

**Revalidating DNA Damage Inducible 1 Homolog II (DDI2) as a Novel Target  
for the Treatment of Cancer**

by

Ibtisam Ibtisam

A dissertation submitted to the Graduate Faculty of  
Auburn University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Auburn, Alabama  
May 6<sup>th</sup>, 2023

Keywords: DDI2, proteasome, CRISPR KO, aspartic protease, SPPS, Drug  
Discovery

Copyright 2023 by Ibtisam Ibtisam

Approved by

Dr. Alexei F. Kisselev, Chair, Professor of Drug Discovery and Development  
Dr. Forrest T. Smith, Co-chair, Professor of Drug Discovery and Development  
Dr. Randall Clark, Gilliland Professor of Drug Discovery and Development  
Dr. Jack DeRuiter, Alumni Professor of Drug Discovery and Development  
Dr. Douglas Goodwin, Professor of Department of Chemistry and Biochemistry

## **Abstract**

*Cancer cells are under proteotoxic stress and are more sensitive to the perturbation of protein quality control pathways than normal cells. The proteasome is a critical component of the protein quality control machinery. Proteasome inhibitors are approved for treating multiple myeloma, a cancer of plasma cells that produce large amounts of immunoglobulins leading to exceptionally high levels of proteotoxic stress. They have not demonstrated clinical efficacy in other cancers. Recovery of proteasome activity after bolus treatment may contribute to intrinsic and acquired resistance to proteasome inhibitors. Upon treatment with sub-toxic doses of proteasome inhibitors, transcription factor NFE2L1 is activated by a novel protease DDI2 and upregulates the proteasome genes. However, it is not known whether this transcriptional upregulation is responsible for activity recovery. We used RNAi and CRISPR KO cells in this dissertation to answer this question. Although we confirmed that DDI2 processes NFE2L1, DDI2 was not required to recover proteasome activity. The recovery depended on the translation of proteasomal mRNA, and we hypothesize that nascent proteasome subunits assemble more efficiently in the presence of proteasome inhibitors. Based on the CRISPR KO screen and transcriptome analysis, we also report transcriptional downregulation of immunoproteasome subunits and decreased immunoproteasome activity at the basal level. Finally, we describe the design and initial steps of synthesizing an activity-based probe of DDI2, a protease that cannot be assayed with traditional peptide substrates.*

## Acknowledgment

“The mind travels faster than the pen; consequently, writing becomes a question of learning to make occasional wing shots, bringing down the bird of thought as it flashes by.” — E.B. White, *The Elements of Style*.

First and always, Thank God for everything, for instilling hope in my heart when the lights are dim, and for keeping me going in challenging moments. I dedicate this Ph.D. dissertation to my parents, my Mama, and (late) Baba for their unconditional love and endless prayers, most extraordinary support, and for making me who I am today.

During my graduate research training, I would like to thank both of my research advisors for their guidance in chemistry and biology research. I am grateful that I had the opportunity to learn from both of you.

First, thank you to my primary research advisor Dr. Alexei F. Kisselev (aka. the Boss), for the opportunity to be part of your research group. Boss, thank you for giving me a place in your laboratory, teaching me cancer cell biology, and allowing me to learn so many valuable skills. Thank you, also, for your professional guidance and support during my graduate career and for training me to be a tough scientist. I appreciate your patience in helping me with valuable critiques and clear feedback during the writing and completion of this Ph.D. dissertation.

Second, thank you to my co-advisor, Dr. Forrest T. Smith, for giving me a place in your laboratory and teaching me how to do synthesis for the first time. Dr. Smith, thank you for always having time when I have questions about chemistry or my graduate study. Also, thank you for giving me wise advice and feedback while writing this Ph.D. dissertation.

My sincere thank you is extended to my committee members. Dr. Randall Clark, Dr. Jack DeRuiter, and Dr. Douglas Goodwin for their feedback and support during my graduate studies.

Thank you, Dr. Clark, for the opportunity to learn separation techniques and mass spectