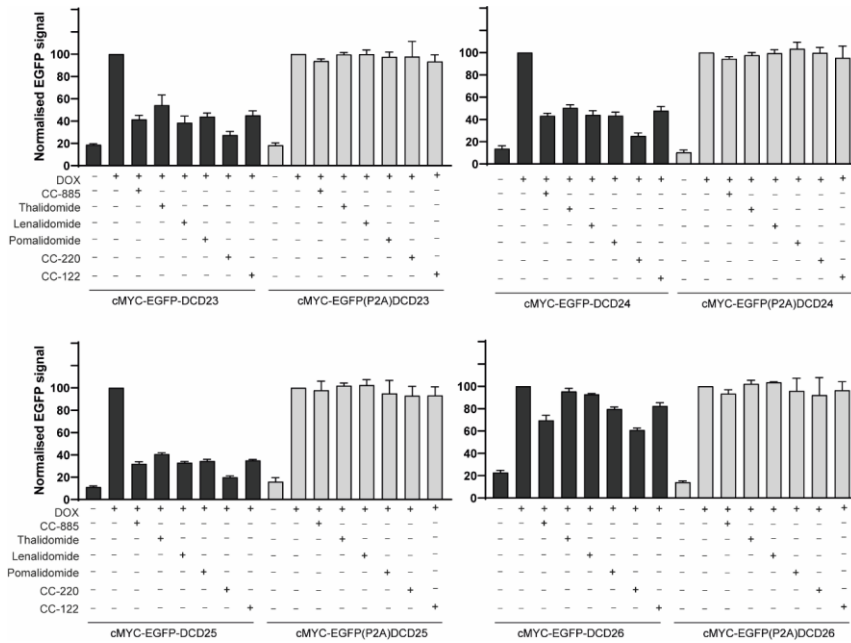


**Figure 3.1. 11 Structure and sequence-based design of DCD23 to DCD26 from *CUL4<sup>CRBN</sup>* E3 ubiquitin ligase neosubstrates Ikaros and ZFP91. (A) Schematic representation of Ikaros and ZFP91 protein sequences. The different ZF domains of Ikaros (ZF1-ZF6) and ZFP91 (ZF1-ZF5) are highlighted in red. (B) Summary of DCD23 to DCD26 sequences evaluated for this thesis. (C) Schematic representation of DCD23 to DCD26 fused to the C-terminus of cMYC-EGFP and the respective P2A controls.**

(**Figure 3.1. 11 A-B**) (An, Charles M. Ponthier, *et al.*, 2017). DCD23 to DCD26 were commercially synthesised and cloned in the pLVX-TetOne lentiviral backbone vector as C-terminal fusions to cMYC-EGFP, under the control of a DOX-inducible promoter (**Figure 3.1. 11 C**). Like the previous two iterations of DCD design, P2A controls were also specifically designed for each of these DCDs and were placed between the EGFP and the DCD motif (**Figure 3.1. 11 C**). Subsequently, these vectors were used to generate HMEC cells stably expressing the fusion constructs.

### ***3.1.8 Chimeric DCDs induced a potent decrease in cMYC-EGFP levels following IMiD/CELMoD treatment***

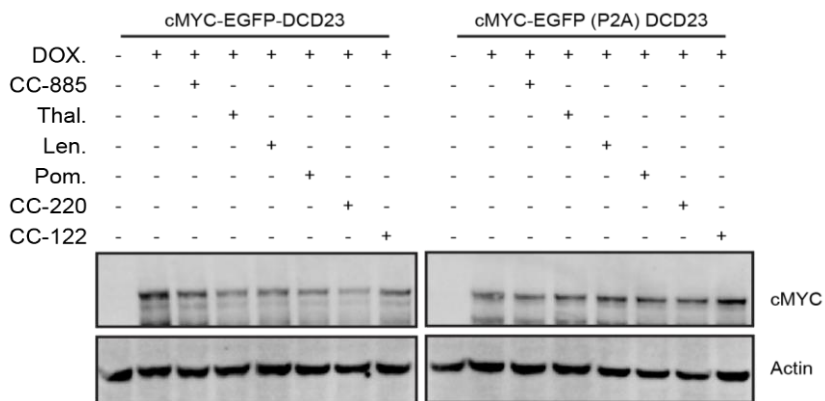
HMEC cells stably expressing cMYC-EGFP-DCD23 to DCD26 were treated with either a vehicle (DMSO) or the following IMiDs/CELMoDs at 10  $\mu$ M for 4h: CC-885, thalidomide, lenalidomide, pomalidomide, CC-220 and CC-122. Flow cytometry analysis of all tested fusion constructs demonstrated a robust reduction in EGFP levels with all IMiDs/CELMoDs (**Figure 3.1. 12**). CC-220 was identified as the most optimal degrader compound, inducing a strong loss of EGFP fluorescence signal in cells expressing cMYC-EGFP fused to DCD23 (residual signal: 27.45%), DCD24 (25.22%), DCD25 (19.84%), and DCD26 (60.9%) (**Figure 3.1. 12**). No decrease of EGFP fluorescence signal was observed in the corresponding P2A controls, due to the uncoupling of the DCD from the cMYC-EGFP fusion protein. ZFP91 ZF4 domain-based DCD26 was less efficient compared to the ZFP91/Ikaros chimeric DCDs in mediating loss of



**Figure 3.1.12 Functional evaluation of DCD23 to DCD26 fused to the C-terminus of cMYC-EGFP.** HMEC cells transduced with cMYC-EGFP fused to DCD23 to DCD26 and their respective P2A controls were treated with 1  $\mu\text{g/ml}$  DOX for 24h to stimulate expression of the fusion constructs, followed by treatment with 10  $\mu\text{M}$  of various IMiDs/CELMoDs for 4h. Median EGFP intensity was measured in each condition using flow cytometry. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments ( $n=3$ ).

cMYC-EGFP following treatment with all IMiDs/CELMoDs (**Figure 3.1.12**). Since all three chimeric DCDs induced a potent loss of EGFP fluorescence signal in the presence of 10  $\mu\text{M}$  CC-220 for 4h, DCD23 with the significantly shorter peptide sequence (60 residues) was further verified by immunoblotting to induce loss of cMYC-EGFP and selected from this cohort of tested DCDs as another good candidate for the development of a novel chemical-biology tool (**Figure 3.1.13**). Fluorescence cytometry data obtained from degradation assays on DCD1 to

DCD26 motifs fused to cMYC-EGFP and the respective P2A controls are summarised in **Table 3.1. 1**.



**Figure 3.1. 13** Loss of cMYC-EGFP-DCD23 following IMiD/CELMoD treatment is verified by immunoblotting. HMEC cells transduced with cMYC-EGFP-DCD23 and cMYC-EGFP(P2A)DCD23 were induced with 1  $\mu\text{g}/\text{ml}$  DOX for 24h, followed by treatment with 10  $\mu\text{M}$  of indicated IMiDs/CELMoDs for 4h. A cMYC antibody was used to probe for the levels of the fusion constructs by Western blotting. Images are representative of 3 independent biological experiments (n=3).

	-DOX	DMSO	CC-885	THAL	LEN	POM	CC-220	CC-122
cMYC-EGFP-25mer	12.15	100.00	63.75	92.23	84.59	76.24	52.99	72.32
cMYC-EGFP(P2A)25mer	11.30	100.00	85.59	95.24	98.92	100.55	102.87	96.03
cMYC-EGFP-DCD1	19.40	100.00	37.63	98.59	100.78	98.19	98.69	96.23
cMYC-EGFP(P2A)DCD1	26.65	100.00	99.16	100.72	101.64	99.78	96.96	97.11
cMYC-EGFP-DCD2	13.02	100.00	35.11	103.02	114.83	111.40	115.30	106.01
cMYC-EGFP(P2A)DCD2	19.27	100.00	101.35	103.98	99.36	99.09	103.72	100.09
cMYC-EGFP-DCD3	21.06	100.00	93.54	99.72	98.13	97.24	97.34	96.77
cMYC-EGFP(P2A)DCD3	20.76	100.00	101.85	99.36	98.41	101.49	97.95	101.97
cMYC-EGFP-DCD4	38.12	100.00	90.70	97.95	100.89	92.94	93.95	91.78
cMYC-EGFP(P2A)DCD4	12.15	100.00	92.45	96.16	112.75	112.71	108.14	112.35
cMYC-EGFP-DCD5	24.19	100.00	93.08	96.13	89.58	87.53	82.80	83.23
cMYC-EGFP(P2A)DCD5	12.43	100.00	95.29	96.50	99.92	100.70	101.12	100.53
cMYC-EGFP-DCD8	49.58	100.00	96.26	98.75	96.37	95.66	93.90	99.04
cMYC-EGFP(P2A)DCD8	28.94	100.00	95.96	104.05	103.62	103.82	104.35	104.84
cMYC-EGFP-DCD10	22.47	100.00	81.07	98.15	90.46	80.20	65.62	79.80
cMYC-EGFP(P2A)DCD10	7.64	100.00	98.97	101.15	106.79	102.43	102.13	99.66
cMYC-EGFP-DCD11	19.72	100.00	52.43		48.83	55.80	44.74	49.95
cMYC-EGFP(P2A)DCD11	10.77	100.00	77.42		96.69	95.17	96.84	93.35
cMYC-EGFP-DCD12	21.24	100.00	76.10	95.95	83.56	76.82	68.59	71.85
cMYC-EGFP(P2A)DCD12	11.58	100.00	92.24	97.74	103.43	97.85	99.90	94.17
cMYC-EGFP-DCD13	20.46	100.00	58.68		66.91	59.48	48.02	54.62
cMYC-EGFP(P2A)DCD13	13.65	100.00	100.90		91.61	88.25	83.59	82.46
cMYC-EGFP-DCD14	13.11	100.00	49.21		59.98	53.22	38.65	45.32
cMYC-EGFP(P2A)DCD14	13.93	100.00	97.42		90.96	89.15	88.70	86.42
cMYC-EGFP-DCD15	27.64	100.00	84.05		83.68	76.26	71.03	74.87
cMYC-EGFP(P2A)DCD15	14.59	100.00	103.33		94.20	90.30	89.93	89.10
cMYC-EGFP-DCD16	13.72	100.00	45.22		64.96	56.35	45.00	50.41
cMYC-EGFP(P2A)DCD16	15.86	100.00	95.14		93.33	95.26	90.96	84.07
cMYC-EGFP-DCD17	12.99	100.00	76.93		85.30	69.71	61.47	63.78
cMYC-EGFP(P2A)DCD17	10.25	100.00	101.19		92.38	90.87	88.13	81.74
cMYC-EGFP-DCD18	12.14	100.00	79.79		97.06	92.82	94.26	87.15
cMYC-EGFP(P2A)DCD18	17.33	100.00	83.83		97.32	94.51	93.76	88.48
cMYC-EGFP-DCD19	24.34	100.00	36.20	80.44	57.17	56.73	33.85	52.64
cMYC-EGFP(P2A)DCD19	21.53	100.00	90.81	102.61	101.60	94.63	89.44	88.50
cMYC-EGFP-DCD20	17.70	100.00	84.74	103.05	88.94	81.00	66.41	84.34
cMYC-EGFP(P2A)DCD20	10.00	100.00	103.50	100.97	102.62	101.82	99.16	101.32
cMYC-EGFP-DCD21	15.97	100.00	47.67	75.76	52.91	53.90	44.37	50.36
cMYC-EGFP(P2A)DCD21	8.17	100.00	89.79	101.07	93.14	90.74	85.61	82.64
cMYC-EGFP-DCD22	10.39	100.00	93.39	98.98	99.32	101.59	97.29	99.37
cMYC-EGFP(P2A)DCD22	9.66	100.00	102.76	108.53	102.43	97.34	96.51	98.83
cMYC-EGFP-DCD23	18.69	100.00	41.46	54.18	38.58	44.06	27.45	45.07
cMYC-EGFP(P2A)DCD23	18.31	100.00	93.74	99.55	99.91	97.57	97.72	93.36
cMYC-EGFP-DCD24	13.75	100.00	43.23	50.35	44.00	43.39	25.22	47.61
cMYC-EGFP(P2A)DCD24	10.48	100.00	94.55	97.76	99.42	103.37	99.77	95.27
cMYC-EGFP-DCD25	11.11	100.00	31.97	40.64	32.90	34.42	19.84	34.92
cMYC-EGFP(P2A)DCD25	15.87	100.00	97.80	101.93	102.36	95.08	92.93	93.26
cMYC-EGFP-DCD26	22.80	100.00	69.63	95.41	92.85	79.66	60.90	82.44
cMYC-EGFP(P2A)DCD26	14.21	100.00	93.40	102.14	103.50	95.80	92.12	96.39

Degradation normalised to DMSO treated control



not tested

Table 3.1. 1 Summary of flow cytometry data on the degradation of all the DCDs tested.

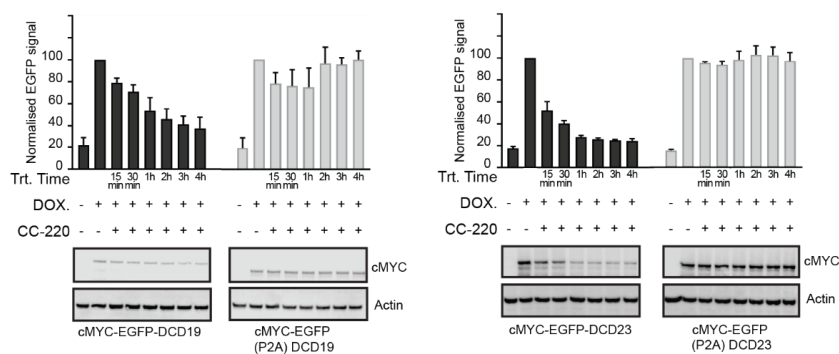
Data are representative of 3 independent biological experiments (n=3).

Commented [SN7]: Modified the table to include numbers. Highlighted the 3 leading DCDs.



### **3.1.9 Acute degradation of cMYC-EGFP fused to DCD19 and DCD23 following IMiD/CELMoD treatment**

An advantage of chemical-biology tools over genetic methods is rapid depletion of the target protein, which prevents adaptation of the cells to target protein loss and the manifestation of phenotypes not associated with loss of the target's functions. Therefore, to determine the earliest time-point of robust and potent degradation of the cMYC-EGFP fusion construct, HMEC cells stably expressing cMYC-EGFP fused C-terminally to either DCD19 or DCD23 and their corresponding P2A controls were treated with 10  $\mu$ M CC-220 in a time-course assay. Cells were harvested at different time-points and subsequently analysed by flow cytometry. Flow cytometry analysis showed that within 15min of CC-220 treatment there is a reduction of EGFP fluorescence signal in HMEC cells expressing cMYC-EGFP-DCD19 (residual signal 79.12%) and cMYC-EGFP-DCD23 (52.19%) (**Figure 3.1. 14**). Maximum loss of EGFP fluorescence signal was achieved at 4h of CC-220 treatment in cells expressing cMYC-EGFP-DCD19 (residual signal: 37.62%) and cMYC-EGFP-DCD23 (24.36%) (**Figure 3.1. 14**). CC-220-treatment of HMEC cells expressing cMYC-EGFP(P2A) fused to either DCD19 or DCD23 did not induce a decrease in fusion protein levels, as expected. Furthermore, these flow cytometry observations were confirmed by immunoblot analysis of cellular lysates, showing a decrease in cMYC-EGFP protein levels fused to either DCD19 or DCD23, as early as 15min following CC-220 treatment, but not of their corresponding P2A controls (**Figure 3.1. 14**).

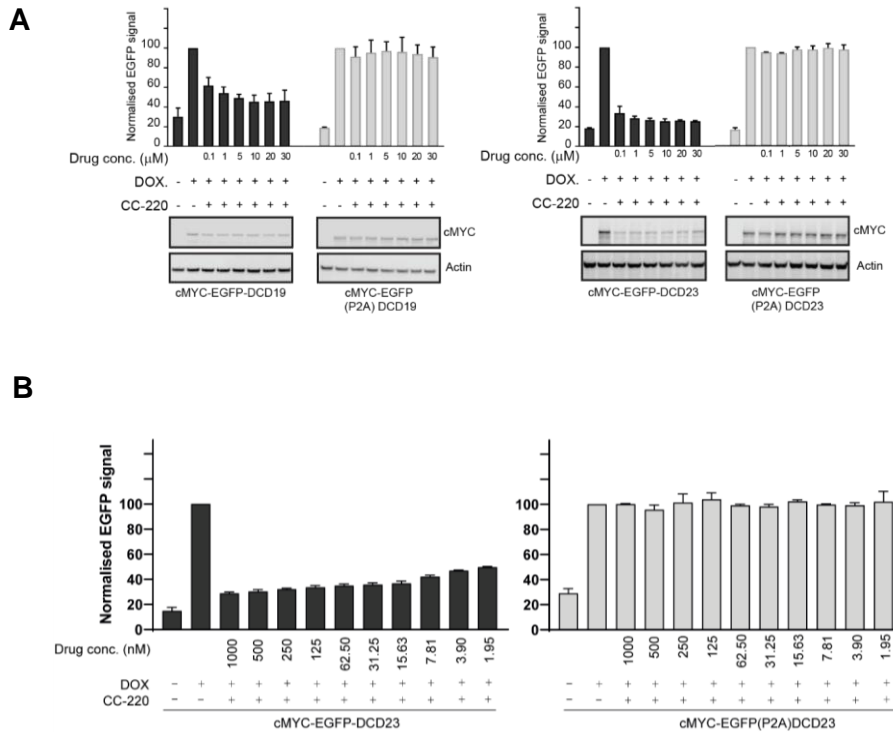


**Figure 3.1.14 Determining the earliest timepoint of IMiD/CELMoD-dependent target protein degradation.** HMEC cells transduced with cMYC-EGFP-DCD19, cMYC-EGFP-DCD23 and their respective P2A controls were treated with 1  $\mu\text{g}/\text{ml}$  DOX for 24h to stimulate expression of the fusion constructs. Subsequently, cells were subjected to 10  $\mu\text{M}$  CC-220 treatment in a time-course assay. Loss of target protein was validated via measuring EGFP median intensity by flow cytometry and immunoblotting for the cMYC protein. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

### 3.1.10 Sub-micromolar concentrations of IMiDs/CELMoDs induced a potent degradation of cMYC-EGFP fused to DCD19 and DCD23

An important aspect in developing a novel chemical-biology tool is to employ compounds that can induce the effective degradation of a target protein when administered at low  $\mu\text{M}$  concentrations, so as to minimise off-target effects that could mask true phenotypes following target loss. Therefore, to determine the minimum concentration of CC-220 required to achieve a robust and potent degradation of the cMYC-EGFP fusion construct, HMEC cells stably expressing

cMYC-EGFP fused C-terminally to either DCD19 or DCD23 and their corresponding P2A controls were treated with increasing CC-220 concentrations for 4h starting at 0.1  $\mu$ M. Concentrations as low as 0.1  $\mu$ M of CC-220 caused a measurable loss of cMYC-EGFP fused to DCD19 (residual signal: 62.05%) and DCD23 (33.6%), demonstrating the high potency of this CELMoD (**Figure 3.1. 15 A**). Maximum loss of EGFP signal at 4h was achieved following treatment with 10  $\mu$ M CC-220 in HMEC cells expressing cMYC-EGFP fused to DCD19 (residual signal: 45.4%) and DCD23 (25.54%) (**Figure 3.1. 15 A**). Conversely, treatment of HMEC cells expressing cMYC-EGFP(P2A) fused to either DCD19 or DCD23 with CC-220 did not induce a decrease in fusion protein levels. These flow cytometry observations were also confirmed by immunoblot analysis of cellular lysates (**Figure 3.1. 15 A**). Since CC-220 was found to effectively decrease EGFP fluorescence levels when administered at 0.1  $\mu$ M for 4h, a second dose titration assay was performed to determine the minimum effective concentration for this compound. Similarly, CC-220 was administered at concentrations spanning 1.95 nM to 1  $\mu$ M for 4h in DOX-treated HMEC cells expressing cMYC-EGFP-DCD23 and cMYC-EGFP(P2A)DCD23. Interestingly, even 1.95 nM of CC-220 was sufficient to induce strong loss of EGFP fluorescence signal in cMYC-EGFP-DCD23 expressing cells (residual signal: 49.88%) with no effects observed in its corresponding P2A control (**Figure 3.1. 15 B**). Collectively, this set of data demonstrated both the high degradation potency of CC-220 even at nM concentrations and the tunability of the system, with higher CC-220 concentrations inducing a stronger loss of EGFP fluorescence signal.

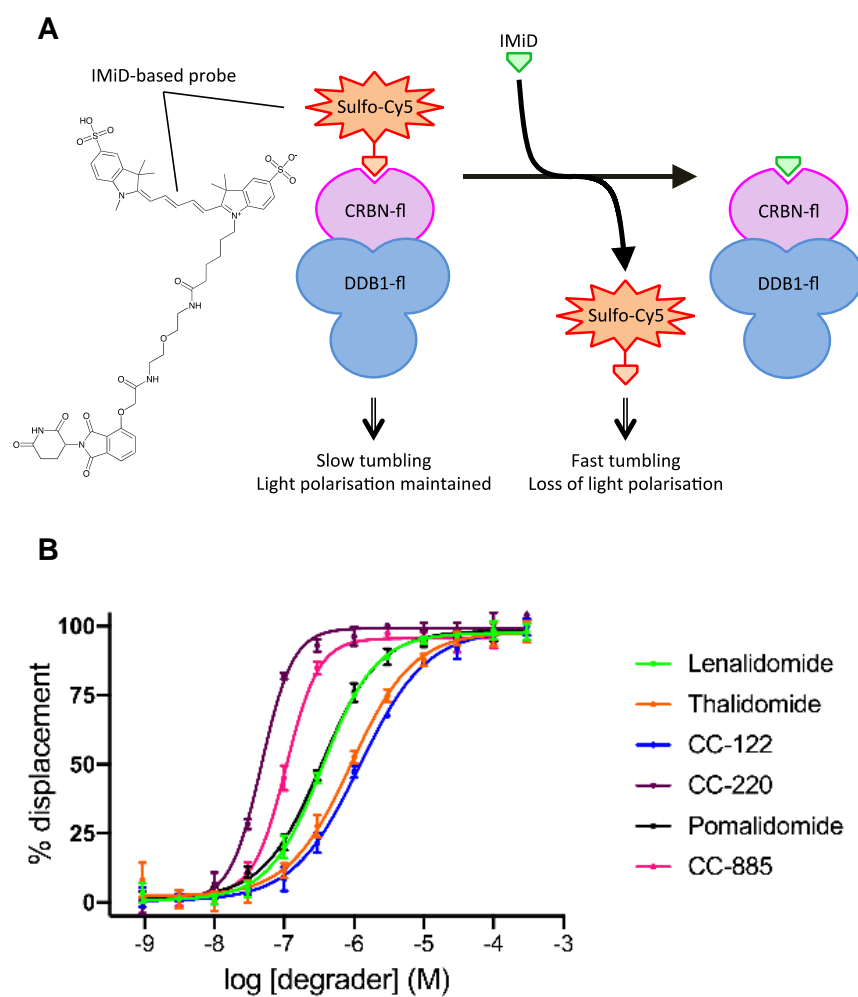


**Figure 3.1. 15 Determining the minimum IMiD/CELMoD concentration required for target protein degradation.** (A) HMEC cells transduced with either cMYC-EGFP-DCD19 or cMYC-EGFP-DCD23 and their respective P2A controls were treated with 1  $\mu\text{g/ml}$  DOX for 24h to stimulate expression of the fusion constructs. Cells were subsequently treated with increasing concentrations of CC-220, ranging from 0.1 to 30  $\mu\text{M}$  for 4h. Loss of target protein was validated via measuring EGFP median intensity by flow cytometry and immunoblotting for the cMYC protein. (B) Treatment of HMEC cells stably expressing cMYC-EGFP-DCD23 and cMYC-EGFP(P2A)DCD23 with CC-220 for 4h in a dose-titration assay (CC-220 concentrations range from 0.0195 to 1  $\mu\text{M}$ ). Median EGFP intensity was measured in each condition using flow cytometry. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments ( $n=3$ ).

### 3.1.11 DCDs efficiently engage CRBN/DDB1 in vitro

In order to measure the binding affinities of IMiDs/CELMoDs to CRBN and of candidate DCDs to the IMiD/CELMoD-CRBN complex, two types of in vitro assays were developed by our in-house collaborators and used to quantify these interactions. First, a FP (Fluorescence Polarisation) assay was established for validating and measuring the binding affinities of the following IMiDs/CELMoDs to CRBN: CC-885, thalidomide, lenalidomide, pomalidomide, CC-220 and CC-122 (**Figure 3.1. 16 A**). In principle, in an FP assay, plane-polarized light is used to excite a fluorescent molecule, with the subsequent change in polarization of emitted light providing a measurement of binding affinity (Lea and Simeonov, 2011). For this FP assay, an IMiD-based probe labelled with SulfoCyanine5 fluorophore was synthesized and found to exhibit a strong binding affinity to the CRBN/DDB1 complex, with an apparent equilibrium dissociation constant ( $K_d$ ) of 36 nM. When this fluorescently-labelled IMiD-probe was bound to CRBN/DDB1 complex light polarisation was maintained. Addition of an IMiD/CELMoD to the system would displace the probe from the complex causing a concomitant change in the polarization of emitted light due to the slower tumbling of the fluorescent molecule SulfoCyanine5, effectively allowing to measure the binding efficiency of each IMiD/CELMoD. Therefore, the relative affinity of each IMiD/CELMoD was subsequently assessed by measuring its ability to displace the IMiD-based probe from the CRBN/DDB1 complex (**Figure 3.1. 16 B**). All tested IMiDs/CELMoDs were able to bind to CRBN with half maximal inhibitory concentrations ( $IC_{50}$ ) spanning from 0.048 to 1.208  $\mu$ M, demonstrating the high

binding affinity of these degrader compounds to the CRBN/DDB1 complex (Figure 3.1. 16 B; Table 3.1. 2).



**Figure 3.1. 16** In vitro measurement of various IMiD/CELMoD affinities for CRBN/DDB1 complex. (A) Schematic representation of the IMiD-based FP assay used to evaluate the relative affinities of the various IMiDs/CELMoDs towards CRBN/DDB1 complex. (B) IMiD/CELMoD-

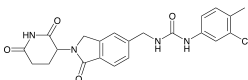
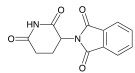
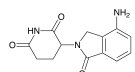
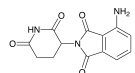
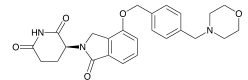
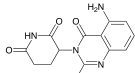
based FP assay curves obtained for the various IMiDs/CELMoDs. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent experiments (n=3).

The second type of in vitro assay developed was a TR-FRET (time-resolved fluorescence resonance energy transfer) assay, which was employed to measure the ability of DCD1, DCD19 and DCD23 in binding a pre-formed IMiD/CELMoD-CRBN/DDB1 complex (**Figure 3.1. 17 A**). To achieve this, an N-terminally SulfoCyanine5 fluorescent-labelled Aiolos peptide was synthesized, based on the sequence of DCD10 (Aiolos ZF2 domain) and was found to bind CRBN with an apparent  $K_d$  of 0.8  $\mu\text{M}$  in the presence of 10  $\mu\text{M}$  lenalidomide (**Figure 3.1. 7 A-B**). The displacement of the Aiolos peptide-based probe was used as a surrogate to assess the relative binding affinities of DCDs for the different IMiD/CELMoD-CRBN/DDB1 complexes. Even though the DCD1/CC-885 pair was identified as a non-optimal DCD sequence due to the cytotoxicity of CC-885 in non-haematological cell lines, it was selected for in vitro evaluation to further explore its CC-885 binding specificity. DCD1 was found to bind the IMiD/CELMoD-CRBN/DDB1 complex with high affinity only when CC-885 was used ( $\text{IC}_{50}$  of 7 nM with CC-885) (**Figure 3.1. 17 B; Table 3.1. 2**). This result was in line with cellular data showing that only CC-885 efficiently induces the degradation of cMYC-EGFP fused to either DCD1 or its shorter variant DCD2 (**Figure 3.1. 1; Figure 3.1. 5 B**). On the contrary, DCD19 was found to engage the various IMiD/CELMoD-CRBN/DDB1 complexes with similar affinities ( $\text{IC}_{50}$ s ranging between 0.09 and 0.2  $\mu\text{M}$ ), contrary to the cellular degradation assay which identified CC-220 as the most potent degrader (**Figure 3.1. 9**). Unlike DCD19, the shorter construct DCD23 was found to bind to the various

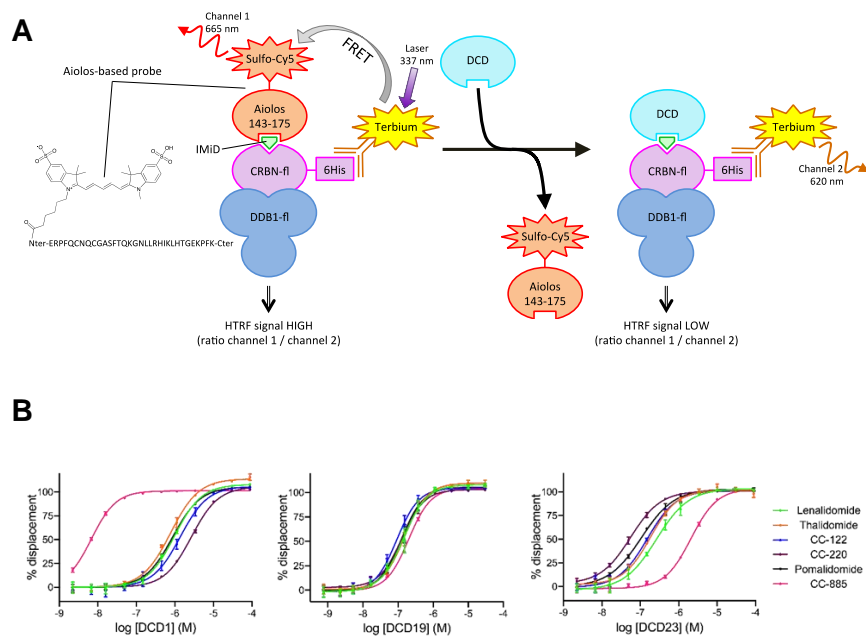
IMiD/CELMoD-CRBN/DDB1 complexes with different affinities depending on which IMiD/CELMoD was used. The least potent was CC-885 ( $IC_{50}$ , 2.1  $\mu$ M) and the most potent was CC-220 ( $IC_{50}$ , 0.06  $\mu$ M). While the CC-220 data were in line with the cellular degradation assays in HMEC cells, the CC-885 data were harder to correlate as CC-885 was a stronger degrader compared to other tested IMiD/CELMoDs in the cellular degradation assay but exhibited the weakest binding affinity for DCD23 ( $IC_{50}$  of 2.087  $\mu$ M) (**Figure 3.1. 17 B; Table 3.1. 2**).

Collectively, the in vitro data on DCD19 and DCD23 highlight that the binding affinities of DCD19 and DCD23 to the IMiD/CELMoD-CRBN/DDB1 complex and degradation efficiency for the target protein are not directly correlated (**Figure 3.1. 9; Figure 3.1. 12; Table 3.1. 2**). Despite this, the data confirmed the direct binding of DCD1, DCD19 and DCD23 to the IMiD/CELMoD-bound CRBN/DDB1 complex with high affinity and further supported the use of both DCD19 and DCD23 as ideal candidates for the development of a novel chemical-biology tool.



Degradar	2D structure	FP IC <sub>50</sub> ( $\mu$ M)	TR-FRET IC <sub>50</sub> ( $\mu$ M)			Percentage of residual EGFP signal at 10 $\mu$ M degrader			
			DCD1	DCD19	DCD23	DCD1	DCD2	DCD19	DCD23
CC-885		0.107 $\pm$ 0.011	0.007 $\pm$ 0.0003	0.200 $\pm$ 0.002	2.087 $\pm$ 0.111	37.63 +/- 1.30	35.11 +/- 11.84	36.20 +/- 2.73	41.46 +/- 3.60
Thalidomide		0.911 $\pm$ 0.137	0.780 $\pm$ 0.114	0.157 $\pm$ 0.032	0.184 $\pm$ 0.012	98.59 +/- 8.19	103.02 +/- 18.62	80.44 +/- 3.98	54.18 +/- 9.11
Lenalidomide		0.353 $\pm$ 0.039	0.986 $\pm$ 0.023	0.144 $\pm$ 0.026	0.265 $\pm$ 0.017	100.78 +/- 10.83	114.83 +/- 21.17	57.17 +/- 4.97	38.58 +/- 5.78
Pomalidomide		0.342 $\pm$ 0.021	0.874 $\pm$ 0.031	0.127 $\pm$ 0.027	0.107 $\pm$ 0.004	98.19 +/- 6.69	111.40 +/- 22.85	56.73 +/- 2.95	44.06 +/- 3.18
CC-220		0.048 $\pm$ 0.004	2.681 $\pm$ 0.106	0.130 $\pm$ 0.037	0.057 $\pm$ 0.011	98.69 +/- 5.82	115.30 +/- 29.08	33.85 +/- 6.13	27.45 +/- 3.25
CC-122		1.208 $\pm$ 0.119	1.401 $\pm$ 0.260	0.094 $\pm$ 0.025	0.153 $\pm$ 0.021	96.23 +/- 6.29	106.01 +/- 12.66	52.64 +/- 4.62	45.07 +/- 4.17

**Table 3.1. 2 Summary of the data obtained in vitro and in cells.** All results presented in this table are mean values of 3 independent determinations.

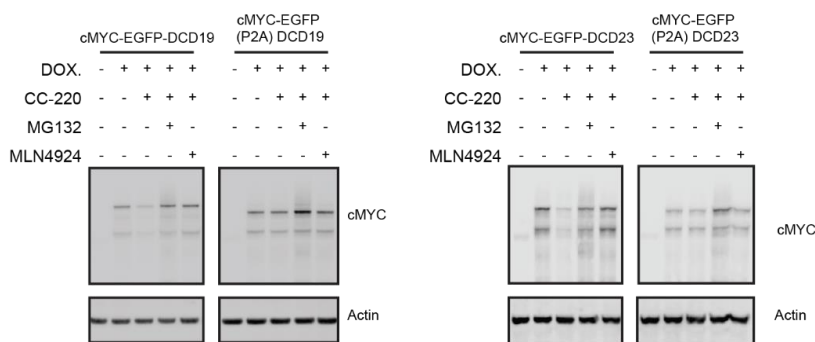


**Figure 3.1. 17** *In vitro* measurement of IMiD/CELMoD efficacy at recruiting DCDs to the CRBN/DDB1 complex. **(A)** Schematic representation of the Aiolos peptide-based TR-FRET assay used to evaluate the ability of IMiDs/CELMoDs to promote the binding of DCDs to CRBN/DDB1 complex. **(B)** Aiolos peptide-based TR-FRET assay curves obtained with DCD1, DCD19 and DCD23 and the various IMiDs/CELMoDs used in this thesis. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent experiments ( $n=3$ ).

### 3.1.12 cMYC degradation is mediated by the UPS

Validating that DCD19 and DCD23-mediated degradation of cMYC-EGFP following treatment with an IMiD/CELMoD is due to the UPS (Ubiquitin-Proteasome System), was essential for demonstrating that these DCDs retain their neosubstrate sequence-specific ability to promote proteasomal degradation. Since IMiDs/CELMoDs have been previously demonstrated to engage the CUL4<sup>CRBN</sup> E3 ubiquitin ligase (Ito *et al.*, 2010), a component of the UPS system, the pan-Cullin inhibitor MLN4924 was employed to confirm binding of DCD-tagged proteins to the E3 ligase complex. MLN4924 is a potent and selective small molecule NAE (NEDD8-activating enzyme) inhibitor, effectively blocking neddylation of cullins, thereby preventing the activation of the CUL4<sup>CRBN</sup> E3 ubiquitin ligase and subsequently ubiquitination of binding targets. Further to MLN4924, the compound MG132 was also employed to explore the role of the 26S proteasome in the degradation of the target proteins. MG132 functions by binding to the active site of the beta ring of the 20S proteasome core thus effectively blocking the proteolytic activity of the 26S proteasome complex. To that extent, HMEC cells stably expressing these fusion proteins were treated with 10  $\mu$ M CC-220 alone or in the presence of either 0.5  $\mu$ M MG132 or 0.1  $\mu$ M MLN4924 for 4h. Immunoblot analysis of HMEC cells expressing cMYC-EGFP fused to either DCD19 or DCD23 showed protein loss in the presence of CC-220 alone, as compared to vehicle control (**Figure 3.1. 18**). As expected, treatment with CC-220 did not affect the protein levels of the corresponding P2A controls. However, co-treatment of CC-220 with either MG132 or MLN4924 inhibited

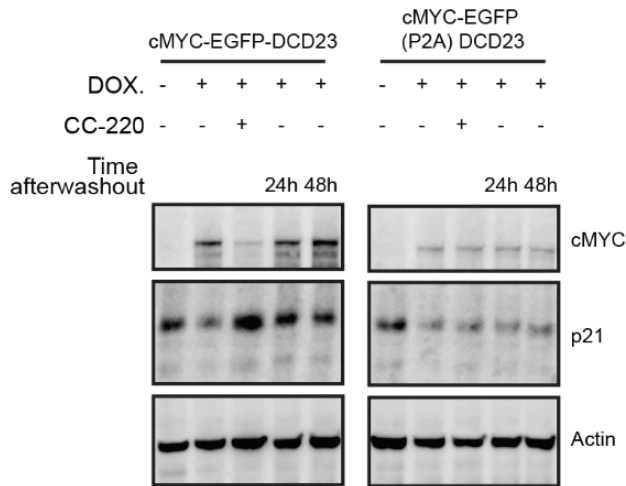
protein loss, demonstrating that loss of cMYC-EGFP-DCD19 and cMYC-EGFP-DCD23 occurs through the activity of the UPS system (**Figure 3.1. 18**).



**Figure 3.1. 18 DCD19 and DCD23-induced loss of target protein is dependent on the UPS system.** HMEC cells transduced with cMYC-EGFP-DCD19 or cMYC-EGFP-DCD23 and their respective P2A controls were treated with 1  $\mu$ g/ml DOX for 24h to induce expression of the fusion constructs, followed by treatment with 10  $\mu$ M CC-220 in the absence or presence 0.5  $\mu$ M MG132 and 0.1  $\mu$ M MLN4924 for 4h. A cMYC antibody was used to probe for the levels of the fusion constructs by immunoblotting. Images are representative of 3 independent biological experiments (n=3).

### 3.1.13 cMYC degradation is reversible

A major advantage of chemical-biology tools over genetic methods is that they offer the ability to reverse the loss of a target protein in a one-step approach. This is a critical aspect in the application of chemical-biology tools as it allows causal relationships to be attributed between target protein degradation caused by a perturbant (e.g. IMiDs/CELMoDs) and observed phenotypic consequences (Kaelin, 2017). To examine whether IMiD/CELMoD washout would restore cMYC levels back to those of the untreated controls, HMEC cells expressing cMYC-EGFP-DCD23 or cMYC-EGFP(P2A)DCD23 were treated with 10  $\mu$ M CC-220 for 4h, followed by CC-220 washout and re-incubation of cells with DOX only for 24h or 48h. CC-220 treatment alone induced a robust and potent degradation of cMYC-EGFP-DCD23 which was not recapitulated in cells expressing cMYC-EGFP(P2A)DCD23 (**Figure 3.1. 19**). Loss of cMYC-EGFP-DCD23 resulted in a concomitant decrease in the protein levels of the cyclin-dependent kinase inhibitor p21, a previously described and well-characterised transcriptional repression target of cMYC (**Figure 3.1. 19**) (Gartel *et al.*, 2001). Following CC-220 removal, cMYC-EGFP-DCD23 levels were restored back to baseline after both 24h and 48h of drug washout, which also resulted in restoration of p21 repression (**Figure 3.1. 19**).



**Figure 3.1. 19 DCD-induced loss of target protein is reversible.** HMEC cells transduced with cMYC-EGFP-DCD23 or cMYC-EGFP(P2A)DCD23 were treated with 1  $\mu\text{g/ml}$  DOX for 24h to induce expression of the fusion constructs, followed by treatment with 10  $\mu\text{M}$  CC-220 for 4h. Subsequently, CC-220 was washed out and cells were collected and reseeded in a new cell culturing plate and re-treated with 1  $\mu\text{g/ml}$  DOX for either 24h or 48h. Protein lysates were analysed by immunoblotting and probed with cMYC and p21 antibodies. Images are representative of 3 independent biological experiments (n=3).

### **3.1.14 Interim Discussion Results Subchapter 3.1**

In this results subchapter, a set of twenty-six DCD motifs were designed based on the protein sequences of known CRBN neosubstrates and assessed in both a cellular degradation assay and an in vitro TR-FRET assay, with DCD2, DCD19 and DCD23 identified to induce a potent and robust degradation when fused to cMYC-EGFP, following IMiD/CELMoD treatment. From those three DCDs, DCD2 was eliminated as a suitable degron for the development of a novel chemical-biology tool since it only induced target protein degradation in the presence of CC-885, which was found to be toxic to the cells. DCD19 and DCD23 paired with CC-220 were selected for further characterisation and demonstrated to mediate a reversible degradation of cMYC-EGFP in both a time- and dose-dependent manner. Interestingly, compared to the previously described Aiolos ZF2-based degron, both DCD19 and DCD23 induced a more potent degradation of the target protein (Koduri *et al.*, 2019). Since both DCDs contain additional Ikaros ZF domains to the core ZF2 harbouring the  $\beta$ -hairpin loop with the glycine at the apex, this demonstrates that the observed degradation differences are due to the presence of these additional ZF domains. Moreover, this signifies that these ZF domains could form interactions with residues on the surface of CRBN leading to the stabilization of the tertiary conformation of the target protein-DCD – IMiD/CELMoD – CRBN complex, promoting target protein ubiquitination and subsequently proteasomal degradation. It should also be noted that, even though DCDs solely based on the core ZF domain containing the key  $\beta$ -hairpin loop induced partial degradation of the target protein, for non-ZF neosubstrates

GSPT1 and CK1 $\alpha$  the minimal  $\beta$ -hairpin loop was insufficient in degrading the target protein following IMiD/CELMoD treatment. This highlights that additional residues in the tertiary structure of these neosubstrates flanking the  $\beta$ -hairpin loop are critical for degradation by forming interactions with residues on the surface of CRBN. Furthermore, a disconnect between the in vitro TR-FRET assay and the cellular degradation assay was observed; for instance, CELMoDs such as CC-885 and CC-220 which induced a potent degradation of cMYC-EGFP-DCD19 in cells were not identified to mediate a more potent binding of DCD19 in the in vitro TR-FRET assay. This highlights the necessity for a wide range of orthogonal approaches for selecting the optimal tag and IMiD/CELMoD agent and validates the approach undertaken in defining an optimal DCD; with high binding potency and efficient target protein degradation. In summary, the data here support the use of both DCD19 and DCD23 along with the potent CELMoD CC-220 for the development of a novel chemical-biology tool, which was named iTAG (Inducible and TArgeted protein deGradation).

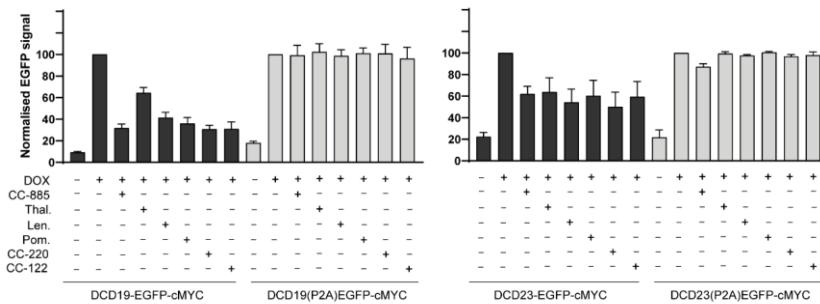


***Subchapter 3.2 : The iTAG system is a versatile tool for studying target protein function***

### ***3.2.1 Degradation of N-terminally tagged cMYC following IMiD/CELMoD treatment allows to study functional consequences of nuclear target protein loss***

One of the key aspects of an optimal DCD motif is its ability to induce degradation when fused to either target protein terminus, following treatment with IMiDs/CELMoDs. To further characterise the newly developed iTAG system and demonstrate its potential to selectively degrade N-terminally tagged proteins, DCD19, DCD23 and their corresponding P2A controls were fused at the N-terminus of EGFP-cMYC fusion construct and commercially cloned in the same lentiviral backbone vector used for assessing the functionality of DCD1 to DCD26. These vectors were subsequently used to generate HMEC cells stably expressing the fusion constructs. Treatment with 10  $\mu$ M of various IMiDs/CELMoDs for 4h led to a potent and robust decrease in EGFP fluorescence signal in HMEC cells expressing DCD19/ DCD23-EGFP-cMYC protein, but not in cells expressing their respective P2A controls (**Figure 3.2. 1**). Maximum loss of EGFP signal was achieved following treatment with 10  $\mu$ M CC-220, in cells expressing DCD19-EGFP-cMYC (residual signal: 30.78%) or DCD23-EGFP-cMYC (50.06%) (**Figure 3.2. 1**). Interestingly, N-terminal fusions of DCD19 led to a more potent degradation of target protein compared to DCD23, which is in contrast with the degradation assays of the cMYC-EGFP C-terminal fusions (**Figure 3.1. 9; Figure 3.1. 12**).

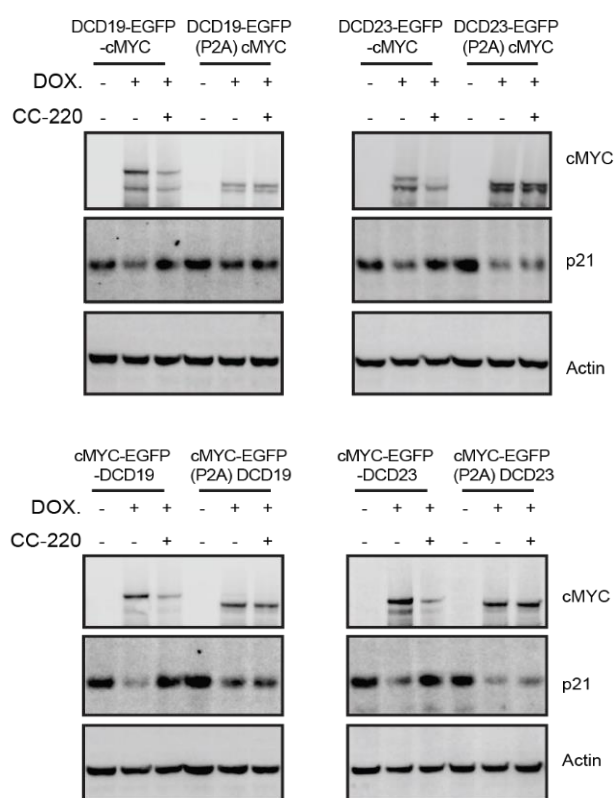
In order, to demonstrate that fusion of the DCD degron motifs to either the N- or C- terminus of cMYC does not abrogate its cellular functions, the expression levels of the cyclin-dependent kinase inhibitor p21 were selected as a read-out



**Figure 3.2.1 Fusion of DCD19 or DCD23 to the N-terminus of cMYC induced its degradation following IMiD/CELMoD treatment.** HMEC cells transduced with DCD19-EGFP-cMYC, DCD23-EGFP-cMYC and their respective P2A controls were treated with 1  $\mu$ g/ml DOX for 24h to stimulate expression of the fusion constructs, followed by treatment with 10  $\mu$ M of various IMiDs/CELMoDs for 4h. Median EGFP intensity was measured in each condition using flow cytometry. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

for DCD-tagged cMYC activity. cMYC has been previously shown to bind to the promoter region of p21 and induce its transcriptional repression (Gartel *et al.*, 2001). Treatment with DOX for 24h of HMEC cells transduced with either DCD19/ DCD23-EGFP-cMYC, cMYC-EGFP-DCD19/ DCD23 or their corresponding P2A controls induced cMYC expression and a concomitant decrease in p21 levels (**Figure 3.2. 2**). Subsequent treatment with 10  $\mu$ M of CC-220 for 16h induced the degradation of both DCD19/ DCD23-EGFP-cMYC and cMYC-EGFP-DCD19/ DCD23, followed by a restoration of p21 levels, thereby demonstrating that the iTAG system can be used to study the functional consequences following loss of a nuclear target protein. On the contrary, for the HMEC cells expressing the respective P2A controls, CC-220 did not induce degradation of cMYC, effectively preventing the restoration of p21 levels (**Figure 3.2. 2**). This result emphasizes

the relevance of the P2A ribosomal skipping sequences as controls for any potential off-target cellular effects arising from the activity of IMiDs/CELMOs.

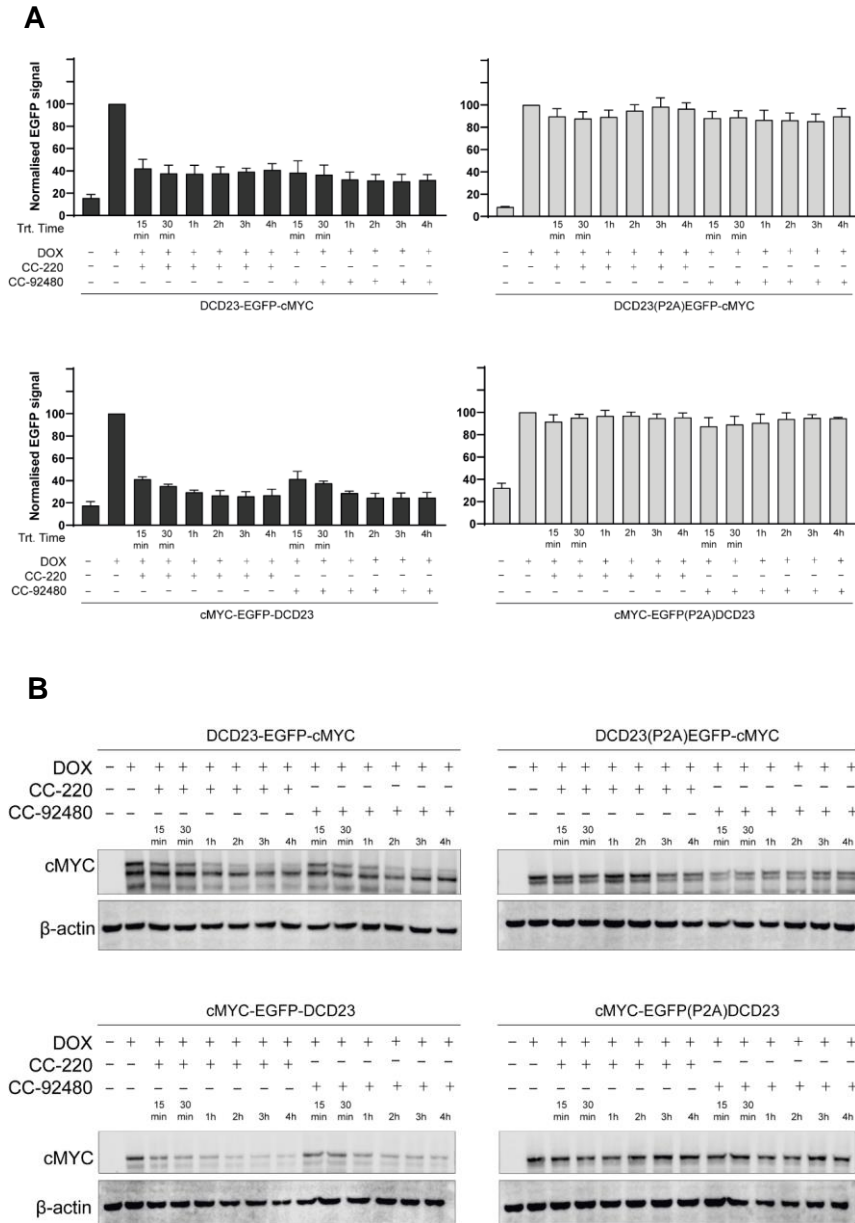


**Figure 3.2. 2 Functional validation of cMYC degradation following IMiD/CELMO treatment.**

HMEC cells transduced with DCD19/ DCD23-EGFP-cMYC, cMYC-EGFP-DCD19/ DCD23 and their respective P2A controls were treated with 1  $\mu$ g/ml DOX for 24h to stimulate expression of the fusion constructs, followed by treatment with 10  $\mu$ M of CC-220 for 16h. Protein lysates were collected and probed for cMYC and p21 using Western blotting. Images are representative of 3 independent biological experiments (n=3).

### **3.2.2 CC-220 and CC-92480 exhibit a similar degradation profile of iTAG-tagged cMYC**

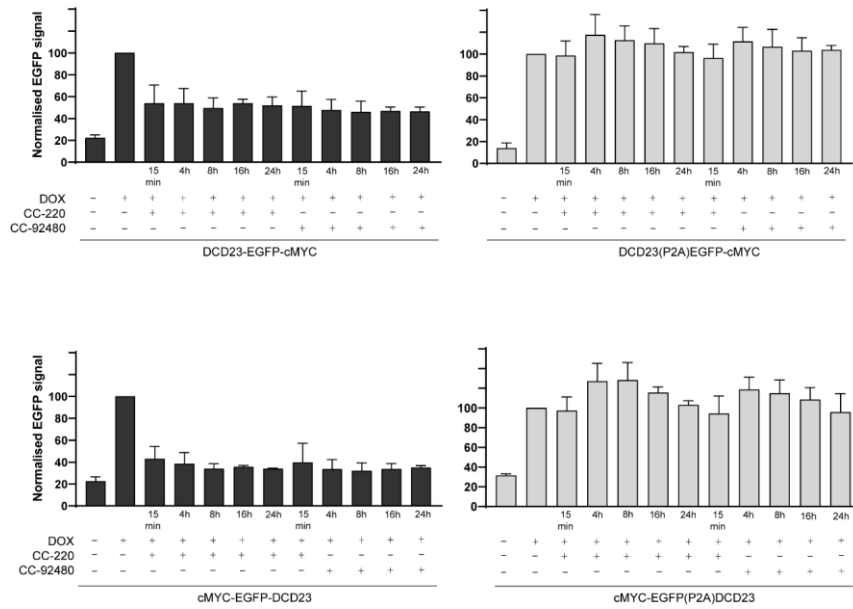
A recently developed CELMoD, namely CC-92480, has been characterised by a rapid and enhanced degradation efficiency of both Ikaros and Aiolos when compared to other IMiDs, such as lenalidomide and pomalidomide (Hansen *et al.*, 2020). To examine whether CC-92480 induces a stronger target protein degradation compared to CC-220, a time-course assay was performed. HMEC cells stably expressing DCD23-EGFP-cMYC, cMYC-EGFP-DCD23 and their respective P2A controls were treated with 10  $\mu$ M of either CC-220 or CC-92480 in a set of drug treatment time-points spanning 15min to 4h. HMEC cells were harvested at different time-points and subsequently analysed by both flow cytometry and immunoblotting. Flow cytometry analysis showed within 15min of either CC-220 or CC-92480 treatment a comparable reduction of EGFP fluorescence signal in HMEC cells cell expressing DCD23-EGFP-cMYC (residual signal: CC-220, 42.15%; CC-92480, 38.45%) and cMYC-EGFP-DCD23 (CC-220, 41.14%; CC-92480 41.49%) (**Figure 3.2. 3 A**). Maximum loss of EGFP signal was achieved at 2h of either CC-220 or CC-92480 treatment in cells expressing DCD23-EGFP-cMYC (residual signal: CC-220, 37.88%; CC-92480, 31.4%) and cMYC-EGFP-DCD23 (CC-220, 26.62%; CC-92480, 24.64%) (**Figure 3.2. 3 A**). CC-220 or CC-92480 treatment of HMEC cells expressing the corresponding P2A controls did not induce loss of cMYC fusion protein, as expected. In addition, immunoblot analysis of cellular lysates also confirmed the decrease in protein levels of both DCD23-EGFP-cMYC and cMYC-EGFP-



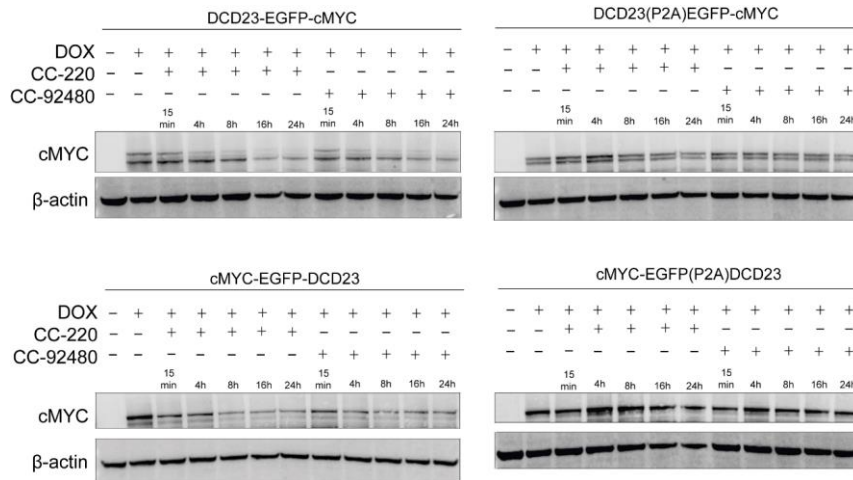
respective P2A controls were treated with 1 µg/ml DOX for 24h to stimulate expression of the fusion constructs. Subsequently cells were subjected to 10 µM CC-220 or CC-92480 treatment in a time-course assay. Loss of target protein was validated via **(A)** measuring EGFP median intensity by flow cytometry and **(B)** immunoblotting for the cMYC protein. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

DCD23, as early as 15min following CC-220 or CC-92480 treatment, but not of their corresponding P2A controls **(Figure 3.2. 3 B)**. Even though both CC-220 and CC-92480 exhibited similar efficiencies at inducing the degradation of iTAG-tagged cMYC at early time-points, loss of target protein remained partial. At the time, it was speculated that prolonged treatment with either CC-220 or CC-92480 could induce the complete loss of cMYC target protein and to assess this, a similar time-course assay to the one described above was performed, however, spanning a set of drug treatment time-points from 15min to 24h. Flow cytometry analysis showed that even 24h of 10 µM CC-220 or CC-92480 treatment, induced a comparable reduction of EGFP fluorescence signal in HMEC cells expressing DCD23-EGFP-cMYC (residual signal: CC-220, 51.93%; CC-92480, 46.46%) and cMYC-EGFP-DCD23 (CC-220, 34.04%; CC-92480, 34.92%), disproving this hypothesis **(Figure 3.2. 4 A)**. As expected, cells expressing the corresponding P2A controls had EGFP levels similar to vehicle-treated control cells for both N- and C-terminal fusions. Furthermore, immunoblotting of cellular lysates also confirmed that the degradation cMYC fusion constructs at 24h treatment with 10 µM CC-220 or CC-92480 is comparable but incomplete **(Figure 3.2. 4 B)**. Collectively, the data from both time-course assays supported the use of either CC-220 or CC-92480 as the optimal IMiD/CELMoD partners to DCD23-tagged target proteins.

**A**



**B**



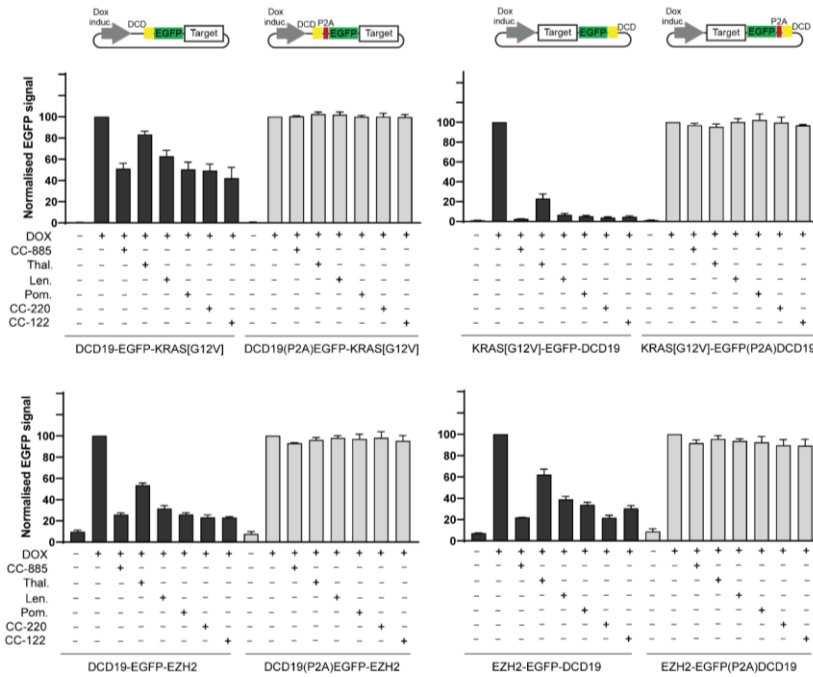
**Figure 3.2.4 Evaluation of the efficiency of cMYC degradation with prolonged CC-220 or CC-92480 treatment of cells. (A-B) HMEC cells transduced with DCD23-EGFP-cMYC, cMYC-**



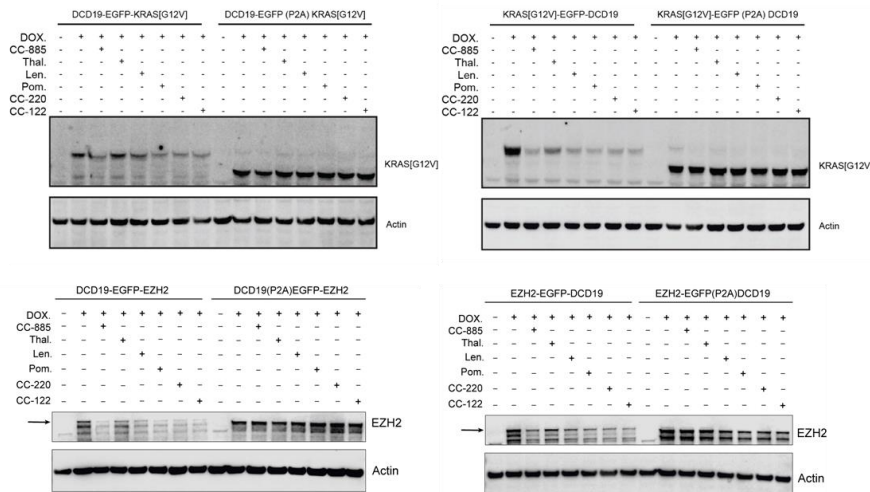
EGFP-DCD23 and their respective P2A controls were treated with 1 µg/ml DOX for 24h to stimulate expression of the fusion constructs. Subsequently cells were subjected to 10 µM CC-220 or CC-92480 treatment in a time-course assay. Loss of target protein was validated via (A) measuring EGFP median intensity by flow cytometry and (B) immunoblotting for the cMYC protein. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

### ***3.2.3 iTAG is a versatile tool to degrade both cytoplasmic and nuclear proteins***

In order to generalise the newly developed iTAG system as a versatile chemical-biology tool used to mediate the selective degradation of target proteins localised in various cellular compartments (cytoplasm and nucleus), the ability of both DCD19 and DCD23 to induce the degradation of other targets besides cMYC was assessed. The oncogenic version of the small GTPase KRAS, namely KRAS[G12V], harbouring a glycine to valine mutation at position 12, was selected as a relevant cytoplasmic target and the histone-lysine N-methyltransferase enzyme EZH2 (Enhancer of zeste homolog 2), which constitutes the functional enzymatic component of the polycomb repressive complex 2 (PRC2), was selected besides cMYC as another nuclear target protein. For the initial characterisation, DCD19 and its P2A control were selected to be fused along with EGFP to both termini of KRAS[G12V] and EZH2. These fusion constructs were commercially cloned in the same lentiviral backbone vector used for the cMYC studies and subsequently used to generate HMEC cells stably expressing them following DOX addition.



**Figure 3.2. 5 Functional evaluation of DCD19 fused to either terminus of KRAS[G12V] and EZH2.** HMEC cells transduced with DCD19-EGFP-KRAS[G12V], DCD19-EGFP-EZH2, KRAS[G12V]-EGFP-DCD19, EZH2-EGFP-DCD19 or their corresponding P2A controls were treated with 1 µg/ml DOX for 24h to stimulate expression of the fusion constructs, followed by treatment with 10 µM of various IMiDs/CELMoDs for 4h. Median EGFP intensity was measured in each condition using flow cytometry. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).



**Figure 3.2. 6 Fusion of DCD19 to either terminus of KRAS[G12V] and EZH2 induced their degradation following IMiD/CELMoD treatment.** HMEC cells transduced with DCD19-EGFP-KRAS[G12V], DCD19-EGFP-EZH2, KRAS[G12V]-EGFP-DCD19, EZH2-EGFP-DCD19 or their corresponding P2A controls were treated with 1  $\mu$ g/ml DOX for 24h to induce expression of the fusion constructs, followed by treatment with 10  $\mu$ M of indicated IMiDs/CELMoDs for 4h. KRAS[G12V] and EZH2 antibodies were used to probe for the levels of the fusion constructs by Western blotting. Images are representative of 3 independent biological experiments (n=3).

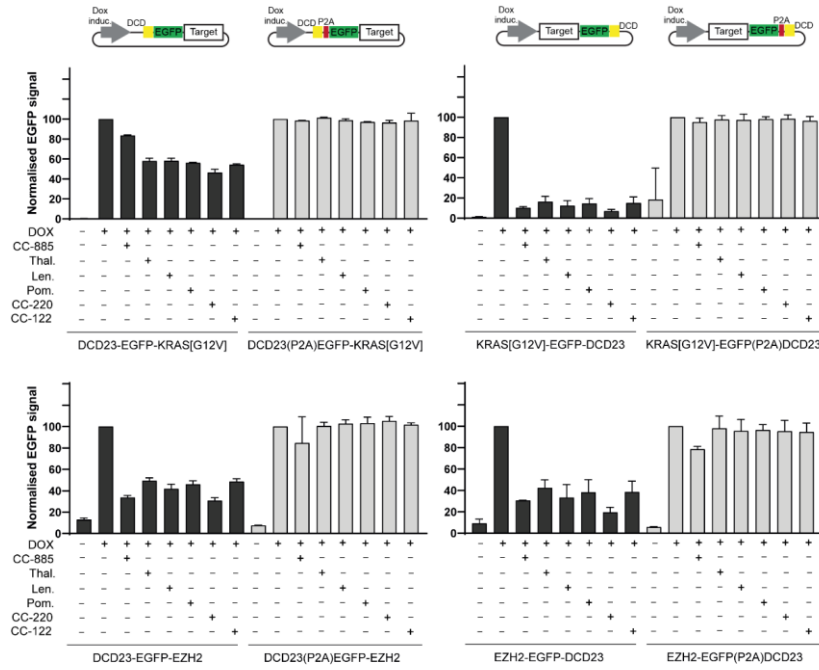
Subsequently, the newly generated HMEC cells expressing these fusion constructs were subjected in a flow cytometry assay to treatment with 10  $\mu$ M of various IMiDs/CELMoDs for 4h, to assess the degradation efficiency of DCD19. All IMiDs/CELMoDs tested induced a robust loss of EGFP fluorescent signal of both DCD19-tagged KRAS[G12V] and EZH2 target proteins (**Figure 3.2. 5**). Both CC-220 and CC-122 were identified as the most potent IMiDs/CELMoDs, with CC-220 inducing a potent loss of EGFP signal in HMEC cells expressing DCD19-

EGFP-KRAS[G12V] (residual signal: 47.45%), KRAS[G12V]-EGFP-DCD19 (3.83%), DCD19-EGFP-EZH2 (23.17%) and EZH2-EGFP-DCD19 (21.64%) (**Figure 3.2. 5**). The corresponding P2A controls for both KRAS[G12V] and EZH2 exhibited no significant loss of EGFP fluorescence signal. These flow cytometry observations were also validated by immunoblotting of extracts of HMEC cells expressing the fusion constructs, further demonstrating the versatility of the iTAG system (**Figure 3.2. 6**).

### ***3.2.4 Potent degradation of DCD23-tagged KRAS[G12V] and EZH2 following IMiD/CELMoD treatment***

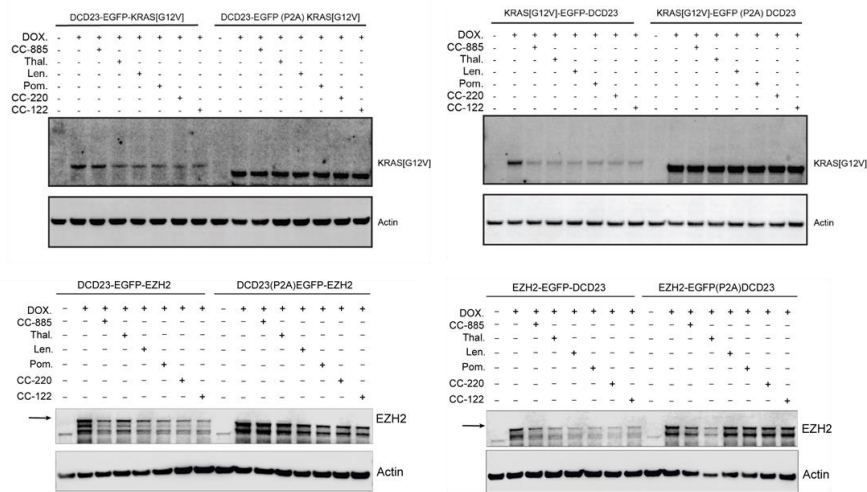
Having demonstrated that DCD19 (131 residues) can mediate the degradation of other target proteins besides cMYC, the shorter in length DCD23 (60 residues) was also assessed for its ability to degrade KRAS[G12V] and EZH2 when fused to either terminus. Lentiviral backbone vectors harbouring the fusion constructs under a DOX-inducible promoter were generated in-house, as described in detail in the Materials and Methods section (**Figure 2. 5 A-D**). Subsequently, these lentiviral vectors were used to generate HMEC cells stably expressing the fusion constructs and were also treated with various IMiDs/CELMoDs for 4h to assess the relative degradation efficiency of each compound by flow cytometry. All IMiDs/CELMoDs tested induced a robust loss of EGFP fluorescent signal of both DCD23-tagged KRAS[G12V] and EZH2, comparable to their DCD19-tagged counterparts (**Figure 3.2. 5; Figure 3.2. 7**).

Maximum loss of EGFP signal at 4h was observed following treatment with 10  $\mu$ M CC-220 in HMEC expressing DCD23-EGFP-KRAS[G12V] (residual signal: 46.41%), KRAS[G12V]-EGFP-DCD23 (7.05%), DCD23-EGFP-EZH2 (30.69%) and EZH2-EGFP-DCD23 (19.38%) (**Figure 3.2. 7**). The loss of EGFP signal observed by flow cytometry for this set of constructs was also validated by immunoblotting (**Figure 3.2. 8**). Similarly to DCD19-tagged KRAS[G12V] and EZH2, the corresponding P2A controls exhibited no significant loss of EGFP fluorescence signal (**Figure 3.2. 6; Figure 3.2. 8**). Interestingly, C-terminal fusion of DCD23 to both KRAS[G12V] and EZH2 induced a more effective loss of these target proteins compared to N-terminal fusions of DCD23, possibly due to structural conformation differences at each terminus dictating efficient binding to the IMiD/CELMoD-bound CRL4<sup>CRBN</sup> E3 ubiquitin ligase. In addition, N-terminally DCD19-tagged KRAS[G12V] compared to DCD23-tagged KRAS[G12V] was more efficiently degraded following treatment with 10  $\mu$ M CC-220 (**Figure 3.2. 5; Figure 3.2. 7**). This highlights the relevant importance of having more than one DCD tag variant to select for different target proteins, based on degradation efficiency. Collectively, the data on both DCD19- and DCD23-tagged KRAS[G12V] and EZH2 supported the use of both DCDs for further development of the iTAG system and demonstrated that it constitutes a modular, versatile, and efficient chemical-biology tool for acutely degrading target proteins.



**Figure 3.2. 7 Functional evaluation of DCD23 fused to either terminus of KRAS[G12V] and EZH2.** HMEC cells transduced with DCD23-EGFP-KRAS[G12V], DCD23-EGFP-EZH2, KRAS[G12V]-EGFP-DCD23, EZH2-EGFP-DCD23 or their corresponding P2A controls were treated with 1 µg/ml DOX for 24h to stimulate expression of the fusion constructs, followed by treatment with 10 µM of various IMiDs/CELMoDs for 4h. Median EGFP intensity was measured in each condition using flow cytometry. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

## Results



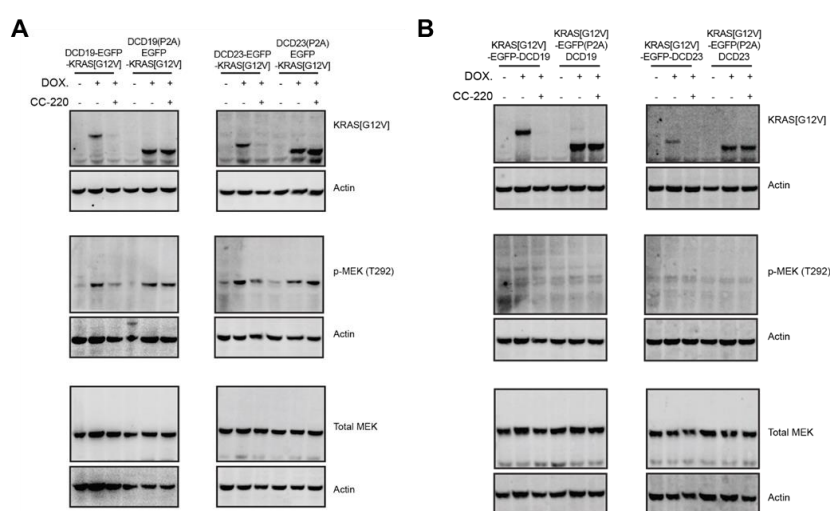
**Figure 3.2. 8 Fusion of DCD23 to either terminus of KRAS[G12V] and EZH2 induced their degradation following IMiD/CELMoD treatment.** HMEC cells transduced with DCD23-EGFP-KRAS[G12V], DCD23-EGFP-EZH2, KRAS[G12V]-EGFP-DCD23, EZH2-EGFP-DCD23 or their corresponding P2A controls were treated with 1  $\mu$ g/ml DOX for 24h to induce expression of the fusion constructs, followed by treatment with 10  $\mu$ M of indicated IMiDs/CELMoDs for 4h. KRAS[G12V] and EZH2 antibodies were used to probe for the levels of the fusion constructs by Western blotting. Images are representative of 3 independent biological experiments (n=3).

### **3.2.5 Degradation of N-terminally tagged KRAS[G12V] following IMiD/CELMoD treatment allows to study the functional consequences of cytoplasmic target protein loss**

In order to study the functional consequences following degradation of iTAG-tagged KRAS[G12V], the downstream activation of MAPK (mitogen-activated protein kinase)/ ERK (extracellular signal-regulated kinase) signalling pathway components was assessed. Mutant KRAS[G12V] is persistently active and is known to induce activation of the proto-oncogene serine/threonine-protein kinase RAF, which in turn, can phosphorylate MEK1/2 (mitogen-activated protein kinase kinase 1/2) that subsequently phosphorylates further downstream components, eventually leading to activation of the MAPK/ERK signalling pathway (Waters and Der, 2018). It has been previously shown that in a negative feedback loop, activated ERK1 phosphorylates activated MEK1 at position T292 (Eblen *et al.*, 2004). Therefore, as a downstream marker for KRAS[G12V] activation, the phosphorylation levels of MEK1 (T292) were examined in HMEC cells stably expressing DCD19/ DCD23-EGFP-KRAS[G12V], KRAS[G12V]-EGFP-DCD19/ DCD23 and their respective P2A controls. Interestingly, DCD19/ DCD23-EGFP-KRAS[G12V] expression promoted MEK1 phosphorylation, whilst this was not observed in C-terminal fusion constructs, possibly due to interference of farnesylation of the KRAS[G12V] C-terminal CAAX motif by the fused iTAG motifs, thus hindering its activation (Choy *et al.*, 1999) **(Figure 3.2. 9)**. Treatment of cells with 10  $\mu$ M CC-220 for 48h resulted in degradation of both N- and C-terminally iTAG-tagged KRAS[G12V] followed by a concomitant decrease in MEK1 phosphorylation levels of DCD19/ DCD23-EGFP-



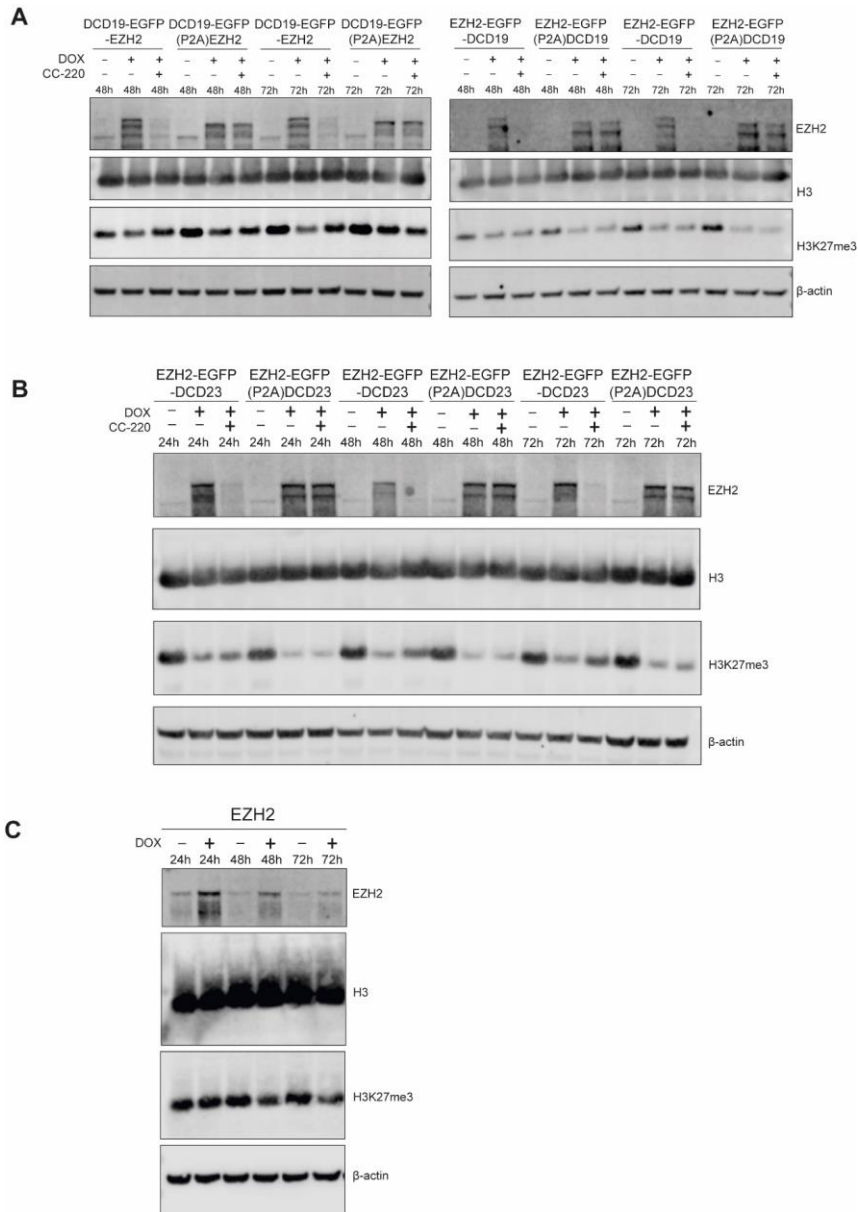
KRAS[G12V] expressing cells (**Figure 3.2. 9**). As anticipated, no degradation of the P2A controls for both N- and C-terminal fusions was observed by immunoblotting, with MEK1 phosphorylation levels of CC-220-treated DCD19/DCD23(P2A)EGFP-KRAS[G12V] remaining elevated as the untreated group (**Figure 3.2. 9**). Overall, these data demonstrated the versatility and applicability of the iTAG system to study the functional consequences of a cytoplasmic target protein loss following IMiD/CELMoD-mediated degradation.



**Figure 3.2. 9 Functional validation of KRAS[G12V] degradation following IMiD/CELMoD treatment.** (A-B) HMEC cells transduced with DCD19/ DCD23-EGFP-KRAS[G12V], KRAS[G12V]-EGFP-DCD19/ DCD23 and their respective P2A controls were treated with 1  $\mu\text{g/ml}$  DOX for 24h to stimulate expression of the fusion constructs, followed by treatment with 10  $\mu\text{M}$  of CC-220 for 48h. Protein lysates were collected and probed for KRAS[G12V], phospho-MEK1(T292) and MEK1 using Western blotting. Images are representative of 3 independent biological experiments (n=3).

### **3.2.6 Expression of iTAG-tagged EZH2 causes a decrease of H3K27me3 levels**

The histone-lysine N-methyltransferase enzyme EZH2 acts as the catalytic component of PRC2, a complex which is comprised in mammalian cells of the core subunits EED (Embryonic Ectoderm Development) and SUZ12 (Suppressor Of Zeste 12 Protein Homolog) (Laugesen, Højfeldt and Helin, 2019). Along with a histone-binding protein, RBBP4 (RB Binding Protein 4) or RBBP7 (RB Binding Protein 7), the PRC2 core complex forms distinct subcomplexes that function by mono-, di-, and tri-methylating Lys27 on histone H3 (H3K27me1, H3K27me2, and H3K27me3) (Cao *et al.*, 2002; Czermin *et al.*, 2002; Kuzmichev *et al.*, 2002; Müller *et al.*, 2002). Previous studies probing the function of EZH2 focus on H3K27me3 which is considered a hallmark of PRC2 activity (Müller *et al.*, 2002; Pengelly *et al.*, 2013). Therefore, in order to study the functional consequences following degradation of iTAG-tagged EZH2, the downstream trimethylation of H3K27 was assessed. HMEC cells stably expressing DCD19-EGFP-EZH2, EZH2-EGFP-DCD19/ DCD23 and their respective P2A controls were treated with DOX for either 24h, 48h or 72h to induce expression of the fusion proteins. Longer timepoints of EZH2 expression were selected as it has been previously shown that PRC2 is more efficient at catalysing the initial methylation of H3K27 to generate H3K27me1 and H3K27me2, whilst conversion to the trimethylated form requires more time (Sneeringer *et al.*, 2010; Zee *et al.*, 2012). Interestingly, expression of both N- and C-terminally iTAG-tagged EZH2 lead to a strong and robust decrease in the trimethylation levels of H3K27,



**Figure 3.2. 10 Functional validation of EZH2 degradation following IMiD/CELMoD treatment. (A-B)** HMEC cells transduced with DCD19-EGFP-EZH2, EZH2-EGFP-DCD19/DCD23 and their respective P2A controls were treated with 1 µg/ml DOX for 24h to stimulate

expression of the fusion constructs, followed by treatment with 10  $\mu$ M of CC-220 for 24h, 48h and 72h. Protein lysates were collected and probed for EZH2, total H3 and H3K27me3 using Western blotting. **(C)** HMEC cells transduced with EZH2 were treated with 1  $\mu$ g/ml DOX for 24h to stimulate expression of the fusion constructs, followed by treatment with 10  $\mu$ M of CC-220 for 24h, 48h and 72h. Protein lysates were collected and probed for EZH2, total H3 and H3K27me3 using Western blotting. Images are representative of 3 independent biological experiments (n=3).

contrary to previous observations **(Figure 3.2. 10 A-B)** (Müller *et al.*, 2002; Pengelly *et al.*, 2013). Treatment with 10  $\mu$ M CC-220 for either 24h, 48h or 72h partially restored the levels of H3K27me3 for cells expressing DCD19-EGFP-EZH2 and EZH2-EGFP-DCD19/ DCD23 to those of the non-induced controls, but not for their respective P2A controls **(Figure 3.2. 10 A-B)**. A possible explanation for this was that fusion of the iTAG motifs to either terminus of EZH2 affected its function by inhibiting the formation of the PRC2 complex or by blocking its catalytic activity. To test this hypothesis, lentiviral backbone vectors expressing untagged EZH2 under the control of a DOX-inducible promoter were generated and used to transduce HMEC cells. Details of the cloning procedure are presented in the Materials and Methods section **(Figure 2. 6 A-B)**. However, H3K27me3 levels were still decreased following expression of EZH2 in HMEC cells for 24h, 48h or 72h, demonstrating that the observed effects are not due to the fusion of the iTAG degron motifs to EZH2 but could be either an artifact of the exogenous overexpression of EZH2 or related to the function of EZH2 in this particular cell line **(Figure 3.2. 10 C)**.

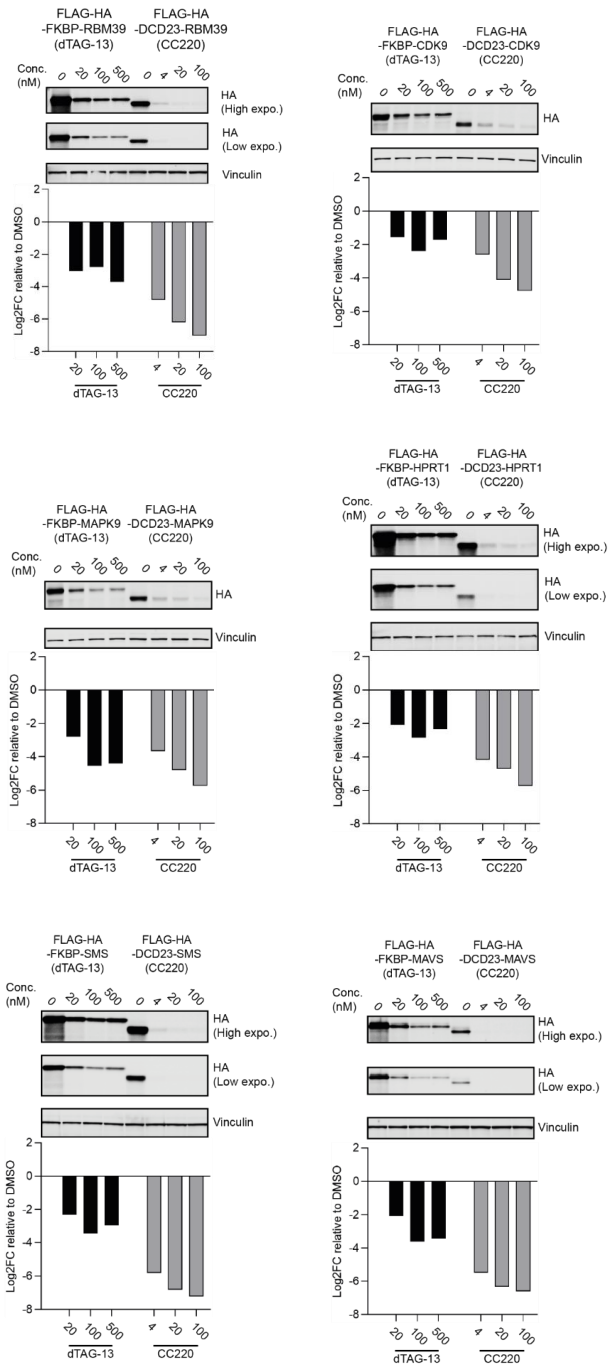
***Subchapter 3.3 : The iTAG system constitutes a modular tool for both in vitro and in vivo target protein evaluation***

### **3.3.1 iTAG-induced degradation of target proteins is more potent and efficient compared to the dTAG system**

One of the points to address, was to demonstrate the degradation of target proteins directly fused to the DCD motifs (i.e. without an EGFP sequence) following IMiD/CELMoD treatment, and to compare the degradation efficiency of the iTAG system to other well-established chemical-biology tools. Currently, the dTAG method developed by Nathanael Gray's and James Bradner's laboratories in 2018, constitutes a broadly adopted chemical-biology tool (**Figure 1. 15 A-D**) (Nabet *et al.*, 2018, 2020; Nabet, 2021). Details of how the dTAG system functions are provided in the introductory chapter of this thesis, but in brief, the mutant FKBP12<sup>F36V</sup> is fused to a target protein and degradation is induced following treatment with a PROTAC such as dTAG-7, dTAG-13, or dTAG<sup>V</sup>-1, which target the FKBP12<sup>F36V</sup>-tagged protein to either the CRL4<sup>CRBN</sup> or CRL2<sup>VHL</sup> E3 ubiquitin ligases (**Figure 1. 15 A-D**) (Nabet *et al.*, 2018, 2020). The dTAG system allows to study the functional consequences of acute target protein degradation both in vitro and in vivo and has been already successfully applied to several targets (Erb *et al.*, 2017; Huang *et al.*, 2017). Our collaborators (Trey Westbrook's laboratory, Baylor College of Medicine, Texas, USA), selected the following target proteins to compare the degradation efficiency of the iTAG system to the dTAG system: RBM39 (RNA-binding protein 39), CDK9 (cyclin-dependent kinase 9), MAPK9 (mitogen-activated protein kinase 9), HPRT1 (hypoxanthine phosphoribosyltransferase 1), SMS (Spermine synthase) and MAVS (mitochondrial antiviral-signalling protein). Targets were selected to cover a broad range of cellular localisations in order to allow for a more complete and

direct comparison in degradation efficiency between the iTAG and dTAG system. From the selected target proteins, RBM39, CDK9 and MAPK9 are present in the nucleus, HRPT1 and SMS in the cytoplasm and MAVS in mitochondria. The epitope tags FLAG (5' DYKDDDDK 3') and HA (human influenza hemagglutinin, 5' YPYDVPDYA 3') were also fused in tandem to the N-terminus of both FKBP12<sup>F36V</sup> and DCD23. The rationale behind this was twofold, first to be used as markers for target protein levels, and second, they have a significantly smaller molecular weight (~1 kDa) compared to EGFP (29.6 kDa). These target proteins were synthesised and cloned by our collaborators in a lentiviral backbone vector with either FLAG-HA-FKBP12<sup>F36V</sup> or FLAG-HA-DCD23 directly fused at their N-terminus. Subsequently, SUM159 mesenchymal triple-negative breast cancer cells stably expressing these fusion constructs were generated. SUM159 cells were treated with various concentrations of either dTAG13 or CC-220 for 24h with loss of target protein determined by immunoblotting. Both FKBP12<sup>F36V</sup> and DCD23 induced the degradation of target proteins, with FLAG-HA-DCD23 fusions inducing a more potent and robust degradation of target proteins, even at low nM concentrations of CC-220 (**Figure 3.3. 1; Table 3.3. 1**). Furthermore, the fact that DCD23 maintained its ability to induce target protein degradation in the absence of EGFP, demonstrated that degradation was not dependent on the presence of EGFP. In addition, fusion of FLAG-HA to the N-terminus of DCD23 did not affect its functionality, further demonstrating that DCD23-mediated degradation is not dependent on its position within the fusion target protein. Interestingly, the degradation efficiency of FLAG-HA-FKBP12<sup>F36V</sup>-CDK9, FLAG-HA-FKBP12<sup>F36V</sup>-HPRT1, and FLAG-HA-FKBP12<sup>F36V</sup>-SMS was slightly reduced

## Results





**Figure 3.3. 1 The iTAG system induced a more potent and robust degradation of multiple target proteins compared to the dTAG system.** SUM159 cells transduced with various target proteins N-terminally tagged with either the dTAG degron motif FLAG-HA-FKBP12<sup>F36V</sup> or the iTAG degron motif FLAG-HA-DCD23 were subjected to dTAG13 or CC-220 treatment, respectively for 24h at the indicated doses. Protein lysates were probed for HA on immunoblots, as a marker for target protein levels. Band intensities were also quantified and represented as Log<sub>2</sub> Fold Change relative to DMSO. Data are representative of 3 independent biological experiments (n=3). Data for this figure were generated by our collaborators, Trey Westbrook's laboratory at Baylor College of Medicine.

at high concentrations (500 nM) of dTAG13 treatment, possibly due to the "hook effect" for which PROTACs are known to suffer from when administered at high concentrations (**Figure 3.3. 1; Table 3.3. 1**) (Bondeson *et al.*, 2018). Collectively, these results demonstrated that the iTAG system is a highly applicable tool for inducing the selective and potent degradation of differentially localised target proteins, and that it can be employed for studying the functional consequences following acute target protein loss.

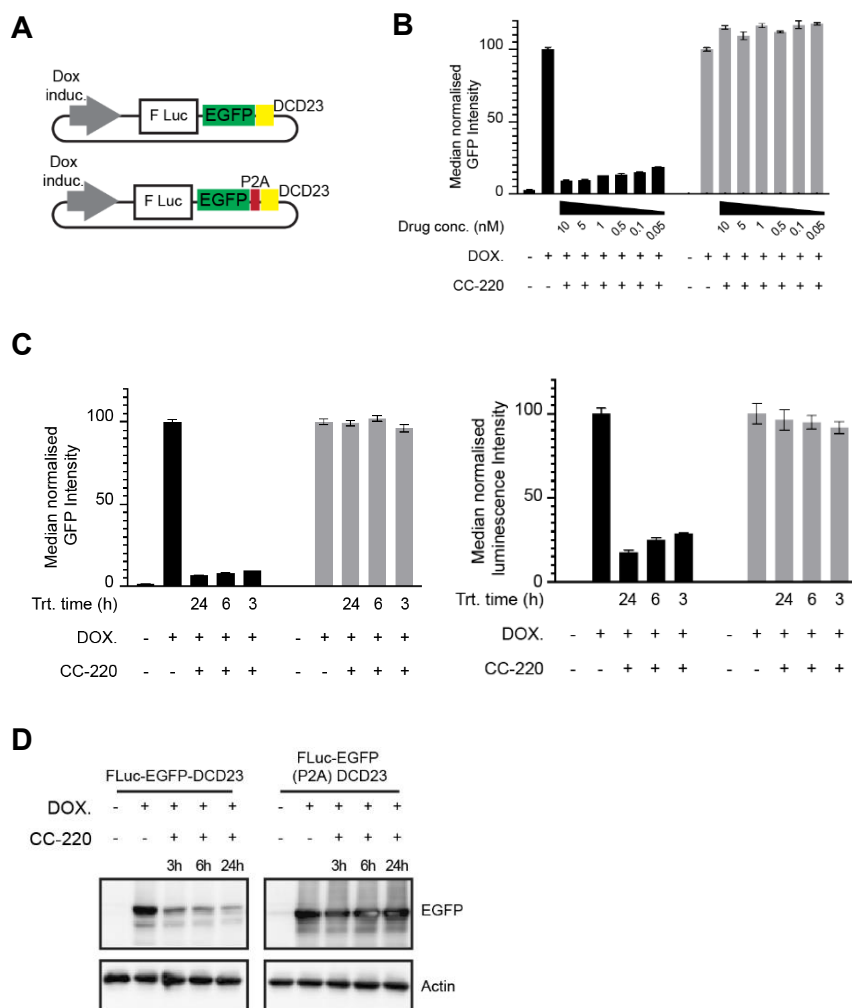
Drug Conc. (nM)	dTAG degradation (Log <sub>2</sub> FC relative to DMSO)			iTAG degradation (Log <sub>2</sub> FC relative to DMSO)		
	20	100	500	4	20	100
RBM39	-3.02	-2.78	-3.73	-4.81	-6.21	-7.01
CDK9	-1.54	-2.38	-1.71	-2.60	-4.13	-4.78
MAPK9	-2.80	-4.55	-4.40	-3.66	-4.79	-5.74
HPRT1	-2.08	-2.85	-2.34	-4.15	-4.69	-5.72
SMS	-2.30	-3.45	-2.96	-5.81	-6.81	-7.22
MAVS	-2.10	-3.62	-3.43	-5.47	-6.34	-6.61

**Table 3.3. 1 Summary table of quantified immunoblot band intensities representing the range of degradation in Log<sub>2</sub> Fold Change relative to DMSO using either the dTAG or iTAG system.** Data are representative of 3 independent biological experiments (n=3). Data for this table were generated by our collaborators, Trey Westbrook's laboratory at Baylor College of Medicine.

### ***3.3.2 In vivo degradation of an iTAG-tagged target protein***

A final point to address in developing iTAG as a tool to study protein function was to demonstrate its applicability in the in vivo setting. For the animal studies, Fluc (firefly luciferase) was selected as the target protein, as it would facilitate monitoring of its in vivo expression levels by bioluminescence imaging. Fluc was commercially synthesised fused in tandem to EGFP-DCD23 or EGFP(P2A)DCD23 at its C-terminus, cloned in the pLVX-TetOne lentiviral backbone expression vector under the control of a DOX-inducible promoter and used to generate MDA-MB-231 breast cancer cells stably expressing the fusion constructs (**Figure 3.3. 2 A**). Treatment with various CC-220 concentrations for 24h of the MDA-MB-231 cells showed a robust and potent degradation of Fluc-EGFP-DCD23, whilst no loss of target protein was observed in cells expressing Fluc-EGFP(P2A)DCD23 (**Figure 3.3. 2 B**). In addition, a time-course assay of 10  $\mu$ M CC-220 treatment for 3h, 6h and 24h was performed to further characterise the degradation of Fluc-EGFP-DCD23. Fluc-EGFP-DCD23 degradation was confirmed by flow cytometry and bioluminescence measurements as well as immunoblotting within 3h of CC-220 treatment (**Figure 3.3. 2 C-D**).

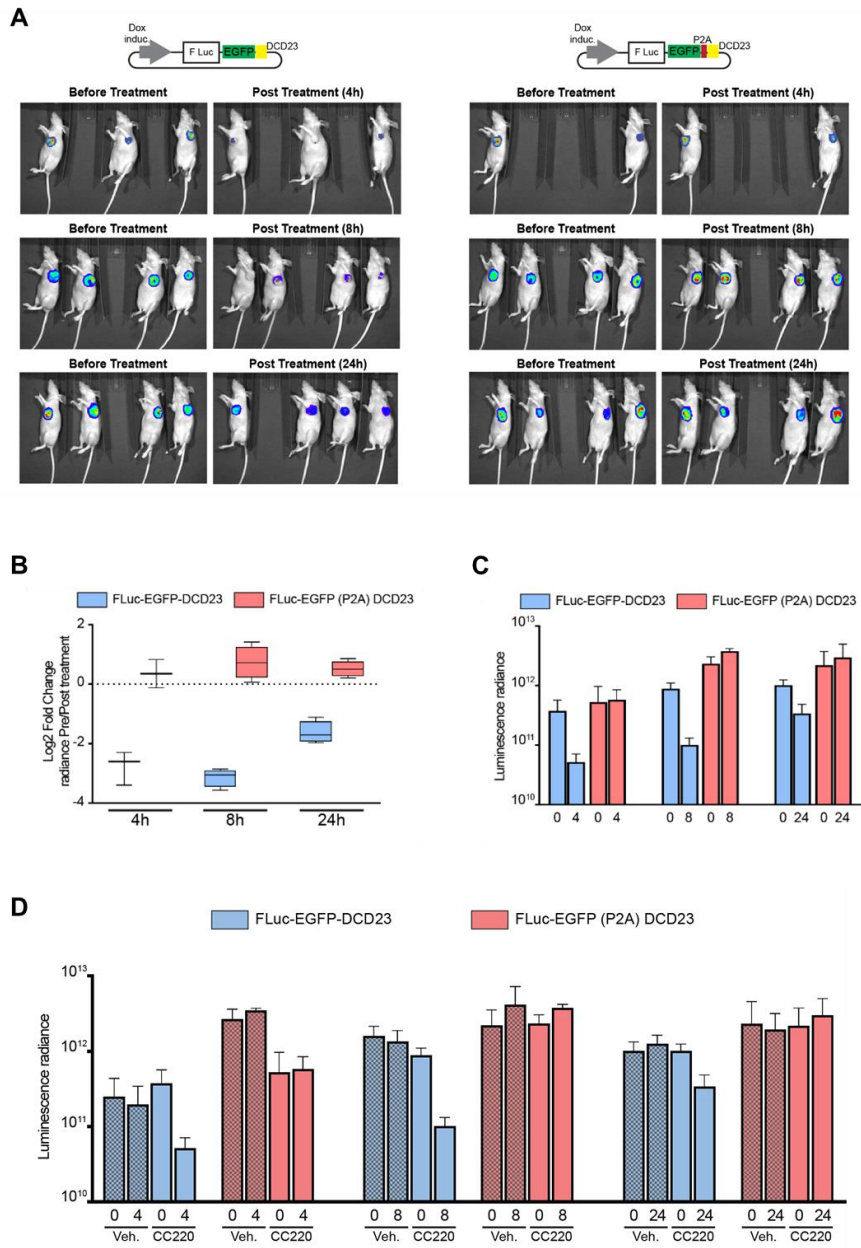
Having validated the degradation of iTAG-tagged Fluc in the in vitro setting, our collaborators (Trey Westbrook's laboratory, Baylor College of Medicine, Texas, USA) orthotopically injected MDA-MB-231 cells expressing either Fluc-EGFP-DCD23 or Fluc-EGFP(P2A)DCD23 in the fat mammary pads

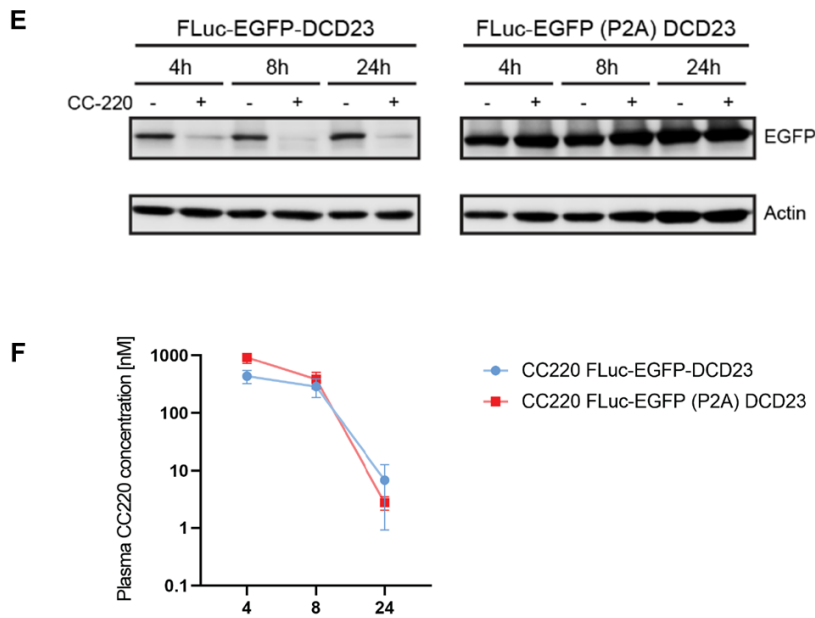


**Figure 3.3. 2 Evaluation of iTAG-mediated degradation of Fluc in MDA-MB-231 cells.** (A) Fluc-EGFP-DCD23 and Fluc-EGFP(P2A)DCD23 cloned in a lentiviral backbone vector under the control of a DOX-inducible promoter. (B) MDA-MB-231 cells transduced with Fluc-EGFP-DCD23 or Fluc-EGFP(P2A)DCD23 were treated with 1  $\mu\text{g}/\text{ml}$  DOX for 24h to stimulate expression of the fusion constructs. Subsequently, cells were subjected to CC-220 treatment at indicated drug concentrations for 24h. Fluc loss was validated via measuring EGFP median intensity by flow cytometry. (C) Time-course assay of MDA-MB-231 cells expressing Fluc-EGFP-DCD23 or Fluc-EGFP(P2A)DCD23. Loss of target protein following 10  $\mu\text{M}$  CC-220 treatment was assessed by

both measuring EGFP median intensity and bioluminescence intensity. **(D)** Validation of flow cytometry and bioluminescence observations of time-course assay by immunoblotting for EGFP. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3). Data for this figure were generated by me.

of immunocompromised mice. The cells were allowed a time-period for the formation of breast tumours of a size of 300-350 mm<sup>3</sup>, prior to inducing expression of the fusion constructs by administering DOX to the mice. In addition, luciferin was administered to the mice bearing either MDA-MB-231 Fluc-EGFP-DCD23 or Fluc-EGFP(P2A)DCD23 expressing tumours to induce bioluminescence and subsequently the mice were randomised to receive a vehicle or 30 mg/kg CC-220. Bioluminescence measurements were taken prior and post 4h, 8h or 24h treatment with the vehicle or CC-220 (**Figure 3.3. 3 A-D**). Bioluminescence intensity rapidly decreased by a Log<sub>2</sub> fold change of -2.76 and -3.13 at 4h and 8h, respectively (**Figure 3.3. 3 B**). Furthermore, tumours were extracted from culled mice and prepared for immunoblotting, which validated the degradation of Fluc-EGFP-DCD23 (**Figure 3.3. 3 E**). After 24h of CC-220 administration, residual plasma drug levels were in the nanomolar range, yet the bioluminescence signal was still decreased in the Fluc-EGFP-DCD23 tumours compared to pre-treatment levels (Log<sub>2</sub> fold change of -1.63) (**Figure 3.3. B-D, F**). This was also seen on Western blot with almost undetectable Fluc levels in tumours at 24h, suggesting that even nanomolar plasma concentration of CC220 is sufficient to maintain target degradation in vivo (**Figure 3.3. 3 E**). Collectively, the data from this study highlight the applicability of the iTAG system in the in vivo setting by demonstrating the acute and potent degradation of iTAG-tagged Fluc in mice.





**Figure 3.3. 3 The iTAG system induced the selective degradation of a target protein in vivo.**

(A) Bioluminescence images of orthotopic tumours formed at the fat mammary pads in immunocompromised mice from the injection of MDA-MB-231 cells expressing either Fluc-EGFP-DCD23 or Fluc-EGFP(P2A)DCD23. Tumour images of pre- and post-systemic administration of 30 mg/kg CC-220 by oral gavage are shown. (B, C) Quantification of bioluminescence signals from mice represented in either Log<sub>2</sub> fold change pre/post treatment in (B) or in absolute values (C). (D) Quantification of bioluminescence signals from mice represented in (A) and their corresponding vehicle treated mice. (E) Tumours from culled mice were collected, lysed, and probed by an EGFP antibody to examine Fluc levels following CC-220 treatment in vivo. (F) Quantification of CC-220 concentration at various time-points in the plasma of mice injected with MDA-MB-231 cells expressing iTAG-tagged Fluc or its respective P2A control. Data for this figure were generated by our collaborators, Trey Westbrook's laboratory at Baylor College of Medicine.

***Subchapter 3.4 : Engineered mutant DCD23 motifs mediate acute and potent degradation of target proteins***

### **3.4.1 Design and generation of mutant DCD23 degrons**

C2H2 ZFs are common DNA-binding motifs, with amino acids on their surface forming several contacts along the DNA major groove (Wolfe, Nekludova and Pabo, 1999). DNA recognition typically requires two to four tandemly arranged ZF domains with the  $\alpha$ -helical portion of each ZF fitting into the DNA major groove (Wolfe, Nekludova and Pabo, 1999). Amino acids such as lysine, arginine, and histidine in ZF domains are key in mediating DNA binding, with histidine residues also being implicated in stabilising the tertiary structure of ZF domains (Elrod-erickson *et al.*, 1996; Wolfe, Nekludova and Pabo, 1999). The transcription factor Ikaros is classified as a C2H2 ZF domain-containing protein and has been previously demonstrated to bind via its N-terminal ZF1-ZF4 domains at the promoter sites of genes regulating haematopoietic stem cell differentiation (Arpad and Georgopoulos, 1994). Furthermore, posttranslational modifications such as phosphorylation of residues S358 and S361 at its C-terminus by the spleen tyrosine kinase (SYK), have been also shown to augment its nuclear localization and sequence-specific DNA binding activity (Uckun *et al.*, 2012). Since the sequence of DCD23 is primarily based on the N-terminal ZF1, ZF2 and ZF3 domains of Ikaros, I aimed to examine whether fusion of this degraon motif to a target protein would affect its subcellular localisation, by causing the mislocalisation of the target protein to the nucleus. EGFP was selected as the target protein for this set of experiments, first because it has no bias for a specific subcellular compartment and secondly, as a fluorescent protein, it would allow for easy monitoring of both its expression and localisation (Seibel *et al.*, 2007).



**DCD23 WT**

Nter-PNVLMVH**KR**SHTGE**R**P**LQ**CEICGFT**C****E****R****K**GNLL**R**HI**K**LHSGE**K**PFKCHLCNYAC**RRR**DAL**C**ter

ZFP91 ZF2      Ikaros ZF2  
β-hairpin loop      α-helix

**DCD23mut1**

Nter-PNVLMVH**KR**SHTGE**R**P**LQ**CEICGFT**C****E****A**GNLL**N**HI**E**LHSGE**K**PFKCHLCNYAC**RRR**DAL**C**ter

ZFP91 ZF2      Ikaros ZF2  
β-hairpin loop      α-helix

**DCD23mut2**

Nter-PNVLMVH**KR**SHTGE**I**P**LQ**CEICGFT**C****E****A**GNLL**N**HI**E**LHSGE**K**PFKCHLCNYAC**RRR**DAL**C**ter

ZFP91 ZF2      Ikaros ZF2  
β-hairpin loop      α-helix

**DCD23mut3**

Nter-PNVLMVH**NE**SHTGE**I**P**LQ**CEICGFT**C****E****A**GNLL**N**HI**E**LHSGE**I**PFKCHLCNYAC**RRR**DAL**C**ter

ZFP91 ZF2      Ikaros ZF2  
β-hairpin loop      α-helix

**Key:**

**big black font:** important residues for CRBN interaction

**big red font:** crucial residues for DNA interaction

**small red font:** additional residues mediating DNA interaction

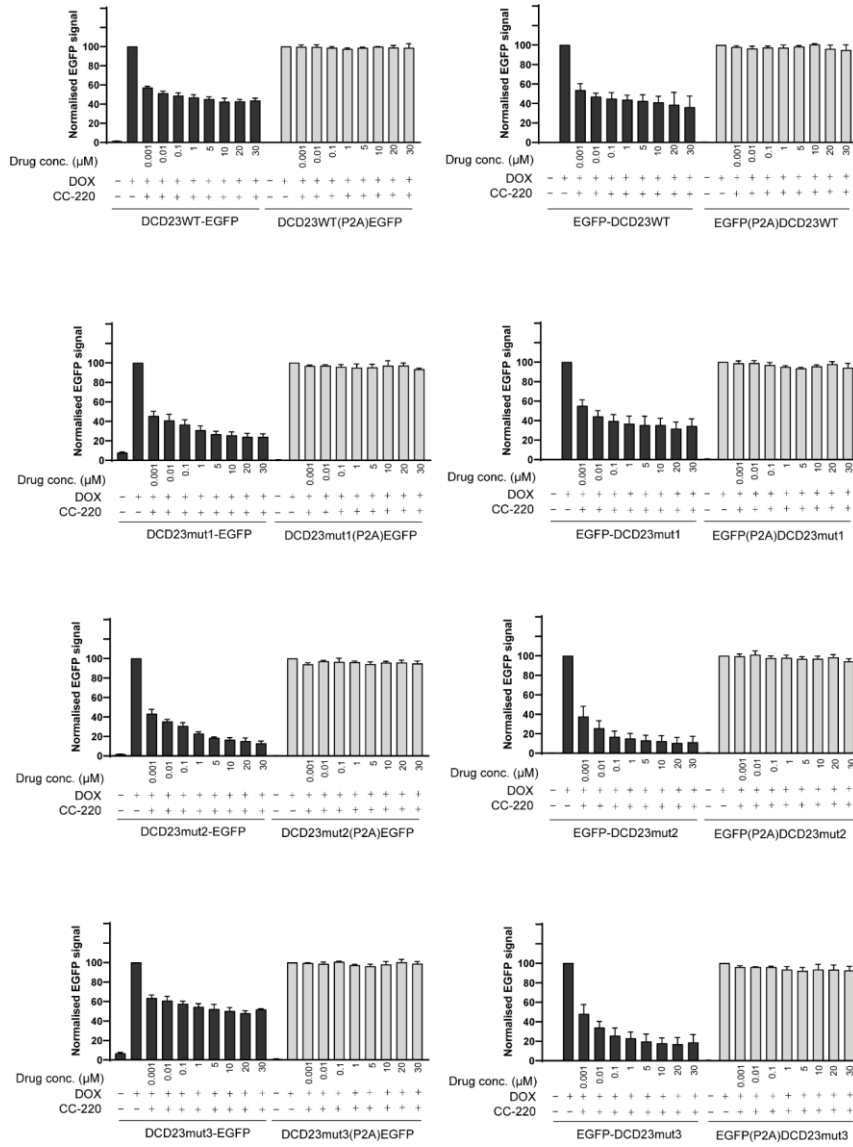
**small green font:** apical Glycine residue of chimeric ZFP91/Ikaros ZF2 domain

**Figure 3.4. 1 Structure and sequence-based design of DCD23mut1-mut3.** Schematic representation of DCD23mut1-mut3 protein sequences evaluated in this thesis. Key DNA binding residues mutated in DCD23 are depicted in red font.

Three DCD23 mutants, namely DCD23mut1-mut3, harbouring point-mutations of key DNA-binding lysine and arginine residues were designed and commercially synthesised (**Figure 3.4. 1**) (Wolfe, Nekludova and Pabo, 1999). These were used to generate lentiviral backbone vectors expressing wild type (WT) DCD23 or DCD23mut1-mut3 N- and C-terminally tagged EGFP under the control of a DOX-inducible promoter. In addition, a set of corresponding P2A controls was designed for each DCD. Details of the cloning procedure are presented in the Materials and Methods section (**Figure 2. 7**). Subsequently, the final lentiviral vectors were used to generate HMEC cells stably expressing the fusion constructs following DOX administration.

### ***3.4.2 Mutant DCD23 motifs induced a potent and robust degradation of EGFP***

To demonstrate that DCD23mut1-mut3 retain the ability to induce degradation of EGFP following IMiD/CELMoD treatment, HMEC cells stably expressing EGFP fused N- or C-terminally to DCD23, DCD23mut1-mut3 or their corresponding P2A controls were treated with increasing CC-220 concentrations for 4h, ranging from 0.001  $\mu\text{M}$  to 30  $\mu\text{M}$ . Treatment concentrations as low as 0.001  $\mu\text{M}$  caused a measurable loss of EGFP fluorescence signal for all fusion constructs, demonstrating both the high potency of CC-220 and the retained ability of the DCD23 mutants to induce degradation of EGFP (**Figure 3.4. 2**). As expected, treatment of HMEC cells expressing the corresponding P2A controls

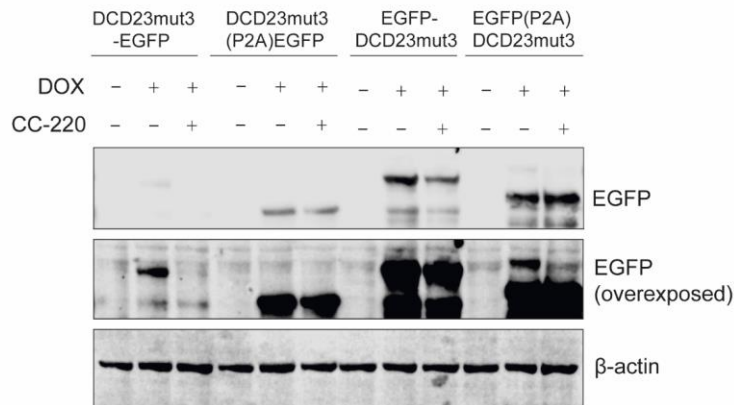
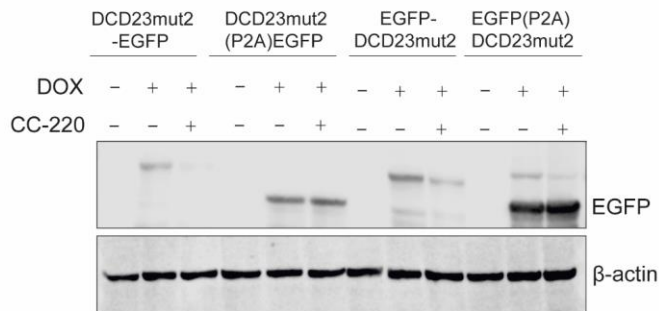
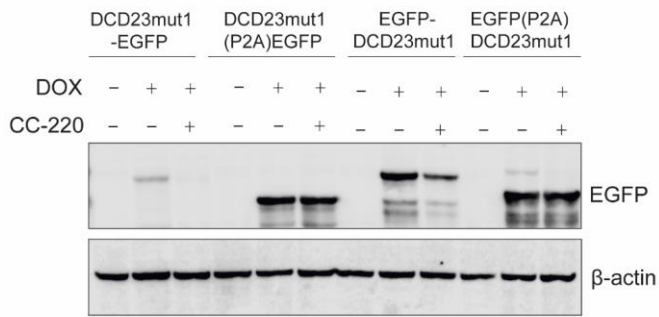
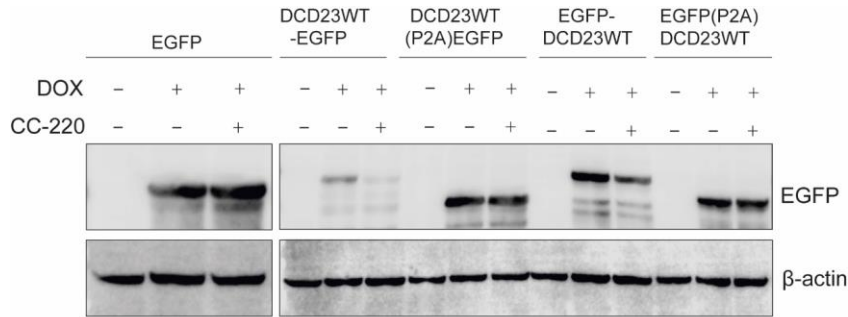


**Figure 3.4. 2 Functional evaluation of DCD23mut1-mut3 fused to the N- or C-terminus of EGFP.** HMEC cells transduced with either DCD23 or DCD23mut1-mut3 -tagged EGFP and their respective P2A controls were treated with 1 µg/ml DOX for 24h to stimulate expression of the fusion constructs. Cells were subsequently treated with increasing concentrations of CC-220, ranging from 0.001 to 30 µM for 4h. Loss of target protein was validated via measuring EGFP

median intensity by flow cytometry. Error bars represent the standard deviation (SD) of the data. Data are representative of 3 independent biological experiments (n=3).

with various IMiDs/CELMoDs did not induce a decrease in EGFP fluorescence levels (**Figure 3.4. 2**) Furthermore, these flow cytometry observations were also confirmed by immunoblot analysis of cellular lysates treated with 10  $\mu$ M CC-220 for 4h (**Figure 3.4. 3**). Collectively, the data here demonstrated that all three DCD23 mutants can induce a potent and robust degradation of a target protein, comparable to the WT DCD23, and support their use to further explore the effects of DCD fusion to the subcellular localisation of EGFP.

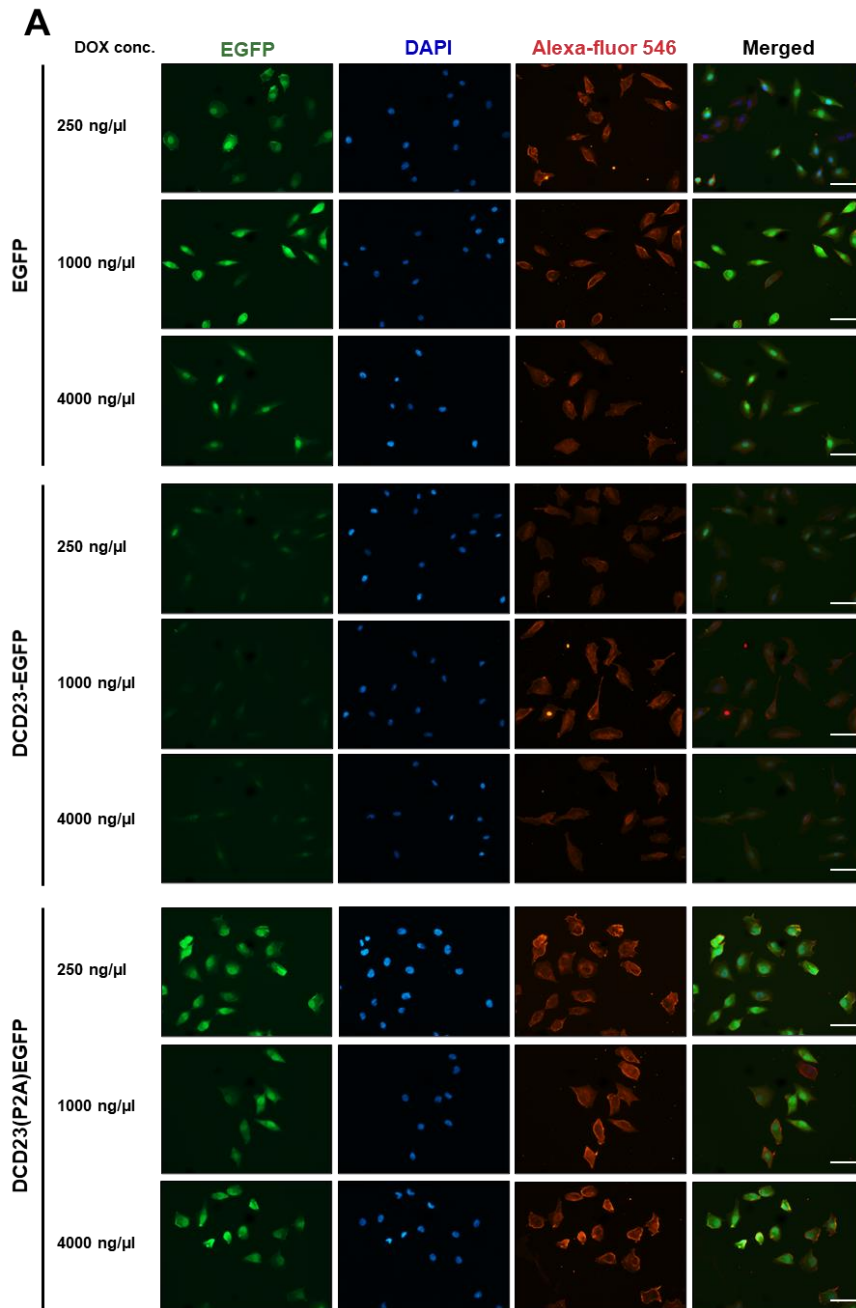
Results

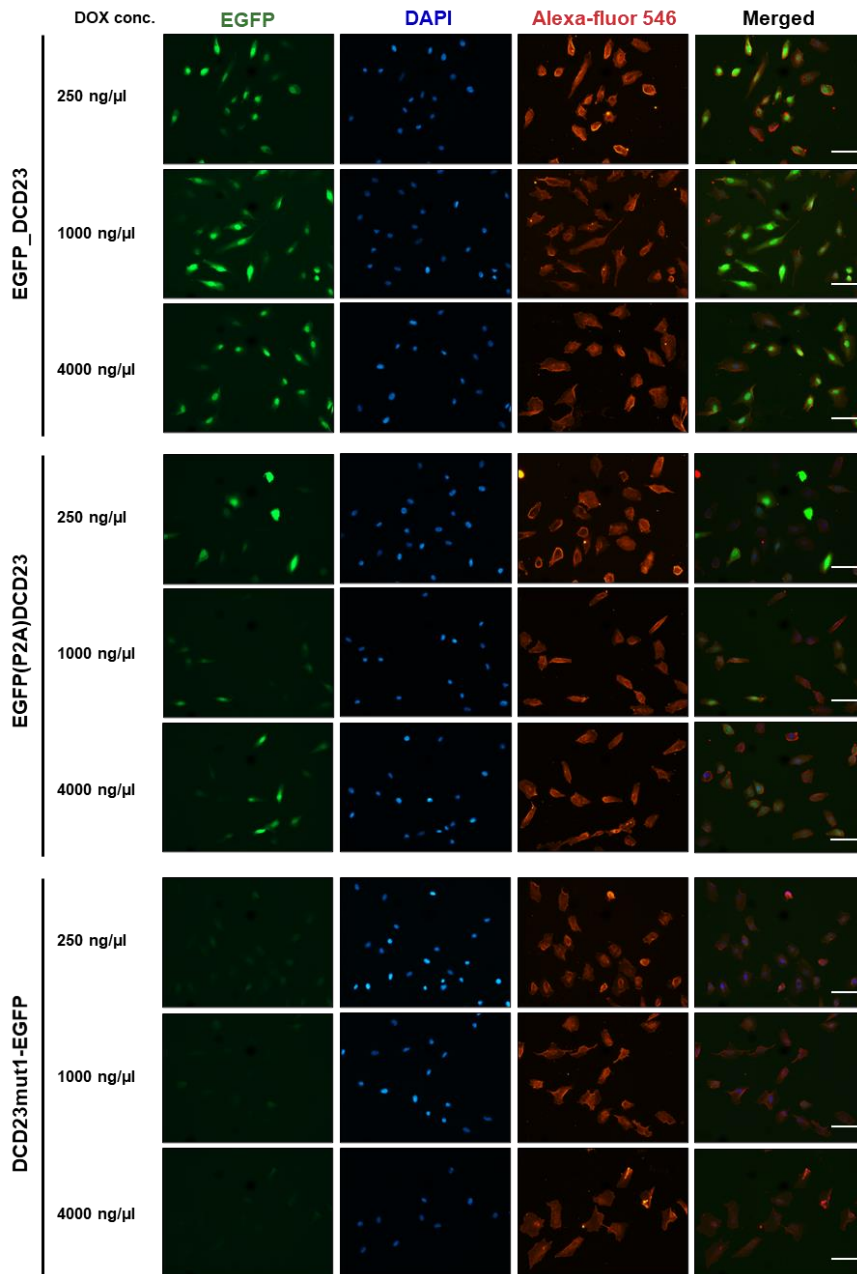


**Figure 3.4. 3 Fusion of DCD23 or DCD23mut1-mut3 to either terminus of EGFP induced its degradation following CC-220 treatment.** HMEC cells transduced with DCD23 or DCD23mut1-mut3 -tagged EGFP and their corresponding P2A controls were treated with 1 µg/ml DOX for 24h to induce expression of the fusion constructs, followed by treatment with 10 µM of CC-220 for 4h. An EGFP antibody was used to probe for the levels of the fusion constructs by Western blotting. Images are representative of 3 independent biological experiments (n=3).

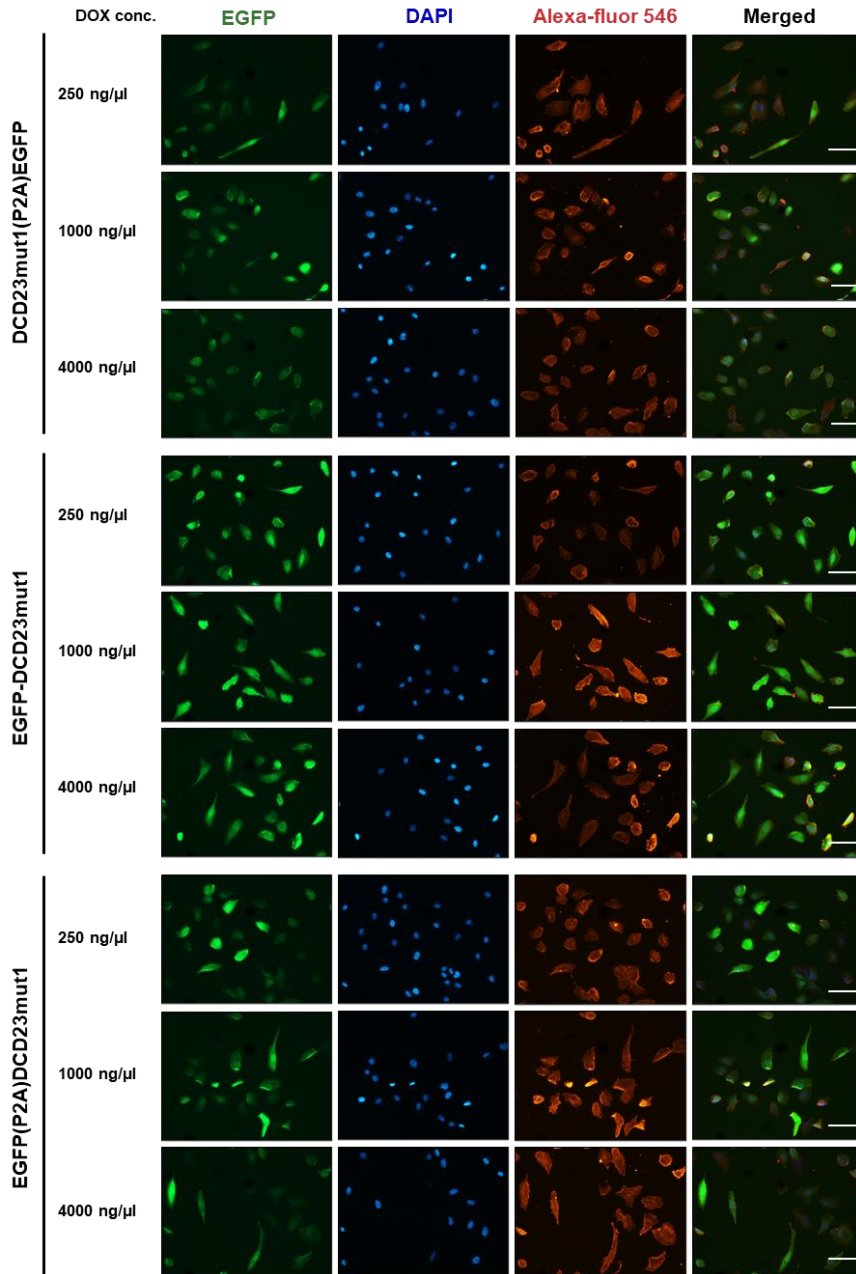
### ***3.4.3 Examining the cellular localisation of mutant DCD23 motifs fused to EGFP***

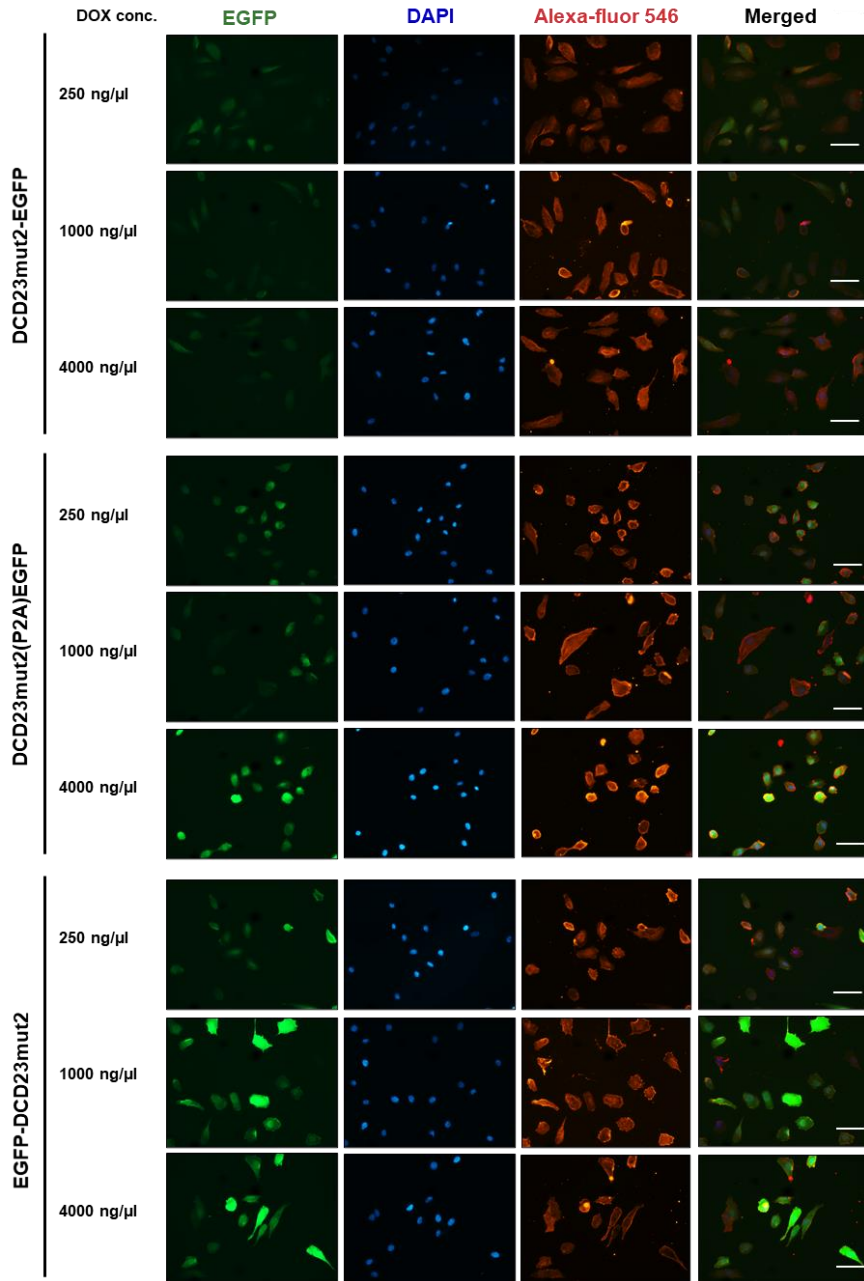
Having demonstrated that DCD23mut1-mut3 can induce EGFP degradation following IMiD/CELMoD treatment, the subcellular localisation of the fusion proteins was examined. First, the expression levels of EGFP fused N- or C-terminally to DCD23, DCD23mut1-mut3 or their corresponding P2A controls were assessed by both fluorescence microscopy and flow cytometry by treating cells with three increasing DOX concentrations (250 ng/ml, 1000 ng/ml, and 4000 ng/ml) for 24h (**Figure 3.4. 4 A-B**). Between the different DOX concentrations administered to each cell line, there was no difference in the expression levels of the fusion proteins, demonstrating that higher DOX concentrations do not result in increased fusion protein expression. The expression levels of N-terminal fusions of DCD23 and DCD23mut1-mut3 to EGFP were lower compared to other fusion proteins, whilst EGFP(P2A)DCD23mut2 was found to be the most highly expressed fusion protein (**Figure 3.4. 4 A-B**). The nuclear/cytoplasmic distribution for each fusion protein was determined via fluorescence microscopy of fixed cells. Untagged EGFP was found to be strongly localised in the

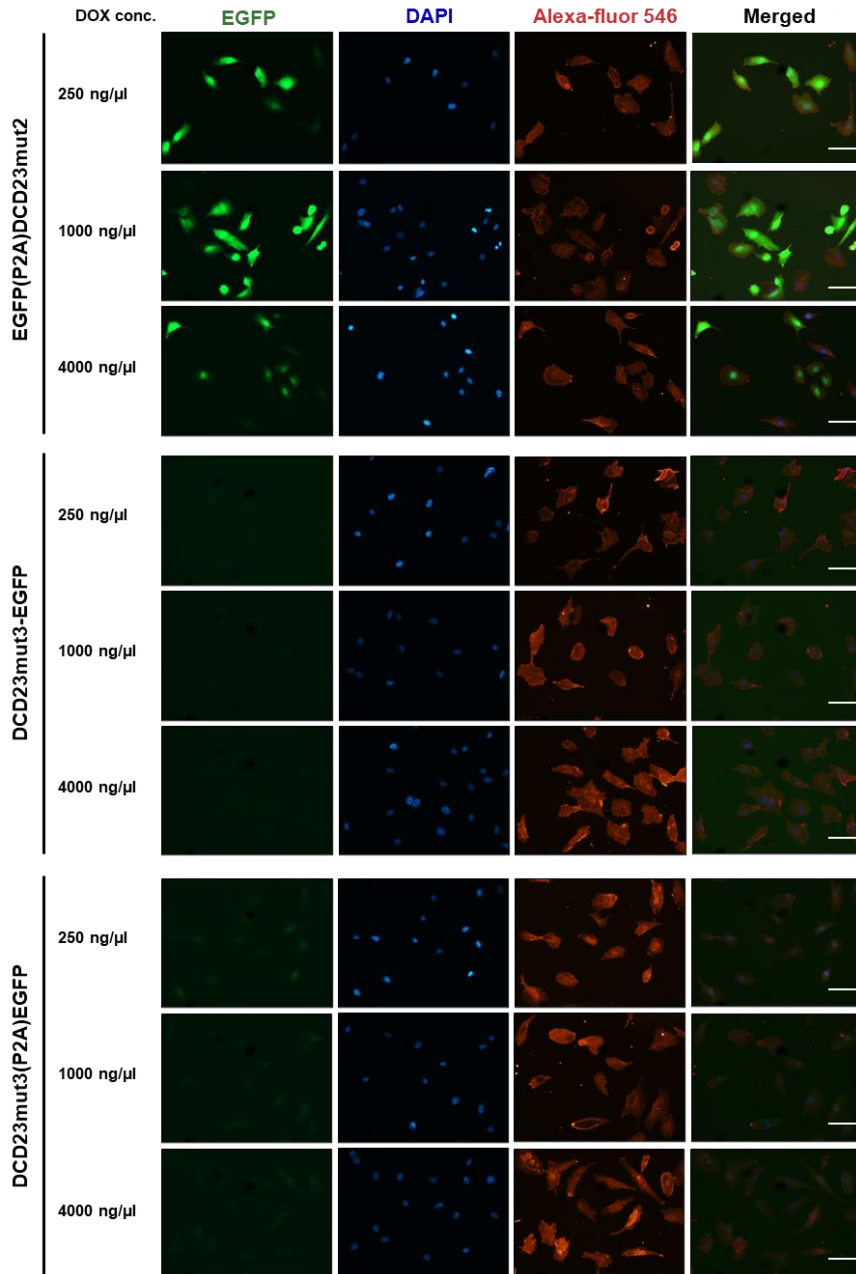


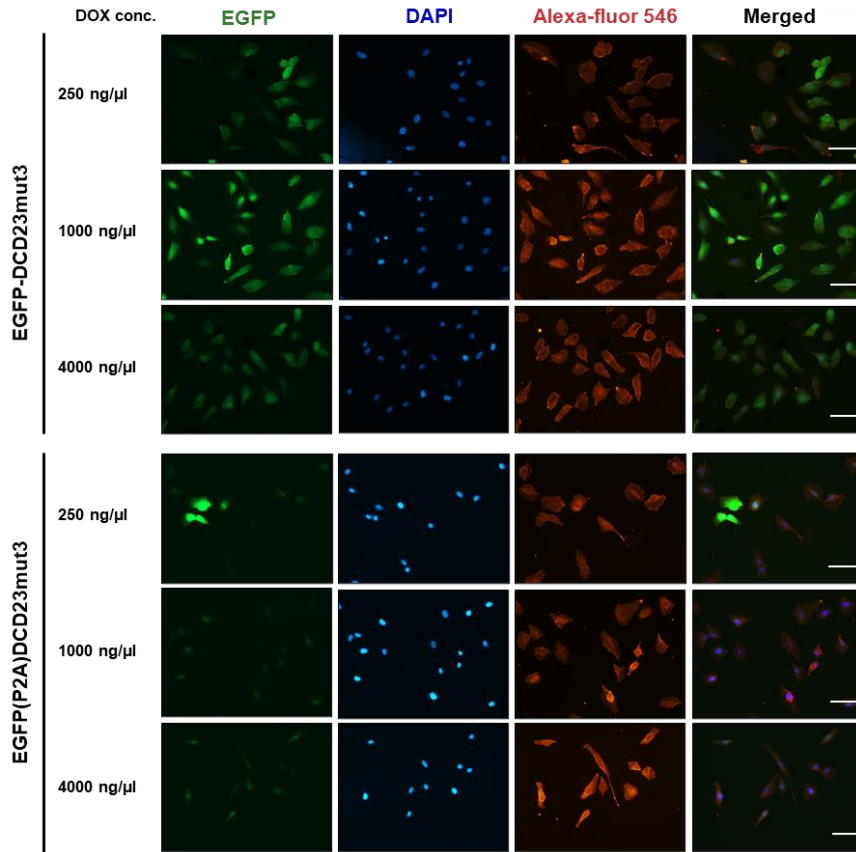




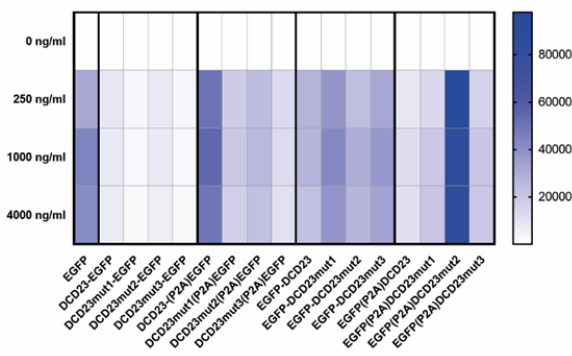




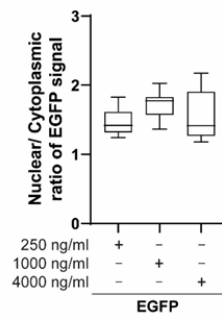


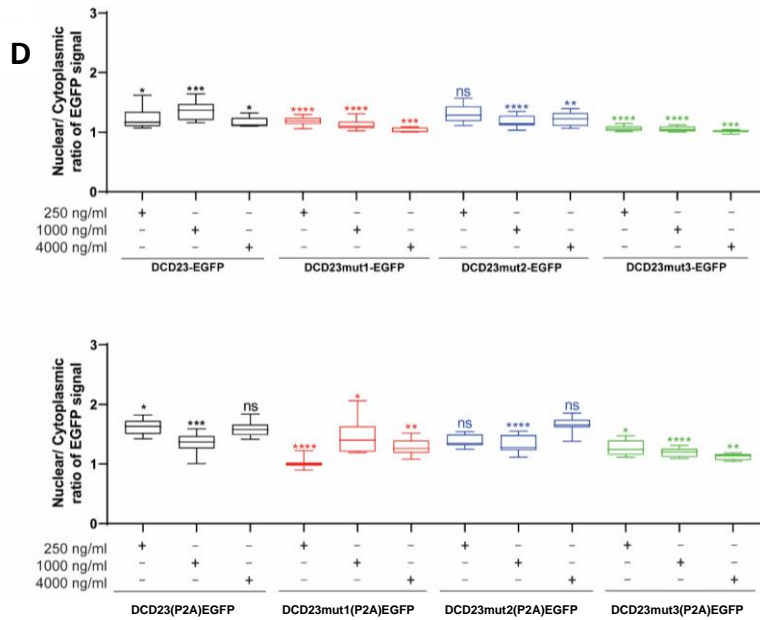


**B**



**C**

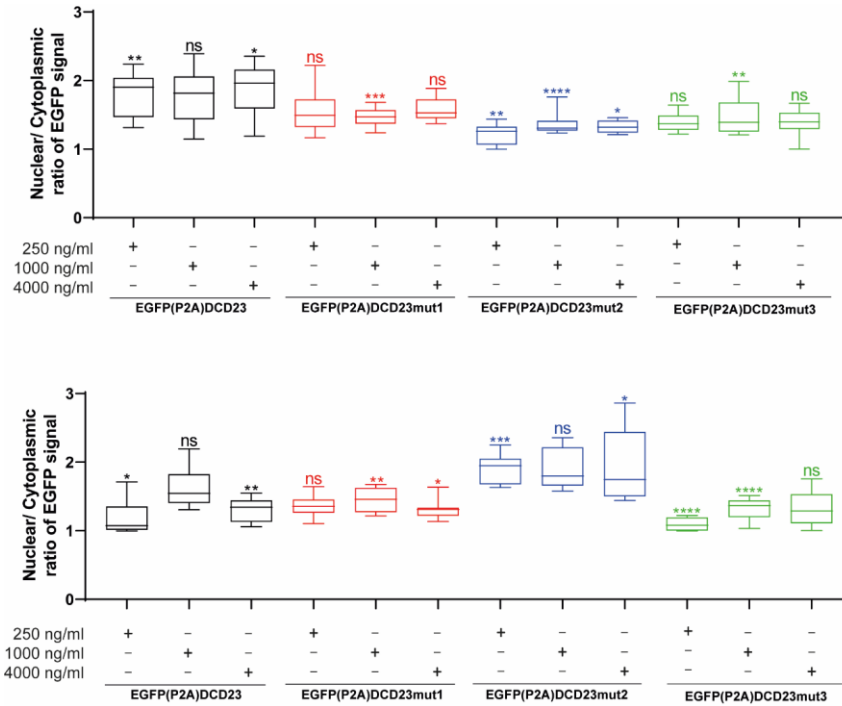




**Figure 3.4.4 Measuring nuclear/cytoplasmic localisation of N-terminally DCD23 and DCD23mut1-mut3-tagged EGFP.** (A) Representative fluorescent microscopy images depicting HMEC cells expressing the various EGFP fusion constructs under different DOX concentrations. HMEC cells stably transduced with EGFP N- or C-terminally tagged with DCD23, DCD23mut1, DCD23mut2, DCD23mut3 or their respective P2A controls were treated with increasing DOX concentrations for 24h. Cell nuclei were stained with DAPI. HMEC cells were also incubated with MS anti-actin antibody followed with Alexa-Fluor 546 -conjugated anti-actin antibody to stain for the cytoplasm. Scale bar, 50  $\mu$ m. (B) Heatmap comparing expression levels of EGFP fusion constructs under different DOX concentrations. EGFP expression was measured using flow cytometry. (C) HMEC cells stably transduced with EGFP were treated with increasing DOX concentrations for 24h. EGFP expression was measured using flow cytometry. (D) HMEC cells stably transduced with DCD23-EGFP, DCD23mut1-EGFP, DCD23mut2-EGFP, DCD23mut3-EGFP and their corresponding P2A controls were treated with increasing DOX concentrations for 24h. EGFP nuclear/cytoplasmic distribution for each fusion protein was determined via fluorescence microscopy of fixed cells. Statistical significance was calculated using unpaired Student's t-tests between untagged EGFP (C) and WT DCD23/ DCD23mut1-mut3 with a two-tailed distribution. Statistically not significant results are denoted with ns ( $\geq 0.05$ ). Statistically significant results are denoted with an asterisk (\*  $p = 0.01 - 0.05$ , \*\*  $p = 0.001 - 0.01$ , \*\*\*  $p = 0.0001-0.001$ , \*\*\*\*  $p < 0.0001$ ). Data are representative of 3 independent biological experiments ( $n=3$ ).

**Commented [SN8]:** Included the epi-fluorescence images of all DCDs tested as part A of this figure.

nucleus contrary to previous observations (**Figure 3.4. 4 C**) (Seibel *et al.*, 2007). DCD23-EGFP and DCD23mut2-EGFP were also found to be more highly localised in the nucleus, whilst DCD23mut1-EGFP and DCD23mut3-EGFP were identified to be equally localised in the nucleus and the cytoplasm (**Figure 3.4. 4 D**). DCD23(P2A)EGFP, DCD23mut1(P2A)EGFP, DCD23mut2(P2A)EGFP and DCD23mut3(P2A)EGFP also demonstrated a stronger preference for nuclear localisation (**Figure 3.4. 4 D**). Similarly, C-terminal fusions of WT DCD23, DCD23mut1-mut3 and their respective P2A controls to EGFP, showed that the target protein primarily resides in the nucleus (**Figure 3.4. 5**). Interestingly, these differences in the localisation of the fusion proteins correlated with their expression levels as measured by flow cytometry, indicating that higher expression of the fusion protein leads to a greater difference in nuclear/cytoplasmic expression, with more expression in the nucleus. Indeed, EGFP(P2A)DCD23mut2 which was identified as the highest expressed fusion protein had also the biggest nuclear/cytoplasmic ratio, whilst the lowest expressed fusion proteins DCD23mut1-EGFP and DCD23mut3-EGFP, were equally localised in the nucleus and the cytoplasm. Collectively, although there is a pattern in the data linking high expression levels of the target protein with nuclear localisation, no evidence of the tested DCD23 motifs affecting cellular localisation was found, thus meriting further investigation.



**Figure 3.4.5** Measuring nuclear/cytoplasmic localisation of C-terminally DCD23 and DCD23mut1-mut3 -tagged EGFP. HMEC cells stably transduced with EGFP-DCD23, EGFP-DCD23mut1, EGFP-DCD23mut2, EGFP-DCD23mut3 and their corresponding P2A controls were treated with increasing DOX concentrations for 24h. EGFP nuclear/cytoplasmic distribution for each fusion protein was determined via fluorescence microscopy of fixed cells. Statistical significance was calculated using unpaired Student's t-tests between untagged EGFP (Figure 3.4.4 C) and WT DCD23/ DCD23mut1-mut3 with a two-tailed distribution. Statistically not significant results are denoted with ns ( $\geq 0.05$ ). Statistically significant results are denoted with an asterisk (\*  $p = 0.01 - 0.05$ , \*\*  $p = 0.001 - 0.01$ , \*\*\*  $p = 0.0001 - 0.001$ , \*\*\*\*  $p < 0.0001$ ). Data are representative of 3 independent biological experiments ( $n=3$ ).

### **3.4.4 Interim Discussion Results Subchapter 3.4**

In order to exert their physiological functions, a diverse set of proteins such as cell cycle regulators, transcription factors and histones need to be efficiently transported from the cytoplasm where they are synthesised through the nuclear pore complex to their destination nuclear compartments (Lu *et al.*, 2021). To achieve this, most nuclear proteins contain nuclear localisation signals (NLS) as part of their amino acid sequence. NLS are recognised by nuclear transporters which interact with nucleoporins facilitating the transfer of the protein through the nuclear pore complex (Lu *et al.*, 2021). There are two types of NLS, the classical NLS, which can be further categorised into monopartite and bipartite and the non-classical NLS. Monopartite NLS are composed of a single cluster of 4-8 basic amino acids with four or more positively charged residues and can be characterised by the K (K/R) X (K/R) motif (where X denotes any amino acid) (Lu *et al.*, 2021). On the contrary, bipartite NLS are composed of two clusters of 2-3 positively charged residues interspaced by a 9-12 proline-rich amino acid-linker region, with the consensus sequence expressed as R/K(X)<sub>10-12</sub>KRXX (Lu *et al.*, 2021). Non-classical NLS as the name indicates are atypical NLS that are not based on arginine or lysine residues; an example being the proline-tyrosine (PY)-NLS which is characterised by a stretch of 20-30 residues, consisting of N-terminal hydrophobic or basic motifs and C-terminal R/K/H(X)<sub>2-5</sub>PY motifs (where X<sub>2-5</sub> denotes any sequence of 2–5 residues) (Lu *et al.*, 2021).

The C2H2 ZF transcription factor Ikaros lacks a defined NLS but its subcellular localisation is postulated to be regulated via posttranslational



modifications and heterodimerization with other members of its family of DNA binding proteins (Cobb *et al.*, 2000; Uckun *et al.*, 2012). Since the sequence of the ZFP91/Ikaros chimeric DCD23 is primarily based on the N-terminal ZF1, ZF2 and ZF3 domains of Ikaros, the effects on the subcellular localisation of a target protein fused to DCD23 was assessed by examining the localisation of EGFP fused N- and C- terminally to DCD23. Along with WT DCD23, a set of three mutants (DCD23mut1, DCD32mut2 and DCD23mut3) containing point-mutations of key DNA-binding lysine and arginine residues were designed to be tested when fused to either terminus of EGFP (Wolfe, Nekludova and Pabo, 1999). These point mutants were generated as a contingency plan in case WT DCD23 was found to promote the nuclear localisation of EGFP. Interestingly, all three mutant DCD23 sequences were demonstrated to retain their functionality by inducing degradation of EGFP following CC-220 treatment similarly to WT DCD23. The degradation of EGFP fused to either of the DCD23 mutants also highlights that even though lysine and arginine residues are important for DNA binding, their role in engaging the IMiD/CELMoD-CRBN complex could be redundant. In addition, these data support previous studies, by further corroborating that the geometric arrangement of the  $\beta$ -hairpin loop rather than its consensus sequence is critical for binding to the IMiD/CELMoD-CRBN complex (Krönke *et al.*, 2014; Matyskiela *et al.*, 2016; Petzold, Fischer and Thomä, 2016; Sievers *et al.*, 2018). When the subcellular localisation of the fusion proteins was examined, a correlation between high levels of expression and increased nuclear localisation was observed. Interestingly, even untagged EGFP, which was also highly expressed, was found to be highly concentrated within the nucleus compared to the cytoplasm, despite the fact that EGFP lacks a NLS (Seibel *et*

*al.*, 2007). The above observations could denote an artefact of the overexpression of untagged and iTAG-tagged EGFP, which promotes the accumulation of EGFP by passive diffusion into the nucleus, thus revealing a localisation event that occurs in a particular condition. Indeed, EGFP has a small molecular weight (27 kDa) and has been previously described to translocate into the nucleus via nuclear pores (Seibel *et al.*, 2007). Moreover, in a previous study, even EGFP homotetramers and homoexamers were found in the nuclei of five analysed mammalian cell lines, further highlighting that fusion of DCD23 with either monomeric or multimeric EGFP should not be used for examining nuclear import processes (Seibel *et al.*, 2007). Therefore, the data from the localisation assay, failed to show whether fusion of the DCD23 motif to a target protein affects subcellular localisation, thus meriting further investigations to explore whether this is a possibility.

---

## 4. DISCUSSION

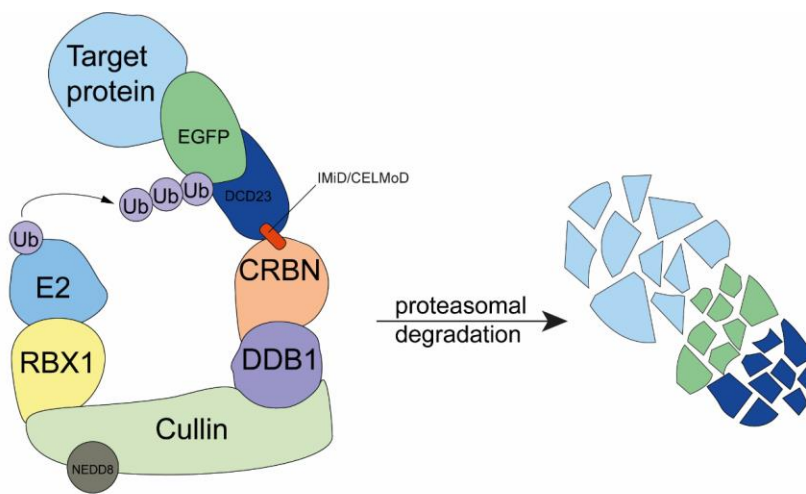
Chemical-biology systems capable of precise temporal control of target protein expression can provide unique insights of a target's functions, therefore making their application invaluable both in basic research and during the process of target validation in drug discovery. Over the past two decades, the development and application of several such systems for studying the functions of target proteins, has allowed to overcome some of the key limitations of genetic methods. For instance, regarding the application of the RNAi technology in the study of target function, incomplete target knockdown in certain cases could convolute observed phenotypes and lead to erroneous conclusions regarding the target protein's functions (Sigoillot and King, 2011; Boettcher and Mcmanus, 2015). Moreover, several reports have raised the issue of molecular compensation and cellular adaptation following the knockout (KO) of a target gene with CRISPR/Cas genome editing tools; which can lead to observed phenotypes unrelated to the loss of target protein expression and the formulation of false conclusions (Boettcher and Mcmanus, 2015; Olive *et al.*, 2018). The use of well-established chemical-biology tools, such as the AID and dTAG systems, has overcome such drawbacks by allowing for an acute, modulated, and reversible loss of target protein expression. These tools have been also successfully implemented in elucidating primary roles of transcriptional regulators, via early time-resolved analysis (minutes to hours) of the evoked phenotypes and by also capitalizing on their reversibility (Jaeger and Winter, 2021). For example, the use of the AID system in studying regulators of 3D

chromatin structure, such as the cohesin complex, has enhanced our mechanistic understanding of transcription regulation by the spatial conformation of the chromatin fibre (Rao *et al.*, 2017). In addition, studies employing both the AID and dTAG systems to acutely degrade the transcriptional co-activator Mediator and RNA polymerase II have unravelled that both targets are dispensable for physically tethering regulatory DNA, contrary to previous models, whilst also demonstrating that they are critical for maintaining nascent transcription (Khattabi *et al.*, 2019; Jaeger *et al.*, 2020). Moreover, a recent study employed the AID system for rapid degradation of MYC and bromodomain-containing protein 4 (BRD4) coupled with metabolic sequencing of nascent mRNA to study the principal consequences following acute depletion of these essential DNA-binding transcription factors (Muhar *et al.*, 2018). MYC was identified to primarily control metabolic processes such as ribosome biogenesis and de novo purine synthesis, whilst BRD4 degradation revealed a global down-regulation of transcription via impaired RNA polymerase II pause release (Muhar *et al.*, 2018). Similarly, the dTAG system has been applied for rapid degradation of endogenous oncogenic transcription factor AML1-ETO (RUNX1-RUNX1T1), enabling the identification of direct gene targets that constitute a core AML1-ETO regulatory network (Stengel *et al.*, 2021).

Even though from the above studies, it is evident that chemical-biology tools have provided novel mechanistic insights into biological processes, current well-established systems (DD, LID, AID, SMASh, Trim-Away and PROTAC-based systems) face some key limitations related to their application. For instance, auxin-based systems (AID, AID-ARF16, miniIAA7 and AID2) lack appropriate controls for any toxicity and off-target effects mediated by the use of

auxins (NAA/ IAA/ 5-Ph-IAA) (Nishimura *et al.*, 2009; Li *et al.*, 2019; Sathyan *et al.*, 2019; Yesbolatova *et al.*, 2020). Indeed, IAA has been previously reported to cause kidney toxicity when it is converted to indoxyl sulphate in the liver, therefore potentially limiting the use of these systems for *in vivo* studies (Yesbolatova *et al.*, 2020). Moreover, the auxin-based systems and the Trim-Away system exhibit multicomponent complexity since the first requires the exogenous overexpression of TIR1 to form a functional E3 ubiquitin ligase for target protein degradation and the latter the exogenous overexpression of TRIM21 to form a functional Trim-Away system (Nishimura *et al.*, 2009; Clift *et al.*, 2017, 2018). Other systems such as the LID system and Aiolos ZF2 domain-based 25mer degron, exhibit low versatility as they only mediate degradation of a target when fused at its C-terminus, therefore, excluding the study of target proteins that can only functionally tolerate genetic fusions at their N-termini (Bonger *et al.*, 2011; Koduri *et al.*, 2019). Therefore, an ideal system would involve the use of non-toxic compounds that mediate the potent degradation of a minimally modified target protein, thus allowing the study of acute protein removal prior to physiological adaptation. Based on our increasing understanding of the molecular mechanisms underpinning the UPS (ubiquitin proteasome system) and newly identified small molecules that engage key molecular components of the UPS (e.g. E3 ubiquitin ligases), I developed iTAG (Inducible and TArgeted protein deGradation). iTAG encompasses the fusion of a functional degron motif from known CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates to a target protein and the use of commercially available potent IMiDs/CELMoDs for acute and reversible target protein degradation (**Figure 4.1**). Two candidate degradation tags were identified from a cellular degradation assay of twenty-six degron-

containing domain (DCD) motifs, namely Ikaros-based DCD19 (131 residues) and chimeric ZFP91-Ikaros DCD23 (60 residues), with the latter selected as the optimal degradation tag as it was demonstrated to degrade a broad range of targets both in vitro and in vivo.



**Figure 4. 1 The iTAG system.** Schematic representation of the iTAG system which involves genetic fusion of EGFP-DCD23 or DCD23 to a target protein and the use of an IMiD/CELMoD (e.g. CC-220) to mark the target protein for proteasome-mediated degradation.

As the rules of IMiD/CELMoD-mediated protein degradation are largely unknown, an empirical approach was implemented to identify a functional and potent degron motif, derived from known CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates. Previous structural and in vitro mutagenesis studies had shown that a  $\beta$ -hairpin loop with a critical glycine at the apex is required for neosubstrate

binding to the IMiD/CELMoD-CRBN complex (Krönke *et al.*, 2014; Matyskiela *et al.*, 2016, 2020; Petzold, Fischer and Thomä, 2016; An, Charles M Ponthier, *et al.*, 2017; Sievers *et al.*, 2018). This  $\beta$ -hairpin loop constitutes a geometric degron motif which can be found as part of the tertiary structure of neosubstrates such as GSPT1 and CK1 $\alpha$  and at the turn of two anti-parallel  $\beta$ -strands in a C2H2 ZF domain of ZF transcription factor neosubstrates, such as Ikaros, Aiolos and ZFP91. The first set of DCDs designed, DCD1 to DCD9, were based on this  $\beta$ -hairpin loop along with a flanking variable-length sequence from the non-ZF domain-containing neosubstrates GSPT1 and CK1 $\alpha$ . Regarding DCD8 (18 residues) which was based on the CK1 $\alpha$   $\beta$ -hairpin loop (residues 35-41, isoform 1) found in the kinase N-lobe and a minimal flanking sequence from CK1 $\alpha$  isoform 1 protein sequence (residues 30-47), no degradation of cMYC-EGFP was observed following IMiD/CELMoD treatment. To obtain a functional degron motif based on CK1 $\alpha$ , further DCDs need to be designed and tested based on the whole or extended CK1 $\alpha$  kinase N-lobe (residues 1-93) (Petzold, Fischer and Thomä, 2016; Jiang *et al.*, 2018). The further contacts from the extra CK1 $\alpha$  residues could be required to stabilize the  $\beta$ -hairpin loop in the correct geometry for effective binding and also form additional interactions with residues on the surface of CRBN. DCD3 which was solely based on the GSPT1  $\beta$ -hairpin loop, failed to induce degradation of cMYC-EGFP following IMiD/CELMoD treatment, demonstrating that the  $\beta$ -hairpin loop alone cannot act as a functional degron motif. Similarly, target protein degradation in the presence of IMiDs/CELMoDs was not observed for fusions with DCD4 and DCD5, which were based on both the GSPT1  $\beta$ -hairpin loop and a minimal flanking sequence forming an unstructured  $\beta$ -barrel domain from the C-terminus of GSPT1 protein. On the

other hand, DCD2 which was designed to incorporate the whole GSPT1 C-terminal  $\beta$ -barrel domain containing this  $\beta$ -hairpin loop, as well as DCD1, which encompassed an extra GSPT1 C-terminal  $\beta$ -barrel domain, induced a potent, robust, and acute degradation of the target protein in the presence of CC-885. This indicates that in order to form a functional degron motif, at least the whole GSPT1 C-terminal  $\beta$ -barrel domain containing the  $\beta$ -hairpin loop is necessary. An intact  $\beta$ -barrel domain could be required to stabilize the tertiary structure of the  $\beta$ -hairpin loop in the correct geometry for engaging the CC-885-bound CRL4<sup>CRBN</sup> E3 ubiquitin ligase, as well as, providing additional residues that form interactions with amino acids on the surface of CRBN for efficient binding. Indeed, the latter is supported by previous in vitro mutagenesis assays and structural studies of GSPT1 binding to CC-885/ CRBN complex, showing that GSPT1 forms direct protein-protein interactions with CRBN residues N351, H357, W400, Y355, H397, and V388 via both hydrogen and Van der Waal's bonds (Matyskiela *et al.*, 2016). These data are also further corroborated from a previous co-immunoprecipitation assay of GSPT1 with CRBN surface mutants, identifying the critical role of CRBN residues N351, H357, E377, V388 and H397 for recruiting GSPT1 to the CC-885/CRBN complex (Matyskiela *et al.*, 2016). Even though DCD2 was identified as a potential candidate degron motif for developing iTAG, CC-885 was shown to be cytotoxic. Indeed, CC-885-mediated degradation of GSPT1 has been previously identified as a culprit for the observed cytotoxicity effects of this thalidomide analogue (Matyskiela *et al.*, 2016). However, it would be interesting to test a recently developed analogue of CC-885, CC-90009, which has been demonstrated to solely induce the degradation of GSPT1 (unlike CC-885 which also recruits other neosubstrates for



degradation) and to be non-toxic at nM concentrations in THLE-2, a hepatocyte cell line (Matyskiela *et al.*, 2016; Hansen *et al.*, 2021; Surka *et al.*, 2021).

The second set of DCDs designed, DCD10-DCD22, were based on the ZF2 domain of either Aiolos, Ikaros or Eos, which encompasses the  $\beta$ -hairpin loop with a critical glycine at the apex (Krönke *et al.*, 2014). The ZF2 domain of Aiolos and Ikaros has been previously identified to be sufficient as a minimal unit for IMiD/CELMoD-induced CRL4<sup>CRBN</sup>-dependent degradation in cells (Krönke *et al.*, 2014). Therefore, DCD10-DCD21 were designed to either incorporate the minimal Aiolos or Ikaros ZF2 domain or extended to contain additional whole or partial ZF domains (Aiolos ZF1, ZF3 and ZF4; Ikaros ZF1 and ZF3) flanking it. Regarding DCD22, which was based on the ZF2 domain of Eos (residues 187-209), no degradation of cMYC-EGFP was observed following IMiD/CELMoD treatment. In contrast, DCD10 and DCD20 based on the ZF2 domain of Aiolos and Ikaros respectively, induced partial loss of target protein in the presence of some of the tested IMiDs/CELMoDs. Both Eos/IKZF4 and Helios/IKZF2 (another ZF transcription factor of the Ikaros family), compared to Aiolos/IKZF3 and Ikaros/IKZF1, contain a single amino acid difference in their ZF2 domain, a Gln to His at position 188. This single amino acid variability in the ZF2 domain of Eos and Helios has been previously demonstrated to prevent their degradation following treatment with an IMiD/CELMoD (Krönke *et al.*, 2014). Substitution of the His188 to a Gln in Eos was shown to render Eos susceptible to lenalidomide-induced degradation, whilst, when the corresponding glutamine (Gln147) in Aiolos was changed to a His residue, resistance to lenalidomide-induced degradation was developed (Krönke *et al.*, 2014). Therefore, the data from our cellular degradation assay on Eos ZF2 corroborate previous studies and further

highlight the fact that amino acid changes in this region provide the basis for differential sensitivity to IMiDs/CELMoDs between Ikaros family members. Moreover, the partial loss of target protein observed with DCD10 and DCD20 highlighted the importance of additional ZF domains for optimal degradation. Indeed, Aiolos ZF2-based DCD11 to DCD17 and Ikaros ZF2-based DCD19 and DCD21 which incorporated extra ZF domains from Aiolos or Ikaros induced a more potent and robust loss of target protein in the presence of various IMiDs/CELMoDs. This could also explain why the previously described Aiolos ZF2 domain-based 25mer degron induced partial loss of target protein, when examined in our cellular degradation assay (Krönke *et al.*, 2014; Koduri *et al.*, 2019). The additional ZF domains of these DCDs could therefore confer enhanced binding to the CRL4<sup>CRBN</sup> E3 ubiquitin ligase by interacting with amino acids on the surface of CRBN. This statement is further supported by a previous *in vitro* study which demonstrated that the combination of ZF2 and ZF3 domains of Aiolos to form a degron motif confers higher binding affinity to IMiD/CELMoD-bound CRBN than ZF2 alone (Sievers *et al.*, 2018). From this set of tested DCDs, DCD19 (131 residues) based on Ikaros full length ZF1, ZF2 and ZF3 domains was identified as an ideal candidate degron motif for developing the iTAG system, since it induced the most potent and robust loss of target protein in presence of the non-toxic CELMoD CC-220. Even though CC-220 induced a potent degradation of cMYC-EGFP-DCD19 in the cellular degradation assay, in the *in vitro* TR-FRET assay CC-220 was not identified to induce a stronger binding of DCD19 to the CRBN/DDB1 complex, thus demonstrating a disconnect between the *in vitro* binding affinity and degradation efficiency in cells. This highlighted the importance of examining the degradation efficiency of candidate degron motifs in

cellular degradation assays and further supported our decision to employ two orthogonal approaches for defining an optimal DCD motif.

The third set of DCDs designed, DCD23 to DCD26, were based on a chimeric ZFP91-Ikaros ZF2 domain ( $\beta$ -hairpin loop from ZFP91 ZF4 domain fused in tandem to the Ikaros ZF2  $\alpha$ -helix) and additional flanking ZF domains from Ikaros. From the cellular degradation assay, DCD23 was identified as a second candidate degron motif for developing iTAG, as it was the shortest DCD (60 residues) from this set to induce a potent loss of cMYC-EGFP following IMiD/CELMoD treatment. DCD23 encompassed the Ikaros-ZFP91 chimeric ZF2 domain fused to only the  $\alpha$ -helix of Ikaros ZF1 and the  $\beta$ -hairpin loop of Ikaros ZF3. Consistent with the cellular assays, the CELMoD CC-220 was identified in the in vitro TR-FRET binding affinity assay as the strongest recruiter of DCD23, contrary to DCD19 which was shown to engage the CRBN/DDB1 complex with similar binding affinities for the tested IMiDs/CELMoDs. More importantly, CC-220 unlike CC-885, when tested in the cell viability assay was found to be non-cytotoxic in non-haematological cell lines. Besides CC-220, the newly developed CELMoD CC-92480, which is currently in clinical trials for relapsed and refractory MM, was also identified as an alternative potent compound for inducing the degradation of DCD23-tagged proteins (Hansen *et al.*, 2020). Despite the fact that DCD23 (60 residues) was less than half the size of DCD19 (131 residues) and only contained half of Ikaros ZF1 and ZF3 domains, it induced a comparably potent and robust degradation of cMYC-EGFP following CC-220 treatment. This suggests that the interactions of the chimeric ZF2 domain with the IMiD/CELMoD-bound CRBN complex could be stronger than those of Aiolos and Ikaros ZF2 domains, therefore compensating for the absence of additional ZF

domains forming further interactions with CRBN surface residues. Indeed, this is supported by the data from the cellular degradation assay on the longer DCD23 versions (DCD24 and DCD25) based on the full Ikaros ZF1 and ZF3 domains, which failed to induce a significantly more potent loss of the target protein following IMiD/CELMoD treatment. Moreover, the degradation of cMYC-EGFP fused to DCD23 (ZF1  $\alpha$ -helix, chimeric ZF2 and ZF3  $\beta$ -hairpin loop) following CC-220 treatment was identified to be more potent compared to non-chimeric degron DCDs of similar size and corresponding origin (Aiolos-based DCD11, ZF1  $\alpha$ -helix, ZF2 and ZF3  $\beta$ -hairpin loop; and Ikaros-based DCD21, ZF1  $\alpha$ -helix, ZF2 and ZF3  $\beta$ -hairpin loop). Furthermore, a previous *in vitro* study has identified a stretch of three consecutive arginine residues in ZF3 (IKZF1 Arg183 to Arg185) which when replaced with the corresponding amino acids in ZF1 (IGP), the affinity of the ZF2-ZF3 for the IMiD/CELMoD-CRBN complex was reduced (Sievers *et al.*, 2018). Even though ZF2 is the primary determinant for CRBN recruitment, these three consecutive arginine residues also present in both DCD19 and DCD23 could contribute to high-affinity binding to the IMiD/CELMoD-CRBN complex, thus partly explaining the more potent degradation of the target protein compared to DCDs lacking them. From all tested sets, DCD23 was identified as the best candidate for developing iTAG, as it induced a strong and selective degradation of the target protein upon IMiD/CELMoD treatment and was also based on a short amino acid sequence (60 residues) to minimise steric hindrances when fused to a target protein. Collectively, the data presented here are in line with previous studies, validating that the geometric arrangement of a  $\beta$ -hairpin loop with a critical glycine at the apex rather than its consensus sequence is what mediates binding to the IMiD/CELMoD-CRBN complex

(Krönke *et al.*, 2014; Matyskiela *et al.*, 2016; Petzold, Fischer and Thomä, 2016; Sievers *et al.*, 2018). In addition, the data presented in this thesis further expand our understanding of what constitutes a functional degron by highlighting the importance of additional residues or domains flanking the  $\beta$ -hairpin loop. The number of new consensus sequence properties constituting a functional degron motif in known CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates elucidated here pave the way for further studies to confirm these empirical observations utilising *in silico* methods, structural approaches such as x-ray crystallography and electron microscopy, and site-directed mutagenesis assays.

The iTAG system was also identified as a versatile chemical-biology tool by demonstrating an acute, potent, and robust loss of the nuclear protein EZH2 and the cytoplasmic protein KRAS[G12V], tagged at their N- or C-termini with either DCD19 or DCD23. Contrary to the 25mer Aiolos ZF2 domain-based degron, which was previously shown to induce loss of only C-terminally tagged target proteins, both DCD19 and DCD23 induced a potent and selective degradation of N-terminally tagged targets (Koduri *et al.*, 2019). This could be a limiting factor in the use of the 25mer Aiolos ZF2 domain-based degron for certain target proteins where C-terminal tagging disrupts the function. Further to EZH2 and KRAS[G12V], the iTAG system was also demonstrated to induce the degradation of a number of other targets, in between nuclear (RBM39, CDK9 and MAPK9), cytoplasmic (HRPT1 and SMS), and mitochondrial (MAVS) proteins. These target proteins were N-terminally tagged with DCD23, without EGFP, demonstrating that DCDs can induce degradation when directly fused to a target. In addition, target protein degradation was also shown to be more potent than the previously developed dTAG system. Indeed, a robust degradation for all

target proteins was achieved following treatment with nanomolar concentrations of CC-220 for 4h, compared to partial degradation of all target proteins achieved by the dTAG system. Interestingly, the degradation efficiency for CDK9, HPRT1 and SMS, was slightly reduced at high nanomolar concentrations of dTAG13 treatment, a situation that was previously observed with many PROTACs and has been termed the “hook effect” (Bondeson *et al.*, 2018; Paiva and Crews, 2019; Pettersson and Crews, 2019). At high concentrations, PROTACs saturate binding to the target and to the E3 ubiquitin ligase resulting in the formation of binary complexes instead of the productive ternary complex (E3 ubiquitin ligase – PROTAC - target protein), effectively preventing the ubiquitination and degradation of the target protein (Moreau *et al.*, 2020). Collectively, this data shows that iTAG is a versatile tool for degrading target proteins with different subcellular compartmentalizations.

Modulation of target expression is important for assessing the function and relevance of a target protein both in basic research, and during the drug discovery process of target validation (Hughes *et al.*, 2011). The applicability of the iTAG system as a chemical-biology tool for studying target protein function was evaluated by exploring the functional consequences following degradation of both cMYC and KRAS[G12V] under overexpression conditions. Both targets have been extensively studied and several of their cellular functions have been well-characterised, thus constituting them as ideal proof-of-concept targets to demonstrate that the iTAG system can be employed for studying the functions of proteins (Amati and Sabo, 2014; McMahon, 2014; Muñoz-maldonado, Zimmer and Medová, 2019). To demonstrate that iTAG-tagged cMYC is functional, the expression levels of the cyclin-dependent kinase inhibitor p21 were selected as

a read-out, since cMYC was previously shown to act as a transcriptional repressor for *CDKN1A* (p21 gene locus) (Gartel *et al.*, 2001). CC-220-mediated degradation of iTAG-tagged cMYC induced a concomitant increase in p21 levels, which were reversed following CC-220 washout, thus demonstrating the reversibility of the iTAG system. This constitutes a major advantage in the use of the iTAG system over irreversible genetic systems that require target cDNA re-expression through gene transfer experiments to rescue the phenotypic consequences, as rapidly reversing the phenotype by washout of the IMiD/CELMoD provides a high degree of confidence in the functional role of the target protein. Furthermore, the reversibility of the iTAG system allows researchers to study the function of targets whose genetic KO would otherwise cause cell lethality. Regarding iTAG-tagged KRAS[G12V], the phosphorylation levels of MEK1 (T292) were examined as a downstream marker for KRAS[G12V] activation (Eblen *et al.*, 2004). However, KRAS[G12V] was functional only when DCD19 or DCD23 were fused at its N-terminus possibly because C-terminal fusions affect farnesylation of the KRAS CAAX motif, which is critical for translocation to the cell membrane and initiation of the RAF/MEK/ERK signalling pathway (Choy *et al.*, 1999). This highlights the importance of evaluating both N- and C-terminal fusions to the target protein when employing the iTAG system and further signifies the importance of utilizing degron motifs that are functional when fused to either terminus of the target protein, unlike the previously developed 25mer Aiolos ZF2 domain-based degron (Koduri *et al.*, 2019).

The iTAG system has several advantages compared to existing chemical-biology tools. First, it incorporates a relevant control for compound off-target effects by employing the use of viral P2A sequences to allow for a separate

translation of target protein and degron motif. The use of these controls is important, as they allow the differentiation between observed phenotypes that arise from IMiD/CELMoD off-target effects and functional consequences that result from target protein loss. Second, the iTAG system benefits from the fact that a single genetic modification (DCD fusion to a target) is required to form a functioning targeted protein degradation system, as the CRL4<sup>CRBN</sup> E3 ubiquitin ligase is ubiquitously expressed (J.J *et al.*, 2004; Xin *et al.*, 2008; Schapira *et al.*, 2019). Moreover, the use of potent IMiDs/CELMoDs that induce target protein degradation when administered at nanomolar concentrations confers an additional advantage for utilising the iTAG system. For comparison, the following AID-based systems AID-ARF16 and minilAA7 require high doses (micromolar concentrations) of auxin compounds to achieve desired degradation of target proteins, with the additional exogenous overexpression of the plant TIR1 E3 ubiquitin ligase to form a functioning degradation machinery (Nishimura *et al.*, 2009). Since the iTAG system relies on hijacking the cell's intrinsic degradation machinery, this obviates the need to express additional UPS machinery. Third, protein degradation is achieved upon IMiD/CELMoD administration compared to the DD system, where ligand removal results in target protein loss (Banaszynski *et al.*, 2006). This is particularly relevant in cases where prolonged expression of a target protein is required before the study, as continued ligand treatment is necessary to maintain target protein expression, which can be expensive (Takara Biosciences #632188 £927.00 for 5 mg of Shield-1; Selleck Chemicals #S8760, £136.00 for 5 mg of CC-220). Therefore, it is desirable to have a system where the compounds would induce loss of target protein, as in the case for iTAG. Fourth, removal of target protein with the iTAG system is through well-defined



proteasomal degradation pathways whilst the mechanisms that lead to target protein loss when utilizing chemical-biology tools like the SMASh-tag and DD systems remain elusive. Indeed, even though the SMASh-tag system has been previously demonstrated to mediate target protein degradation through a combination of proteolytic and autophagy pathways and the DD system via the proteasome, the exact UPS machinery involved in the degradation of the targets under investigation have not been determined, yet (Banaszynski *et al.*, 2006; Chung *et al.*, 2015). The ill-defined processes of target protein removal for these systems could introduce an extra layer of complexity and variability when defining the functional consequences following loss of the target protein. Lastly, the DCD23 degron motif employed by the iTAG system is significantly smaller (7kDa) compared to current well-established degron motifs [SMASh-tag (34 kDa), HaloTAG7 (33 kDa), AID tag (25 kDa) and FKBP12<sup>F36V</sup> (12 kDa)] (Nishimura *et al.*, 2009; Chung *et al.*, 2015; England, Luo and Cai, 2015; Nabet *et al.*, 2018). This could be important in fusions with target proteins that cannot tolerate large tags to either their termini or with target proteins that form part of scaffolding complexes. A small tag such as DCD23 might reduce the risk of steric hindrances therefore possibly not affecting target protein function. However, it should also be stated that one of the main limitations of the iTAG system is that it cannot be applied for the study of target proteins in haematological cancer cell lines, as IMiDs/CELMoDs exhibit cytotoxic effects (Alan *et al.*, 2005; Fenaux *et al.*, 2011; Richardson *et al.*, 2002, 2006, 2014; Hagner *et al.*, 2015; Cubillos-Zapata *et al.*, 2016; Bjorklund *et al.*, 2020; Hansen *et al.*, 2020, 2021; Polizzotto *et al.*, 2021; Surka *et al.*, 2021). Moreover, the iTAG system has currently only been tested in

---

a limited number of different cellular backgrounds and more work is required to prove its applicability for the functional study of proteins in other cell lines.

In parallel to developing of the iTAG system, Tatsuya Sawasaki and colleagues developed a chemical-biology tool, based on the ZF2 domain of the ZF transcription factor SALL4 (spalt-like transcription factor 4), a previously identified CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrate (Donovan *et al.*, 2018; Matyskiela *et al.*, 2018, 2020; Yamanaka *et al.*, 2020). The 28mer SALL4 ZF2-based degron was named S4D and was demonstrated to degrade target proteins in various subcellular compartments, in-between the nucleus, plasma membrane and the mitochondrial outer membrane. Although the S4D degron was not included in our cellular degradation assay, it was reported in this study to induce degradation of a target protein at comparable levels to previously characterised 25mer Aiolos ZF2-based degron (Koduri *et al.*, 2019; Yamanaka *et al.*, 2020). Contrary to the 25mer degron, this system was also shown to induce degradation of N-terminally tagged target proteins (Koduri *et al.*, 2019; Yamanaka *et al.*, 2020). Their study did not demonstrate whether the S4D degron is functional in the in vivo setting, although the S4D degron was successfully inserted at the genomic locus of two targets by employing the CRISPR/Cas9 genome editing tool in HeLa cells. Rel-like domain-containing protein RELA/p65, a component of the NF- $\kappa$ B transcription factor complex and I $\kappa$ B $\alpha$  (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha), which functions as an inhibitor of NF- $\kappa$ B transcription factor were endogenously tagged with S4D degron and demonstrated to be degraded in a dose and time-dependent manner following IMiD-treatment (Yamanaka *et al.*, 2020).

The in vivo data presented in this thesis further support the use of the iTAG system for exploring the functions of target proteins in animal models, as well. MDA-MB-231 cells exogenously expressing Fluc-EGFP-DCD23 were orthotopically transplanted in the fat mammary pads of immunocompromised mice and allowed to form tumours. Subsequent administration of CC-220 led to the degradation of the target protein and a concomitant loss of bioluminescence, within 4h of treatment. Such rapid in vivo degradation kinetics of a target protein have not been previously described for the 25mer Aiolos ZF2-based degron, which was only shown to induce in vivo degradation following treatment with 4 consecutive doses of pomalidomide for 48h (Koduri *et al.*, 2019). As stated above, the S4D system remains to be tested in the in vivo setting and the PK/PD properties of the optimal degrader 5-HydroxyThalidomide need to be also established to ascertain its effectiveness in inducing in vivo degradation of target proteins (Yamanaka *et al.*, 2020). Moreover, to the best of my knowledge, the following chemical-biology tools, HaloTag-based systems and the SMASh-tag method, have not been implemented in mouse studies, yet (Los *et al.*, 2008; Neklesa *et al.*, 2011; Buckley *et al.*, 2015; Chung *et al.*, 2015; England, Luo and Cai, 2015; Tomoshige *et al.*, 2015; Tovell *et al.*, 2019; Caine *et al.*, 2020). A key advantage for the iTAG system as a suitable tool for in vivo functional evaluation is that the IMiD and CELMoD agents tested here are well characterized and optimized in terms of their efficacy, PK/PD properties and toxicity profile, as they are either currently used for the treatment of several haematological malignancies or are in human clinical studies (Singhal *et al.*, 1999; Richardson *et al.*, 2002, 2006, 2014; Cubillos-Zapata *et al.*, 2016; Matyskiela *et al.*, 2017; Holstein, Hillengass and McCarthy, 2018; Bjorklund *et al.*, 2020). Moreover,

IMiDs/CELMoDs like thalidomide can cross the blood-brain barrier which would facilitate the employment of iTAG system for brain studies in animal models (Muscal *et al.*, 2012; Li *et al.*, 2013). This highlights a possible advantage of IMiDs/CELMoDs over PROTACs for in vivo studies as the latter cannot cross the blood brain barrier due to their relatively high molecular weight and topological polar surface area. This, in combination with the use of the P2A controls could constitute iTAG an attractive tool for mouse research as any systemic IMiD/CELMoD off-target effects could be accounted for. However, one thing to note is that IMiDs/CELMoDs cause teratogenicity (McBride, 1961; Lenz, 1962; Vargesson, 2015). The observed embryonic malformations have been previously attributed to the IMiD/CELMoD-dependent degradation of neosubstrates that drive key developmental processes during embryogenesis (Donovan *et al.*, 2018; Matyskiela *et al.*, 2018; Asatsuma-Okumura *et al.*, 2019; Yamanaka *et al.*, 2021). Although the IMiD/CELMoD teratogenicity effects in animal embryo models, could limit their in vivo application for some animal models, this is not the case for mice. This is because mouse Crbn harbours a pair of polymorphisms proximal to the tri-tryptophan pocket, namely, a valine to isoleucine substitution and a glutamate to valine substitution (V388I, E377V, residues correspond to human CRBN) (Chamberlain *et al.*, 2014). This valine to isoleucine substitution (V388I) in mouse Crbn was previously verified to render mouse cells insensitive to lenalidomide-induced CK1 $\alpha$  and Ikaros degradation (Krönke *et al.*, 2015). Likewise, the single amino acid modification E377V in mouse Crbn was also confirmed to inhibit degradation of GSPT1 by CC-885 (Matyskiela *et al.*, 2016). These mouse Crbn polymorphisms underline the absence of thalidomide-induced embryopathies in rodents, constituting the use of the iTAG system ideal

for both adult and embryo murine studies (Matyskiela *et al.*, 2018). Moreover, since IMiD/CELMoD-binding to mouse Crbn does not induce the degradation of mouse CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates, this makes the effects of these compounds specific to the transplanted human cells, in the in vivo study. Furthermore, a humanised mouse model is also now available, where a single amino acid substitution in murine Crbn (CRBN<sup>I391V</sup>) was demonstrated to sensitize murine CD4<sup>+</sup> T cells to IMiD/CELMoD treatment, which would otherwise be resistant (Gemechu *et al.*, 2018). Such an animal model could be employed for studying murine target proteins in different disease models with IMiDs/CELMoDs, which was previously not possible.

Recently, research efforts have also been concentrated on developing CRL4<sup>CRBN</sup> E3 ubiquitin ligase - IMiDs/CELMoDs – based chemical-biology tools to be applied at the clinical stage, and in particular, in conjunction with chimeric antigen receptor (CAR) T cells (Carbonneau *et al.*, 2021; Jan *et al.*, 2021). CAR T cells are currently used in cancer immunotherapy and are characterised by a genetic modification to produce an artificial T-cell receptor which recognises a specific cell surface antigen (Sadelain, Rivière and Riddell, 2017). Despite the clinical success of CAR T cells in B cell malignancies, they are limited in their use due to toxicity arising from T cell hyperactivation (Sadelain, Rivière and Riddell, 2017). To address this, earlier this year, Max Jan, Marcela Maus, Benjamin Ebert, and colleagues published a study where they engineered CAR T cells with ON and OFF switches under the control of lenalidomide (Jan *et al.*, 2021). The ON system comprised a two-component split CAR controlled by lenalidomide-inducible dimerization, whilst the OFF system incorporated a degron tag into the CAR to enable lenalidomide-induced CAR degradation (Jan *et al.*, 2021). The

group identified a hybrid ZFP91-Aiolos degron, which they termed super-degron, from a library of 22 hybrid C2H2 ZFs that were shown to bind CRBN in the presence of an IMiD (Jan *et al.*, 2021). Furthermore, these ON and OFF switch systems were demonstrated to be functional *in vivo*, by displaying antitumour activity in mice engrafted with mantle cell lymphoma cells expressing the CAR-specific antigen CD19 (Jan *et al.*, 2021). Similarly, a study by Seth Carbonneau, Sujata Sharma, Carla Guimaraes, and colleagues demonstrated that a 60 residues-long Aiolos-based degron fused to the cytosolic C-terminus of a CAR in Jurkat cells mediated the degradation of CAR following lenalidomide treatment (Carbonneau *et al.*, 2021). This system was also shown to be functional in a mouse orthotopic xenograft model of acute lymphoblastic leukaemia, with lenalidomide treatment regulating expression and activity of degron-tagged CARs (Carbonneau *et al.*, 2021). In summary, both studies, demonstrated the potential of CRL4<sup>CRBN</sup> E3 ubiquitin ligase – IMiD/CELMoD-based approaches for regulating the expression levels of CAR T cell receptors in the clinical setting. The current work presented in this thesis offers additional degron motifs and a larger repertoire of IMiDs/CELMoDs from which such systems could benefit from and which could aid in the translation of IMiD/CELMoD-controlled CAR T cells to the clinical stage.

One of the key aims in the development of the iTAG system, which was not addressed in this thesis, is the endogenous knock-in (KI) of DCD23 at the genomic locus of a target protein coupled with the use of an IMiD/CELMoD for acute degradation. This is important as it would allow the functional study of a target within a more physiologically relevant context. Endogenous KI was performed with the S4D system using a CRISPR/Cas9-mediated KI approach

described above (Yamanaka *et al.*, 2020). A novel CRISPR/Cas9 KI method, namely the PITCh (Precise Integration into Target Chromosome) system, was also successfully applied in the dTAG system for locus-specific KI of the FKBP12<sup>F36V</sup> degron to a target gene (Nabet *et al.*, 2018). The PITCh system would constitute a good initial approach to combine a CRISPR/Cas9-mediated KI tool with iTAG as its KI effectiveness has been previously demonstrated with the dTAG system (Nabet *et al.*, 2018). This strategy employs MMEJ (microhomology-mediated end-joining) to achieve selective KI of a degron motif and does not require large homology arms. The PITCh system is based on the construction of two vectors, a donor vector containing the degron motif flanked by very short (5-25 bp) microhomologous sequences, as homology arms, and a programmable nuclease vector encoding the Cas9 and the sgRNAs, followed by co-transfection of these vectors into cultured cell lines and subsequent drug selection of KI cells (Sakuma *et al.*, 2015). Alternatively to the PITCh system, Juan Carlos Izpisua Belmonte's group recently developed two novel methods of CRISPR/Cas9-mediated KI, namely the HITI (homology-independent targeted integration) and SATI (intercellular linearized Single homology Arm donor mediated intron-Targeting Integration) (Suzuki *et al.*, 2016, 2019). The HITI strategy allows for efficient targeted KI in both dividing and non-dividing cells in vitro and in vivo. It is based on the NHEJ (non-homologous end joining) pathway to deliver a donor DNA sequence that lacks a homology arm and is designed to include a Cas9 cleavage site flanking the donor sequence to a specific genomic locus (Suzuki *et al.*, 2016). Contrary, the SATI strategy mediates KI of a DNA donor sequence with a single homology arm via both NHEJ and HDR (Suzuki *et al.*, 2019).

In summary, herein the development of the iTAG system is presented, a novel chemical-biology tool based on a series of DCDs from various CRL4<sup>CRBN</sup> E3 ubiquitin ligase neosubstrates. The iTAG system can be employed as a simple, highly modular, robust, and complimentary method to existing tools for studying the functions of target proteins, both in vitro and in vivo. DCD23 has been identified as the most optimal DCD, based on a short amino acid sequence to minimise steric hindrances when fused to target protein and an observed strong and selective target protein degradation upon IMiD/CELMoD treatment. A generalizable approach to engineer targeted protein degradation; a lentiviral system to exogenously express DCD23 fused to a target protein is also presented. The iTAG system therefore constitutes a transformational modular and versatile tool to specifically degrade proteins of interest, which will enable systematic exploration of protein function across the human proteome.



---

## References

- Adli, M. (2018) 'The CRISPR tool kit for genome editing and beyond', *Nature Communications*. Springer US, 9(1911), pp. 1–13. doi: 10.1038/s41467-018-04252-2.
- Alan, L. *et al.* (2005) 'Efficacy of Lenalidomide in Myelodysplastic Syndromes', *The New England Journal of Medicine*, 352(6), pp. 549–557.
- Amati, B. and Sabo, A. (2014) 'Genome Recognition by MYC', *Cold Spring Harbor Perspectives in Medicine*, pp. 1–14.
- Amerik, A. Y. and Hochstrasser, M. (2004) 'Mechanism and function of deubiquitinating enzymes', *Biochimica et Biophysica Acta*, 1695, pp. 189–207. doi: 10.1016/j.bbamcr.2004.10.003.
- An, J., Ponthier, Charles M., *et al.* (2017) 'PSILAC mass spectrometry reveals ZFP91 as IMiD-dependent substrate of the CRL4 CRBN ubiquitin ligase', *Nature Communications*. Nature Publishing Group, 8(1), pp. 1–11. doi: 10.1038/ncomms15398.
- An, J., Ponthier, Charles M., *et al.* (2017) 'pSILAC mass spectrometry reveals ZFP91 as IMiD-dependent substrate of the CRL4CRBN ubiquitin ligase', *Nature Communications*. Nature Publishing Group, 8(15398), pp. 1–11. doi: 10.1038/ncomms15398.
- An, W. *et al.* (2015) 'Engineering FKBP-Based Destabilizing Domains to Build Sophisticated Protein Regulation Systems', *PLOS One*, 10(12), pp. 1–12. doi: 10.1371/journal.pone.0145783.
- Ananthkrishnan, R. and Menon, S. (2013) 'Design of oncology clinical trials : A review', *Critical Reviews in Oncology / Hematology*. Elsevier Ireland Ltd, 88(1), pp. 144–153. doi: 10.1016/j.critrevonc.2013.03.007.
- Anders, C. *et al.* (2014) 'Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease', *Nature*. Nature Publishing Group, 513,

- pp. 569–573. doi: 10.1038/nature13579.
- Armstrong, B. Y. J. W. (1999) 'A review of high-throughput screening approaches for drug discovery', pp. 26–28.
- Armstrong, C. M. and Goldberg, D. E. (2007) 'An FKBP destabilization domain modulates protein levels in *Plasmodium falciparum*', *Nature Methods*, 4(12), pp. 1007–1009. doi: 10.1038/NMETH1132.
- Arpad, M. and Georgopoulos, K. (1994) 'The Ikaros Gene Encodes a Family of Functionally Diverse Zinc Finger DNA-Binding Proteins', *Molecular and Cellular Biology*, 14(12), pp. 8292–8303. doi: 10.1128/MCB.14.12.8292.
- Asatsuma-Okumura, T. *et al.* (2019) 'p63 is a cereblon substrate involved in thalidomide teratogenicity', *Nature Chemical Biology*. Springer US, 15, pp. 1077–1083. doi: 10.1038/s41589-019-0366-7.
- Atkinson, J. C. and Clark, D. B. (2016) 'Design and interpretation of clinical research studies in oral medicine : a brief review', *Oral Diseases*, 22, pp. 87–92. doi: 10.1111/odi.12385.
- Banaszynski, L. A. *et al.* (2006) 'A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules', *Cell*, 126(5), pp. 995–1004. doi: 10.1016/j.cell.2006.07.025.
- Bard, J. A. M. *et al.* (2018) 'Structure and Function of the 26S Proteasome', *Annu. Rev. Biochem.*, 87, pp. 697–724.
- Behan, F. M. *et al.* (2019) 'Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens', *Nature*. Springer US, 568(7753), pp. 511–516. doi: 10.1038/s41586-019-1103-9.
- Bennardo, N. *et al.* (2008) 'Alternative-NHEJ Is a Mechanistically Distinct Pathway of Mammalian Chromosome Break Repair', *PLOS Genetics*, 4(6), pp. 1–10. doi: 10.1371/journal.pgen.1000110.
- Bjorklund, C. C. *et al.* (2020) 'Iberdomide ( CC-220 ) is a potent cereblon E3 ligase modulator with antitumor and immunostimulatory activities in lenalidomide- and pomalidomide-resistant multiple myeloma cells with dysregulated CRBN', *Leukemia*, 34, pp. 1197–1201. doi: 10.1038/s41375-019-0620-8.

- Boettcher, M. and Mcmanus, M. T. (2015) 'Choosing the Right Tool for the Job : RNAi, TALEN, or CRISPR', *Molecular Cell*. Elsevier Inc., 58(4), pp. 575–585. doi: 10.1016/j.molcel.2015.04.028.
- Bondeson, D. P. *et al.* (2018) 'Lessons in PROTAC Design from Selective Degradation with a Promiscuous Warhead', *Cell Chemical Biology*. Elsevier Ltd., 25(1), pp. 78-87.e5. doi: 10.1016/j.chembiol.2017.09.010.
- Bonger, K. M. *et al.* (2011) 'Small-molecule displacement of a cryptic degron causes conditional protein degradation', *Nature Chemical Biology*. Nature Publishing Group, 7(8), pp. 531–537. doi: 10.1038/nchembio.598.
- Braten, O. *et al.* (2016) 'Numerous proteins with unique characteristics are degraded by the 26S proteasome following monoubiquitination', *Proceedings of the National Academy of Sciences of the United States of America*, 113(32), pp. E4639–E4647. doi: 10.1073/pnas.1608644113.
- Buckley, D. L. *et al.* (2015) 'HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins', *ACS Chemical Biology*, 10(8), pp. 1831–1837. doi: 10.1021/acscchembio.5b00442.
- C. Glenn Begley and Lee M. Ellis (2012) 'Raise standards for preclinical cancer research', *Nature*, 483, pp. 531–533. doi: 10.2165/00128413-199509970-00015.
- Cai, Y. *et al.* (2009) 'A Brief Review on the Mechanisms of miRNA Regulation', *Genomics Proteomics Bioinformatics*. Beijing Institute of Genomics, 7(4), pp. 147–154. doi: 10.1016/S1672-0229(08)60044-3.
- Caine, E. A. *et al.* (2020) 'Targeted Protein Degradation Phenotypic Studies Using HaloTag CRISPR/Cas9 Endogenous Tagging Coupled with HaloPROTAC3', *Current Protocols in Pharmacology*, 91(1), pp. 1–21. doi: 10.1002/cpph.81.
- Cao, R. *et al.* (2002) 'Role of histone H3 lysine 27 methylation in polycomb-group silencing', *Science*, 298(5595), pp. 1039–1043. doi: 10.1126/science.1076997.
- Carbonneau, S. *et al.* (2021) 'An IMiD-inducible degron provides reversible regulation for chimeric antigen receptor expression and activity', *Cell Chemical Biology*. Elsevier Ltd., 28, pp. 1–11. doi: 10.1016/j.chembiol.2020.11.012.

- Carthew, R. W. and Sontheimer, E. J. (2009) 'Origins and Mechanisms of miRNAs and siRNAs', *Cell*. Elsevier Inc., 136(4), pp. 642–655. doi: 10.1016/j.cell.2009.01.035.
- Caussinus, E., Kanca, O. and Affolter, M. (2012) 'Fluorescent fusion protein knockout mediated by anti-GFP nanobody', *Nature Structural and Molecular Biology*. Nature Publishing Group, 19(1), pp. 117–122. doi: 10.1038/nsmb.2180.
- Chamberlain, P. P. *et al.* (2014) 'Structure of the human Cereblon – DDB1 – lenalidomide complex reveals basis for responsiveness to thalidomide analogs', *Nature Publishing Group*. Nature Publishing Group, 21(9), pp. 803–810. doi: 10.1038/nsmb.2874.
- Charpentier, E. and Marraffini, L. A. (2014) 'Harnessing CRISPR-Cas9 immunity for genetic engineering', *Current Opinion in Microbiology*. Elsevier Ltd, 19, pp. 114–119. doi: 10.1016/j.mib.2014.07.001.
- Chen, L. and Madura, K. (2002) 'Rad23 Promotes the Targeting of Proteolytic Substrates to the Proteasome', *Molecular and Cellular Biology*, 22(13), pp. 4902–4913. doi: 10.1128/MCB.22.13.4902.
- Chen, W., Werdann, M. and Zhang, Y. (2018) 'The auxin-inducible degradation system enables conditional PERIOD protein depletion in the nervous system of *Drosophila melanogaster*', *FEBS Journal*, 285(23), pp. 4378–4393. doi: 10.1111/febs.14677.
- Chen, X. *et al.* (2019) 'Degradation of endogenous proteins and generation of a null-like phenotype in zebrafish using Trim-Away technology', *Genome Biology*. Genome Biology, 20(19), pp. 1–6.
- Chen, Z. J. and Sun, L. J. (2009) 'Nonproteolytic Functions of Ubiquitin in Cell Signaling', *Molecular Cell*. Elsevier Inc., 33(3), pp. 275–286. doi: 10.1016/j.molcel.2009.01.014.
- Chessari, G. and Woodhead, A. J. (2009) 'From fragment to clinical candidate — a historical perspective', *Drug Discovery Today*, 14, pp. 668–675. doi: 10.1016/j.drudis.2009.04.007.
- Chessum, N. E. A. *et al.* (2018) 'Demonstrating In-Cell Target Engagement Using

- a Pirin Protein Degradation Probe (CCT367766)', *Journal of Medicinal Chemistry*, 61(3), pp. 918–933. doi: 10.1021/acs.jmedchem.7b01406.
- Choy, E. *et al.* (1999) 'Endomembrane Trafficking of Ras : The CAAX Motif Targets Proteins to the ER and Golgi', *Cell*, 98, pp. 69–80.
- Chung, H. K. *et al.* (2015) 'Tunable and reversible drug control of protein production via a self-excising degron', *Nature Chemical Biology*. Nature Publishing Group, 11(9), pp. 713–720. doi: 10.1038/nchembio.1869.
- Ciechanover, A., Finley, D. and Varshavsky, A. (1984) 'Ubiquitin Dependence of Selective Protein Degradation Demonstrated in the Mammalian Cell Cycle Mutant ts85', *Cell*, 37, pp. 57–66.
- Clackson, T. *et al.* (1998) 'Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity', *Proceedings of the National Academy of Sciences of the United States of America*, 95(18), pp. 10437–10442. doi: 10.1073/pnas.95.18.10437.
- Clague, M. J., Urbe, S. and Komander, D. (2019) 'Breaking the chains : deubiquitylating enzyme specificity begets function', *Nature Reviews Molecular Cell Biology*. Springer US, 20(June), pp. 338–352. doi: 10.1038/s41580-019-0099-1.
- Clift, D. *et al.* (2017) 'A Method for the Acute and Rapid Degradation of Endogenous Proteins', *Cell*. Elsevier Inc., 172, pp. 1692–1706. doi: 10.1016/j.cell.2017.10.033.
- Clift, D. *et al.* (2018) 'Acute and rapid degradation of endogenous proteins by Trim-Away', *Nature Protocols*. Springer US, 13(October), pp. 2149–2175. doi: 10.1038/s41596-018-0028-3.
- Cobb, B. S. *et al.* (2000) 'Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding', *Genes and Development*, 14, pp. 2146–2160. doi: 10.1101/gad.816400.to.
- Collins, I. *et al.* (2017) 'Chemical approaches to targeted protein degradation through modulation of the ubiquitin – proteasome pathway', *Biochemical Journal*, 474, pp. 1127–1147. doi: 10.1042/BCJ20160762.

- Le Cong, F. A. R. *et al.* (2013) 'Multiplex Genome Engineering Using CRISPR/Cas Systems', *Science*, 339, pp. 819–824.
- Crews, C. M. (2010) 'Crosstalk Targeting the Undruggable Proteome: The Small Molecules of My Dreams Crosstalk', *Chemistry & Biology*. Elsevier Ltd, 17(6), pp. 551–555. doi: 10.1016/j.chembiol.2010.05.011.
- Croston, G. E. (2017) 'The utility of target-based discovery', *Expert Opinion on Drug Discovery*. Taylor & Francis, 12(5), pp. 427–429. doi: 10.1080/17460441.2017.1308351.
- Cubillos-Zapata, C. *et al.* (2016) 'CC-122 immunomodulatory effects in refractory patients with diffuse large B-cell lymphoma', *Oncoimmunology*. Taylor & Francis, 5(12), pp. 1–5. doi: 10.1080/2162402X.2016.1231290.
- Cyr, D. M., Höhfeld, J. and Patterson, C. (2002) 'Protein quality control: U-box-containing E3 ubiquitin ligases join the fold', *Trends in Biochemical Sciences*, 27(7), pp. 368–375. doi: 10.1016/S0968-0004(02)02125-4.
- Czermin, B. *et al.* (2002) 'Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites', *Cell*, 111(2), pp. 185–196. doi: 10.1016/S0092-8674(02)00975-3.
- D'Amato, R. J. *et al.* (1994) 'Thalidomide is an inhibitor of angiogenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 91(9), pp. 4082–4085. doi: 10.1073/pnas.91.9.4082.
- Dang, C. V (2013) 'MYC, Metabolism, Cell Growth, and Tumorigenesis', *Cold Spring Harbor Perspectives in Medicine*, pp. 1–16.
- Daniel, K. *et al.* (2018) 'Conditional control of fluorescent protein degradation by an auxin-dependent nanobody', *Nature Communications*. Springer US, 9(3297), pp. 1–13. doi: 10.1038/s41467-018-05855-5.
- Dempster, J. M. *et al.* (2019) 'Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets', *Nature Communications*. Springer US, 10(5817), pp. 1–14. doi: 10.1038/s41467-019-13805-y.
- Dikic, I. (2017) 'Proteasomal and Autophagic Degradation Systems', *Annu. Rev. Biochem.*, 86, pp. 193–224.

- Donovan, K. A. *et al.* (2018) 'Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane Radial Ray syndrome', *eLife*, 7, pp. 1–25.
- Ea, C. *et al.* (2006) 'Activation of IKK by TNF $\alpha$  Requires Site-Specific Ubiquitination of RIP1 and Polyubiquitin Binding by NEMO', *Molecular Cell*, 22, pp. 245–257. doi: 10.1016/j.molcel.2006.03.026.
- Eblen, S. T. *et al.* (2004) 'Mitogen-Activated Protein Kinase Feedback Phosphorylation Regulates MEK1 Complex Formation and Activation during Cellular Adhesion', *Molecular and Cellular Biology*, 24(6), pp. 2308–2317. doi: 10.1128/mcb.24.6.2308-2317.2004.
- Ege, N. *et al.* (2021) 'Phenotypic screening with target identification and validation in the discovery and development of E3 ligase modulators', *Cell Chemical Biology*. Elsevier Ltd., 28(3), pp. 283–299. doi: 10.1016/j.chembiol.2021.02.011.
- Elbashir, S. M. *et al.* (2001) 'Duplexes of 21 - nucleotide RNAs mediate RNA interference in cultured mammalian cells', *Nature*, 411(May), pp. 494–498.
- Elrod-erickson, M. *et al.* (1996) 'Zif268 protein – DNA complex refined at 1 . 6 Å : a model system for understanding zinc finger – DNA interactions', *Cell Structure*, 4(10), pp. 1171–1180.
- England, C. G., Luo, H. and Cai, W. (2015) 'HaloTag Technology: A Versatile Platform for Biomedical Applications', *Bioconjugate Chemistry*, 26(6), pp. 975–986. doi: 10.1021/acs.bioconjchem.5b00191.
- Erb, M. A. *et al.* (2017) 'Transcription control by the ENL YEATS domain in acute leukaemia', *Nature*. Nature Publishing Group, 543(7644), pp. 270–274. doi: 10.1038/nature21688.
- Errington, T. M. *et al.* (2014) 'An open investigation of the reproducibility of cancer biology research', *eLife*, 3, pp. 1–9. doi: 10.7554/eLife.04333.
- Fedotova, A. A. *et al.* (2017) 'C2H2 Zinc Finger Proteins : The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors', *Acta Naturae*, 9(33), pp. 47–58.

- Fenaux, P. *et al.* (2011) 'Plenary paper A randomized phase 3 study of lenalidomide versus placebo in RBC transfusion-dependent patients with Low- / Intermediate-1-risk myelodysplastic syndromes with del5q', *Blood*, 118(14), pp. 3765–3776. doi: 10.1182/blood-2011-01-330126.An.
- Fennell, M. *et al.* (2014) 'Impact of RNA-Guided Technologies for Target Identification and Deconvolution', *Journal of Biomolecular Screening*, 19(10), pp. 1327–1337. doi: 10.1177/1087057114548414.
- Ferreira, L. L. G. and Andricopulo, A. D. (2019) 'ADMET modeling approaches in drug discovery', *Drug Discovery Today*. Elsevier Ltd, 24(5), pp. 1157–1165. doi: 10.1016/j.drudis.2019.03.015.
- Fire, A. *et al.* (1998) 'Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*', *Nature*, 391(February), pp. 806–811.
- Fischer, E. S. *et al.* (2014) 'Structure of the DDB1 – CRBN E3 ubiquitin ligase in complex with thalidomide', *Nature*, 512, pp. 49–53. doi: 10.1038/nature13527.
- Fu, Y. *et al.* (2013) 'High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells', *Nature Biotechnology*. Nature Publishing Group, 31(9), pp. 822–826. doi: 10.1038/nbt.2623.
- Fuchs, O. (2017) 'Immunomodulatory drugs and their therapeutic effect in hematological malignancies through cereblon', *Hematology & Medical Oncology*, 2(3), pp. 1–8. doi: 10.15761/HMO.1000129.
- Fulcher, L. *et al.* (2017) 'Targeting endogenous proteins for degradation through the affinity-directed protein missile system', *Open Biology*, 7, pp. 1–11. doi: 10.1016/j.chembiol.2020.06.012.
- Fulcher, L. J. *et al.* (2016) 'An affinity-directed protein missile system for targeted proteolysis', *Open biology*, 6(10). doi: 10.1098/rsob.160255.
- Gandhi, A. K. *et al.* (2013) 'Immunomodulatory agents lenalidomide and pomalidomide co-stimulate T cells by inducing degradation of T cell repressors Ikaros and Aiolos via modulation of the E3 ubiquitin ligase complex CRL4 CRBN', *British Journal of Haematology*, 164, pp. 811–821. doi: 10.1111/bjh.12708.
- Ganoth, D. *et al.* (1988) 'A Multicomponent System That Degrades Proteins



- Conjugated to Ubiquitin', *Journal of Biological Chemistry*. © 1988 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 263(25), pp. 12412–12419. doi: 10.1016/S0021-9258(18)37771-8.
- Gartel, A. L. *et al.* (2001) 'MYC represses the p21 (WAF1/CIP1) promoter and interacts with Sp1 / Sp3', *PNAS*, 98(8), pp. 4510–4515.
- Gemechu, Y. *et al.* (2018) 'Humanized cereblon mice revealed two distinct therapeutic pathways of immunomodulatory drugs', *PNAS*, 115(46), pp. 11802–11807. doi: 10.1073/pnas.1814446115.
- Ghobrial, I. M. and Rajkumar, S. V. (2003) 'Management of thalidomide toxicity.', *The journal of supportive oncology*, 1(3), pp. 194–205.
- Gilbert, L. A. *et al.* (2013) 'CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes', *Cell*. Elsevier Inc., 154(2), pp. 442–451. doi: 10.1016/j.cell.2013.06.044.
- Hadian, K. *et al.* (2011) 'NF- $\kappa$ B Essential Modulator ( NEMO ) Interaction with Linear and Lys-63 Ubiquitin Chains Contributes to NF- $\kappa$ B Activation', *Journal of Biological Chemistry*. © 2011 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 286(29), pp. 26107–26117. doi: 10.1074/jbc.M111.233163.
- Haglund, K. *et al.* (2003) 'Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation', *Nature Cell Biology*, 5(5), pp. 461–466. doi: 10.1038/ncb983.
- Hagner, P. R. *et al.* (2015) 'CC-122 , a pleiotropic pathway modifier , mimics an interferon response and has antitumor activity in DLBCL', *Blood*, 126(6), pp. 779–789. doi: 10.1182/blood-2015-02-628669.The.
- Hajduk, P. J. and Greer, J. (2007) 'A decade of fragment-based drug design: strategic advances and lessons learned', *Nature Reviews Drug Discovery*, 6, pp. 211–219. doi: 10.1038/nrd2220.
- Hannah, J. and Zhou, P. (2015) 'Distinct and overlapping functions of the cullin E3 ligase scaffolding proteins CUL4A and CUL4B', *Gene*. Elsevier B.V., 573(1),

pp. 33–45. doi: 10.1016/j.gene.2015.08.064.

Hanpude, P. *et al.* (2015) 'Deubiquitinating Enzymes in Cellular Signaling and Disease Regulation Enzymatic Mechanism of DUBs', *IUBMB LIFE*, pp. 544–555. doi: 10.1002/iub.1402.

Hansen, J. D. *et al.* (2020) 'Discovery of CRBN E3 Ligase Modulator CC-92480 for the Treatment of Relapsed and Refractory Multiple Myeloma', *Journal of Medicinal Chemistry*, 63, pp. 6648–6676. doi: 10.1021/acs.jmedchem.9b01928.

Hansen, J. D. *et al.* (2021) 'Cc-90009: A cereblon e3 ligase modulating drug that promotes selective degradation of gspt1 for the treatment of acute myeloid leukemia', *Journal of Medicinal Chemistry*, 64, pp. 1835–1843. doi: 10.1021/acs.jmedchem.0c01489.

Hatakeyama, S. *et al.* (2001) 'U Box Proteins as a New Family of Ubiquitin-Protein Ligases', *Journal of Biological Chemistry*. © 2001 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 276(35), pp. 33111–33120. doi: 10.1074/jbc.M102755200.

He, F. *et al.* (2017) 'Predicting siRNA efficacy based on multiple selective siRNA representations and their combination at score level', *Nature Scientific Reports*. Nature Publishing Group, 7(44836), pp. 1–13. doi: 10.1038/srep44836.

Heilker, R., Lessel, U. and Bischoff, D. (2019) 'The power of combining phenotypic and target-focused drug discovery', *Drug Discovery Today*. Elsevier Ltd, 24(2), pp. 526–532. doi: 10.1016/j.drudis.2018.10.009.

Hershko, A. *et al.* (1980) 'Proposed role of ATP in protein breakdown: Conjugation of proteins with multiple chains of the polypeptide of ATP-dependent proteolysis', *PNAS*, 77(4), pp. 1783–1786.

Hershko, A. *et al.* (1983) 'Components of Ubiquitin-Protein Ligase System', *Journal of Biological Chemistry*. © 1983 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 258(13), pp. 8206–8214. doi: 10.1016/S0021-9258(20)82050-X.

Hicke, L. (2001) 'Protein regulation by monoubiquitin', *Nature Reviews Molecular*

- Cell Biology*, 2(3), pp. 195–201. doi: 10.1038/35056583.
- Holstein, S. A., Hillengass, J. and McCarthy, P. L. (2018) 'Next-Generation Drugs Targeting the Cereblon Ubiquitin Ligase', *Journal of Clinical Oncology*, 36(20), pp. 2101–2104. doi: 10.1200/JCO.2018.77.9637.
- Holstein, S. A. and Medical, N. (2021) 'Next-Generation Drugs Targeting the Cereblon Ubiquitin Ligase', 36(20), pp. 2101–2104. doi: 10.1200/JCO.2018.77.9637.
- Hsu, P. D., Lander, E. S. and Zhang, F. (2014) 'Development and Applications of CRISPR-Cas9 for Genome Engineering', *CELL*. Elsevier, 157(6), pp. 1262–1278. doi: 10.1016/j.cell.2014.05.010.
- Huang, H. T. *et al.* (2017) 'MELK is not necessary for the proliferation of basal-like breast cancer cells', *eLife*, 6, pp. 1–29. doi: 10.7554/eLife.26693.
- Hughes, J. P. *et al.* (2011) 'Principles of early drug discovery', *British Journal of Pharmacology*, 162(6), pp. 1239–1249. doi: 10.1111/j.1476-5381.2010.01127.x.
- Iorns, E. *et al.* (2007) 'Utilizing RNA interference to enhance cancer drug discovery', *nature Reviews*, 6(July), pp. 556–568. doi: 10.1038/nrd2355.
- Ito, T. *et al.* (2010a) 'Identification of a Primary Target of Thalidomide Teratogenicity', *Science*, 327(March), pp. 1345–1351.
- Ito, T. *et al.* (2010b) 'Identification of a primary target of thalidomide teratogenicity', *Science*. *Science*, 327(5971), pp. 1345–1350. doi: 10.1126/science.1177319.
- J.J, H. *et al.* (2004) 'A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation', *Neurology*, 63(10), pp. 1927–1931.
- Jackson, A. L. *et al.* (2006) 'Widespread siRNA " off-target " transcript silencing mediated by seed region sequence complementarity', *RNA*, 12(7), pp. 1179–1187. doi: 10.1261/rna.25706.gene.
- Jacobson, A. D. *et al.* (2009) 'The Lysine 48 and Lysine 63 Ubiquitin Conjugates Are Processed Differently by the 26 S Proteasome', *Journal of Biological Chemistry*. © 2009 ASBMB. Currently published by Elsevier Inc; originally

published by American Society for Biochemistry and Molecular Biology., 284(51), pp. 35485–35494. doi: 10.1074/jbc.M109.052928.

Jaeger, M. G. *et al.* (2020) 'Selective Mediator dependence of cell-type-specifying transcription', *Nature Genetics*. Springer US, 52(July), pp. 719–727. doi: 10.1038/s41588-020-0635-0.

Jaeger, M. G. and Winter, G. E. (2021) 'Fast-acting chemical tools to delineate causality in transcriptional control', *Molecular Cell*. Elsevier Inc., 81(8), pp. 1617–1630. doi: 10.1016/j.molcel.2021.02.015.

Jan, M. *et al.* (2021) 'Reversible ON- and OFF-switch chimeric antigen receptors controlled by lenalidomide', *Science Translational Medicine*, 13, pp. 1–14.

Jiang, S. *et al.* (2018) 'Casein kinase 1  $\alpha$ : biological mechanisms and theranostic potential', *BMC Cell Communication and Signaling*. Cell Communication and Signaling, 16(23), pp. 1–24.

Jinek, M. *et al.* (2012) 'A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337(August), pp. 816–822.

Kaelin, W. G. (2017) 'Common pitfalls in preclinical cancer target validation', *Nature Reviews Cancer*. Nature Publishing Group, 17(7), pp. 441–450. doi: 10.1038/nrc.2017.32.

Kanke, M. *et al.* (2011) 'Auxin-inducible protein depletion system in fission yeast', *BMC Cell Biology*, 12, pp. 1–16. doi: 10.1186/1471-2121-12-8.

Khan, A. A. *et al.* (2009) 'Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs', *Nature Biotechnology*, 27(6), pp. 549–671. doi: 10.1038/nbt.1543.

Khattabi, L. El *et al.* (2019) 'A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers', *CELL*, 178, pp. 1145–1158. doi: 10.1016/j.cell.2019.07.011.

Kim, H. K. *et al.* (2016) 'Cereblon in health and disease', *Pflügers Archiv - European Journal of Physiology*. Pflügers Archiv - European Journal of Physiology, 468, pp. 1299–1309. doi: 10.1007/s00424-016-1854-1.

- Kim, J. H. *et al.* (2011) 'High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines , Zebrafish and Mice', *PLOS ONE*, 6(4), pp. 1–8. doi: 10.1371/journal.pone.0018556.
- Kim, S. S. *et al.* (2009) 'Strategies for targeted nonviral delivery of siRNAs in vivo', *Trends in Molecular Medicine*, 15(11), pp. 491–500. doi: 10.1016/j.molmed.2009.09.001.
- Kloe, G. E. De *et al.* (2009) 'Transforming fragments into candidates : small becomes big in medicinal chemistry', *Drug Discovery Today*, 14, pp. 630–646. doi: 10.1016/j.drudis.2009.03.009.
- Koduri, V. *et al.* (2019) 'Peptidic degron for IMiD-induced degradation of heterologous proteins', *Proceedings of the National Academy of Sciences of the United States of America*, 116(7), pp. 2539–2544. doi: 10.1073/pnas.1818109116.
- Komander, D. and Rape, M. (2012) 'The Ubiquitin Code', *Annu. Rev. Biochem.*, 81, pp. 203–229. doi: 10.1146/annurev-biochem-060310-170328.
- Komor, A. C. *et al.* (2016) 'Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage', *Nature*. Nature Publishing Group, 533(7603), pp. 420–424. doi: 10.1038/nature17946.
- Kosicki, M., Tomberg, K. and Bradley, A. (2018) 'Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements', *Nature Biotechnology*, 36(8). doi: 10.1038/nbt.4192.
- Krönke, J. *et al.* (2014) 'Lenalidomide Causes Selective Degradation of IKZF1 and IKZF3 in Multiple Myeloma Cells', *Science*, 343, pp. 301–305.
- Krönke, J. *et al.* (2015) 'Lenalidomide induces ubiquitination and degradation of CK1 $\alpha$  in del(5q) MDS', *nature*, 523, pp. 183–188. doi: 10.1038/nature14610.
- Kuscu, C. *et al.* (2014) 'Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease', *Nature Biotechnology*. Nature Publishing Group, 32(7), pp. 677–683. doi: 10.1038/nbt.2916.
- Kuzmichev, A. *et al.* (2002) 'Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of zeste protein', *Genes*

- and Development*, 16(22), pp. 2893–2905. doi: 10.1101/gad.1035902.
- Kwon, Y. T. and Ciechanover, A. (2017) 'The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy', *Trends in Biochemical Sciences*. Elsevier Ltd, 42(11), pp. 873–886. doi: 10.1016/j.tibs.2017.09.002.
- Laugesen, A., Højfeldt, J. W. and Helin, K. (2019) 'Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation', *Molecular Cell*. Elsevier Inc., 74(1), pp. 8–18. doi: 10.1016/j.molcel.2019.03.011.
- Lea, W. A. and Simeonov, A. (2011) 'Fluorescence Polarization Assays in Small Molecule Screening', *Expert Opinion on Drug Discovery*, 6(1), pp. 17–32. doi: 10.1517/17460441.2011.537322.Fluorescence.
- Lee, J. and Zhou, P. (2007) 'DCAFs , the Missing Link of the CUL4-DDB1 Ubiquitin Ligase', *Molecular Cell*, 26, pp. 775–780. doi: 10.1016/j.molcel.2007.06.001.
- Lenz W, K. K. (1962) 'Foetal malformations due to Thalidomide.', *German Med. Monthly*, 7, pp. 253–258.
- Li, S. *et al.* (2019) 'An efficient auxin-inducible degron system with low basal degradation in human cells', *Nature Methods*. Springer US, 16(9), pp. 866–869. doi: 10.1038/s41592-019-0512-x.
- Li, T. *et al.* (2006) 'Structure of DDB1 in Complex with a Paramyxovirus V Protein : Viral Hijack of a Propeller Cluster in Ubiquitin Ligase', *CELL*, 124, pp. 105–117. doi: 10.1016/j.cell.2005.10.033.
- Li, Z. *et al.* (2013) 'Pomalidomide Shows Significant Therapeutic Activity against CNS Lymphoma with a Major Impact on the Tumor Microenvironment in Murine Models', *PLOS One*, 8(8), pp. 1–18. doi: 10.1371/journal.pone.0071754.
- Lipinski, C. A. *et al.* (2001) 'Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings', *Advanced Drug Delivery Reviews*, 46, pp. 3–26. doi: 10.1016/j.addr.2012.09.019.
- Liu, T. *et al.* (2017) 'Thalidomide and its analogues: A review of the potential for immunomodulation of fibrosis diseases and ophthalmopathy', *Experimental and*

- Therapeutic Medicine*, 14(6), pp. 5251–5257. doi: 10.3892/etm.2017.5209.
- Lopez-Girona, A. *et al.* (2012) 'Cereblon is a direct protein target for immunomodulatory and antiproliferative activities of lenalidomide and pomalidomide.', *Leukemia*. Nature Publishing Group, 26(May), pp. 2326–2335. doi: 10.1038/leu.2012.119.
- Los, G. V *et al.* (2008) 'HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis', *ACS Chemical Biology*, 3(6), pp. 373–382.
- Lu, G. *et al.* (2014) 'The Myeloma Drug Lenalidomide Promotes the Cereblon-Dependent Destruction of Ikaros Proteins', *Science*, 343, pp. 305–309.
- Lu, J. *et al.* (2021) 'Types of nuclear localization signals and mechanisms of protein import into the nucleus', *Cell Communication and Signaling*. BioMed Central, 19(60), pp. 1–10. doi: 10.1186/s12964-021-00741-y.
- Madeira da Silva, L. *et al.* (2009) 'Regulated expression of the Leishmania major surface virulence factor lipophosphoglycan using conditionally destabilized fusion proteins', *PNAS*, 106(18), pp. 7583–7588.
- Mali, P. *et al.* (2013) 'RNA-Guided Human Genome Engineering via Cas9', *Science*, 339(February), pp. 823–827.
- Matyskiela, M. E. *et al.* (2016) 'A novel cereblon modulator recruits GSPT1 to the CRL4 CRBN ubiquitin ligase', *Nature*. Nature Publishing Group, 535, pp. 252–235. doi: 10.1038/nature18611.
- Matyskiela, M. E. *et al.* (2017) 'A Cereblon Modulator (CC-220) with Improved Degradation of Ikaros and Aiolos', *Journal of Medicinal Chemistry*, 61, pp. 535–542. doi: 10.1021/acs.jmedchem.6b01921.
- Matyskiela, M. E. *et al.* (2018) 'SALL4 mediates teratogenicity as a thalidomide-dependent cereblon substrate', *Nature Chemical Biology*. Springer US, 14(October), pp. 981–987. doi: 10.1038/s41589-018-0129-x.
- Matyskiela, M. E. *et al.* (2020) 'Crystal structure of the SALL4 – pomalidomide – cereblon – DDB1 complex', *Nature Structural & Molecular Biology*. Springer US, 27(April), pp. 319–322. doi: 10.1038/s41594-020-0405-9.

- McMahon, S. B. (2014) 'MYC and the Control of Apoptosis', *Cold Spring Harbor Perspectives in Medicine*, pp. 1–10.
- Metzger, M. B., Hristova, V. A. and Weissman, A. M. (2010) 'HECT and RING finger families of E3 ubiquitin ligases at a glance', *Journal of Cell Science*, 125(3), pp. 531–537. doi: 10.1242/jcs.091777.
- Mikuma, T. *et al.* (2004) 'Overexpression of Dicer enhances RNAi-mediated gene silencing by short-hairpin RNAs ( shRNAs ) in human cells', *Nucleic Acids Symposium Series No.48*, (48), pp. 191–192.
- Miyakawa, T. (2020) 'No raw data , no science : another possible source of the reproducibility crisis', *Molecular Brain*. *Molecular Brain*, 13(24), pp. 1–6.
- Moffat, J. G. *et al.* (2017) 'Opportunities and challenges in phenotypic drug discovery: an industry perspective', *Nature Publishing Group*. *Nature Publishing Group*, 16(8), pp. 531–543. doi: 10.1038/nrd.2017.111.
- Moffat, J. G., Rudolph, J. and Bailey, D. (2014) 'Phenotypic screening in cancer drug discovery — past , present and future', *Nature Publishing Group*. *Nature Publishing Group*, 13(August), pp. 588–602. doi: 10.1038/nrd4366.
- Moreau, K. *et al.* (2020) 'Proteolysis-targeting chimeras in drug development : A safety perspective', *Br J Pharmacol.*, 177(November 2019), pp. 1709–1718. doi: 10.1111/bph.15014.
- Muhar, M. *et al.* (2018) 'SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis', *Science*, 360(May), pp. 800–805.
- Müller, J. *et al.* (2002) 'Histone methyltransferase activity of a Drosophila Polycomb group repressor complex', *Cell*, 111(2), pp. 197–208. doi: 10.1016/S0092-8674(02)00976-5.
- Muñoz-maldonado, C., Zimmer, Y. and Medová, M. (2019) 'A Comparative Analysis of Individual RAS Mutations in Cancer Biology', *Frontiers in Oncology*, 9(October), pp. 1–22. doi: 10.3389/fonc.2019.01088.
- Muscal, J. A. *et al.* (2012) 'Plasma and cerebrospinal fluid pharmacokinetics of thalidomide and lenalidomide in nonhuman primates', *Cancer Chemother Pharmacol*, 69, pp. 943–947. doi: 10.1007/s00280-011-1781-y.



- Nabet, B. *et al.* (2018) 'The dTAG system for immediate and target-specific protein degradation', *Nature Chemical Biology*. Springer US, 14(5), pp. 431–441. doi: 10.1038/s41589-018-0021-8.
- Nabet, B. *et al.* (2020) 'Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules', *Nature Communications*. Springer US, 11(1). doi: 10.1038/s41467-020-18377-w.
- Nabet, B. (2021) 'Charting a New Path Towards Degrading Every Protein', *ChemBioChem*, 22, pp. 483–484. doi: 10.1002/cbic.202000531.
- Napoli, C., Lemieux, C. and Jorgensen, R. (1990) 'Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans', *The Plant Cell*, 2(April), pp. 279–289.
- Natsume, T. *et al.* (2016) 'Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors', *Cell Reports*. The Authors, 15(1), pp. 210–218. doi: 10.1016/j.celrep.2016.03.001.
- Neklesa, T. K. *et al.* (2011) 'Small-molecule hydrophobic tagging-induced degradation of HaloTag fusion proteins', *Nature Chemical Biology*. Nature Publishing Group, 7(8), pp. 538–543. doi: 10.1038/nchembio.597.
- Neklesa, T. K., Winkler, J. D. and Crews, C. M. (2017) 'Pharmacology & Therapeutics Targeted protein degradation by PROTACs ☆', *Pharmacology and Therapeutics*. Elsevier Inc., 174, pp. 138–144. doi: 10.1016/j.pharmthera.2017.02.027.
- Nguyen, T. Van *et al.* (2016) 'Glutamine Triggers Acetylation-Dependent Degradation of Glutamine Synthetase via the Thalidomide Receptor Cereblon', *Molecular Cell*. Elsevier, 61, pp. 809–820. doi: 10.1016/j.molcel.2016.02.032.
- Nishimura, K. *et al.* (2009) 'An auxin-based degron system for the rapid depletion of proteins in nonplant cells', *Nature Methods*. Nature Publishing Group, 6(12), pp. 917–922. doi: 10.1038/nmeth.1401.
- Olive, J. F. *et al.* (2018) 'Accounting for tumor heterogeneity when using CRISPR-Cas9 for cancer progression and drug sensitivity studies', *PLOS ONE*, 13(6), pp. 1–23.

- Paddison, P. J. *et al.* (2002) 'Short hairpin RNAs ( shRNAs ) induce sequence-specific silencing in mammalian cells', *Genes and Development*, pp. 948–958. doi: 10.1101/gad.981002.a.
- Paiva, S. and Crews, C. M. (2019) 'Targeted protein degradation : elements of PROTAC design', *Current Opinion in Chemical Biology*. Elsevier Ltd, 50, pp. 111–119. doi: 10.1016/j.cbpa.2019.02.022.
- Paroo, Z. and Corey, D. R. (2004) 'Challenges for RNAi in vivo', *TRENDS in Biotechnology*, 22(8), pp. 390–394. doi: 10.1016/j.tibtech.2004.06.004.
- Pengelly, A. R. *et al.* (2013) 'A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb', *Science*, 339(February), pp. 698–700.
- Petroski, M. D. and Deshaies, R. J. (2005) 'Function and regulation of cullin-RING ubiquitin ligases', *Nature Reviews Molecular Cell Biology*, 6(1), pp. 9–20. doi: 10.1038/nrm1547.
- Pettersson, M. and Crews, C. M. (2019) 'PROteolysis TArgeting Chimeras (PROTACs) — Past, present and future', *Drug Discovery Today: Technologies*. doi: 10.1016/j.ddtec.2019.01.002.
- Petzold, G., Fischer, E. S. and Thomä, N. H. (2016) 'Structural basis of lenalidomide-induced CK1 $\alpha$  degradation by the CRL4 CRBN ubiquitin ligase', *Nature*, 532, pp. 127–130. doi: 10.1038/nature16979.
- Pickar-Oliver, A. and Gersbach, C. A. (2019) 'The next generation of CRISPR–Cas technologies and applications', *Nature Reviews Molecular Cell Biology*. Springer US, 20(8), pp. 490–507. doi: 10.1038/s41580-019-0131-5.
- Pickart, C. M. and Eddins, M. J. (2004) 'Ubiquitin : structures , functions , mechanisms', 1695, pp. 55–72. doi: 10.1016/j.bbamcr.2004.09.019.
- Polizzotto, M. N. *et al.* (2021) 'Pomalidomide for Symptomatic Kaposi ' s Sarcoma in People With and Without HIV Infection : A Phase I / II Study', *Journal of Clinical Oncology*, 34, pp. 4125–4131. doi: 10.1200/JCO.2016.69.3812.
- Prinz, F., Schlange, T. and Asadullah, K. (2011) 'Believe it or not: How much can we rely on published data on potential drug targets?', *Nature Reviews Drug*

*Discovery*. Nature Publishing Group, 10(9), pp. 712–713. doi: 10.1038/nrd3439-c1.

Qi, L. S. *et al.* (2013) 'Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression', *Cell*. Elsevier, 152, pp. 1173–1183. doi: 10.1016/j.cell.2013.02.022.

Quach, H. *et al.* (2010) 'Mechanism of action of immunomodulatory drugs ( IMiDS ) in multiple myeloma', *Leukemia*. Nature Publishing Group, 24(October 2009), pp. 22–32. doi: 10.1038/leu.2009.236.

Ran, F. A. *et al.* (2013) 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols*, 8(11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.

Rao, S. S. P. *et al.* (2017) 'Cohesin Loss Eliminates All Loop Domains', *Cell*. Elsevier Inc., 171(2), pp. 305–320. doi: 10.1016/j.cell.2017.09.026.

Rath, D. *et al.* (2015) 'The CRISPR-Cas immune system: Biology, mechanisms and applications', *Biochimie*. Elsevier B.V, 117, pp. 119–128. doi: 10.1016/j.biochi.2015.03.025.

Richardson, P. G. *et al.* (2002) 'Plenary paper Immunomodulatory drug CC-5013 overcomes drug resistance and is well tolerated in patients with relapsed multiple myeloma', 100(9), pp. 3063–3067. doi: 10.1182/blood-2002-03-0996.Supported.

Richardson, P. G. *et al.* (2006) 'NEOPLASIA A randomized phase 2 study of lenalidomide therapy for patients with relapsed or relapsed and refractory multiple myeloma', 108(10), pp. 3458–3464. doi: 10.1182/blood-2006-04-015909.Supported.

Richardson, P. G. *et al.* (2014) 'Pomalidomide alone or in combination with low-dose dexamethasone in relapsed and refractory multiple myeloma: a randomized phase 2 study', 123(12), pp. 1826–1832. doi: 10.1182/blood-2013-11-538835.POM.

Röth, S. *et al.* (2020) 'Targeting Endogenous K-RAS for Degradation through the Affinity-Directed Protein Missile System', *Cell Chemical Biology*, 27(9), pp. 1151–1163. doi: 10.1016/j.chembiol.2020.06.012.

Ryu, M. Y. *et al.* (2019) 'Classification of barley U-box E3 ligases and their

- expression patterns in response to drought and pathogen stresses', *BMC Genomics*. BMC Genomics, 20(1), pp. 1–15. doi: 10.1186/s12864-019-5696-z.
- S., A. *et al.* (2006) 'Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery', *Nature*, 443(October), pp. 3–6. doi: 10.1038/nature05175.
- Sadelain, M., Rivière, I. and Riddell, S. (2017) 'Therapeutic T cell engineering', *Nature*. Nature Publishing Group, 545(7655), pp. 423–431. doi: 10.1038/nature22395.
- Sakuma, T. *et al.* (2015) 'MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems', *Nature Protocols*. Nature Publishing Group, 11(1), pp. 118–133. doi: 10.1038/nprot.2015.140.
- Sanson, K. R. *et al.* (2018) 'Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities', *Nature Communications*. Springer US, 9(5416), pp. 1–15. doi: 10.1038/s41467-018-07901-8.
- Sarikas, A., Hartmann, T. and Pan, Z. Q. (2011) 'The cullin protein family', *Genome Biology*, 12(4), pp. 1–12. doi: 10.1186/gb-2011-12-4-220.
- Sathyan, K. M. *et al.* (2019) 'An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion', *Genes and Development*, 33(19–20), pp. 1441–1455. doi: 10.1101/gad.328237.119.
- Sauer, H., Gu, J. and Wartenberg, M. (2000) 'Thalidomide Inhibits Angiogenesis in Embryoid Bodies by the Generation of Hydroxyl Radicals', *American Journal of Pathology*, 156(1), pp. 151–158. doi: 10.1016/S0002-9440(10)64714-1.
- Schapira, M. *et al.* (2019) 'Targeted protein degradation: expanding the toolbox', *Nature Reviews Drug Discovery*. Springer US, 18(12), pp. 949–963. doi: 10.1038/s41573-019-0047-y.
- Schneider, R. K. *et al.* (2014) 'Role of Casein Kinase 1A1 in the Biology and Targeted Therapy of del(5q) MDS', *Cancer Cell*, 26(October), pp. 509–520. doi: 10.1016/j.ccr.2014.08.001.
- Schweitzer, A. *et al.* (2016) 'Structure of the human 26S proteasome at a

resolution of 3.9 Å', *PNAS*, 113(28), pp. 7816–7821. doi: 10.1073/pnas.1608050113.

Scrima, A. *et al.* (2008) 'Structural Basis of UV DNA-Damage Recognition by the DDB1-DDB2 Complex', *Cell*, 135(7), pp. 1213–1223. doi: 10.1016/j.cell.2008.10.045.

Seibel, N. M. *et al.* (2007) 'Nuclear localization of enhanced green fluorescent protein homomultimers', *Analytical Biochemistry*, 368, pp. 95–99. doi: 10.1016/j.ab.2007.05.025.

Shabalina, S. A., Spiridonov, A. N. and Ogurtsov, A. Y. (2006) 'Computational models with thermodynamic and composition features improve siRNA design', *BMC Bioinformatics*, 7(65), pp. 1–16. doi: 10.1186/1471-2105-7-65.

Sharma, B. and Taganna, J. (2020) 'Genome-wide analysis of the U-box E3 ubiquitin ligase enzyme gene family in tomato', *Scientific Reports*, 10(1), pp. 1–15. doi: 10.1038/s41598-020-66553-1.

Sheereen, A. *et al.* (2017) 'A missense mutation in the CRBN gene that segregates with intellectual disability and self-mutilating behaviour in a consanguineous Saudi family', *Journal of Medical Genetics*, 54, pp. 236–240. doi: 10.1136/jmedgenet-2016-104117.

Sheskin (1965) 'Further Observation with Thalidomide in Leprosy Reactions', *Leprosy Review*, 36(4), pp. 83–85.

Shin, H. Y. *et al.* (2017) 'CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome', *Nature Communications*. Nature Publishing Group, 8(May), pp. 1–10. doi: 10.1038/ncomms15464.

Shin, Y. J. *et al.* (2015) 'Nanobody-targeted E3-ubiquitin ligase complex degrades nuclear proteins', *Scientific reports*. Nature Publishing Group, 5(14269), pp. 1–11. doi: 10.1038/srep14269.

Sievers, Q. L. *et al.* (2018) 'Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN', *Science*, 362(558), pp. 1–9. doi: 10.1126/science.aat0572.

Sigoillot, F. D. and King, R. W. (2011) 'Vigilance and Validation: Keys to Success

in RNAi Screening', *ACS Chemical Biology*, 6(1), pp. 47–60.

Simpson, L. M. *et al.* (2020) 'Inducible Degradation of Target Proteins through a Tractable Affinity-Directed Protein Missile System', *Cell Chemical Biology*, 27(9), pp. 1164–1180.e5. doi: 10.1016/j.chembiol.2020.06.013.

Singhal, S. *et al.* (1999) 'ANTITUMOR ACTIVITY OF THALIDOMIDE IN REFRACTORY MULTIPLE MYELOMA', *The New England Journal of Medicine*, 341(21), pp. 1565–1571.

Sneeringer, C. J. *et al.* (2010) 'Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 ( H3K27 ) in human B-cell lymphomas', *PNAS*, 107(49), pp. 20980–20985. doi: 10.1073/pnas.1012525107.

Song, H. W. *et al.* (2015) 'shRNA off-target effects in vivo: Impaired endogenous siRNA expression and spermatogenic defects', *PLoS ONE*, 10(3), pp. 1–23. doi: 10.1371/journal.pone.0118549.

Stengel, K. R. *et al.* (2021) 'Definition of a small core transcriptional circuit regulated by AML1-ETO', *Molecular Cell*. Elsevier Inc., 81(3), pp. 530–545. doi: 10.1016/j.molcel.2020.12.005.

Sternberg, S. H. *et al.* (2014) 'DNA interrogation by the CRISPR RNA-guided endonuclease Cas9', *Nature*. Nature Publishing Group, 507, pp. 62–67. doi: 10.1038/nature13011.

Surka, C. *et al.* (2021) 'CC-90009, a novel cereblon E3 ligase modulator, targets acute myeloid leukemia blasts and leukemia stem cells', *Blood*. American Society of Hematology, 137(5), pp. 661–677. doi: 10.1182/blood.2020008676.

Suzuki, K. *et al.* (2016) 'In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration', *Nature*, 450(December 2016), pp. 144–149. doi: 10.1038/nature20565.

Suzuki, K. *et al.* (2019) 'Precise in vivo genome editing via single homology arm donor mediated intron-targeting gene integration for genetic disease correction', *Cell Research*. Springer US, 29(August), pp. 804–819. doi: 10.1038/s41422-019-0213-0.

- Teale, W. D., Paponov, I. A. and Palme, K. (2006) 'Auxin in action: Signalling, transport and the control of plant growth and development', *Nature Reviews Molecular Cell Biology*, 7(11), pp. 847–859. doi: 10.1038/nrm2020.
- Tomoshige, S. *et al.* (2015) 'Degradation of HaloTag-fused nuclear proteins using bestatin-HaloTag ligand hybrid molecules', *Organic and Biomolecular Chemistry*. Royal Society of Chemistry, 13(38), pp. 9746–9750. doi: 10.1039/c5ob01395j.
- Toure, M. and Crews, C. M. (2016) 'Small-molecule PROTACS: New approaches to protein degradation', *Angewandte Chemie - International Edition*, 55(6), pp. 1966–1973. doi: 10.1002/anie.201507978.
- Tovell, H. *et al.* (2019) 'Rapid and Reversible Knockdown of Endogenously Tagged Endosomal Proteins via an Optimized HaloPROTAC Degradator', *ACS Chemical Biology*, pp. 882–892. doi: 10.1021/acscchembio.8b01016.
- Trost, M., Blattner, A. C. and Lehner, C. F. (2016) 'Regulated protein depletion by the auxin-inducible degradation system in *Drosophila melanogaster*', *Fly*. Taylor & Francis, 10(1), pp. 35–46. doi: 10.1080/19336934.2016.1168552.
- Uckun, F. M. *et al.* (2012) 'Serine phosphorylation by SYK is critical for nuclear localization and transcription factor function of Ikaros', *PNAS*, 109(44), pp. 18072–18077. doi: 10.1073/pnas.1209828109.
- Valencia-sanchez, M. A. *et al.* (2006) 'Control of translation and mRNA degradation by miRNAs and siRNAs', *Genes and Development*, 20, pp. 515–524. doi: 10.1101/gad.1399806.miRNAs.
- Vargesson, N. (2015) 'Thalidomide-Induced Teratogenesis: History and Mechanisms', *Birth Defects Research*, 105. doi: 10.1002/bdrc.21096.
- W. G. McBride (1961) 'Thalidomide and congenital abnormalities', *The Lancet*, p. 1358.
- Wang, Y. *et al.* (2020) 'HECT E3 ubiquitin ligases – emerging insights into their biological roles and disease relevance', *Journal of Cell Science*, 133, pp. 1–13. doi: 10.1242/jcs.228072.
- Waters, A. M. and Der, C. J. (2018) 'KRAS: The critical driver and therapeutic

target for pancreatic cancer', *Cold Spring Harbor Perspectives in Medicine*, 8(9), pp. 1–18. doi: 10.1101/cshperspect.a031435.

Watts, J. K., Deleavey, G. F. and Damha, M. J. (2008) 'Chemically modified siRNA: tools and applications', *Drug Discovery Today*, 13(October), pp. 842–855. doi: 10.1016/j.drudis.2008.05.007.

Weir, E. *et al.* (2021) 'Trim-Away mediated knock down uncovers a new function for Lbh during gastrulation of *Xenopus laevis* .', *Developmental Biology*, 470, pp. 74–83. doi: 10.1016/j.ydbio.2020.10.014.

Wertz, I. E. *et al.* (2004) 'De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF- $\kappa$ B signalling', *Nature*, 430(August), pp. 694–699.

Whittaker, M. *et al.* (2010) 'Fragments: Past, present and future', *Drug Discovery Today: Technologies*. Elsevier Ltd, 7(3), pp. e163–e171. doi: 10.1016/j.ddtec.2010.11.007.

Wiedenheft, B., Sternberg, S. H. and Doudna, J. A. (2012) 'RNA-guided genetic silencing systems in bacteria and archaea', *Nature*, 482(7385), pp. 331–338. doi: 10.1038/nature10886.

Wilson, R. and Doudna, J. A. (2013) 'Molecular mechanisms of RNA interference', *Annu Rev Biophys.*, 42, pp. 217–239. doi: 10.1146/annurev-biophys-083012-130404.Molecular.

Wolfe, S. A., Nekludova, L. and Pabo, C. O. (1999) 'DNA Recognition by Cys2His2 Zinc Finger Proteins', *Annu. Rev. Biophys. Biomol. Struct.*, 3, pp. 183–212.

Wu, T. *et al.* (2020) 'Targeted protein degradation as a powerful research tool in basic biology and drug target discovery', *Nature Structural and Molecular Biology*. Springer US, 27(7), pp. 605–614. doi: 10.1038/s41594-020-0438-0.

Wu, X. *et al.* (2014) 'Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells', *Nature Biotechnology*. Nature Publishing Group, 32(7), pp. 670–676. doi: 10.1038/nbt.2889.

Wu, Y. *et al.* (2020) 'A Small Molecule-Controlled Cas9 Repressible System', *Molecular Therapy: Nucleic Acids*. The Author(s), 19(March), pp. 922–932. doi:



10.1016/j.omtn.2019.12.026.

Xin, W. *et al.* (2008) 'Primary function analysis of human mental retardation related gene CRBN', *Mol Biol Rep*, 35, pp. 251–256. doi: 10.1007/s11033-007-9077-3.

Xu, G., Jiang, X. and Jaffrey, S. R. (2013) 'A Mental Retardation-linked Nonsense Mutation in Cereblon Is Rescued by Proteasome Inhibition \*', *Journal of Biological Chemistry*. © 2013 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 288(41), pp. 29573–29585. doi: 10.1074/jbc.M113.472092.

Yamanaka, S. *et al.* (2020) 'An IMiD-induced SALL4 degron system for selective degradation of target proteins', *Communications Biology*. Springer US, 3, pp. 1–14. doi: 10.1038/s42003-020-01240-5.

Yamanaka, S. *et al.* (2021) 'Thalidomide and its metabolite 5 - hydroxythalidomide induce teratogenicity via the cereblon neosubstrate PLZF', *The EMBO Journal*, pp. 1–25. doi: 10.15252/emj.2020105375.

Yang, J. *et al.* (2018) 'Cereblon suppresses the lipopolysaccharide-induced inflammatory response by promoting the ubiquitination and degradation of c-Jun', *Journal of Biological Chemistry*. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 293(26), pp. 10141–10157. doi: 10.1074/jbc.RA118.002246.

Yesbolatova, A. *et al.* (2019) 'Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole', *Methods*. Elsevier, 164–165(January), pp. 73–80. doi: 10.1016/j.ymeth.2019.04.010.

Yesbolatova, A. *et al.* (2020) 'The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice', *Nature Communications*. Springer US, 11(1). doi: 10.1038/s41467-020-19532-z.

Zee, B. M. *et al.* (2012) 'Origins and Formation of Histone Methylation across the Human Cell', *Molecular and Cellular Biology*, 32(13), pp. 2503–2514. doi: 10.1128/MCB.06673-11.

---

## References

Zhang, L. *et al.* (2015) 'The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in *C. elegans*', *Development (Cambridge)*, 142(24), pp. 4374–4384. doi: 10.1242/dev.129635.

Zheng, W., Thorne, N. and Mckew, J. C. (2013) 'Phenotypic screens as a renewed approach for drug discovery', *Drug Discovery Today*. Elsevier Ltd, 18(21–22), pp. 1067–1073. doi: 10.1016/j.drudis.2013.07.001.

Zhu, Y. X. *et al.* (2011) 'Cereblon expression is required for the antimyeloma activity of lenalidomide and pomalidomide', *Blood*, 118(18), pp. 4771–4779. doi: 10.1182/blood-2011-05-356063.The.

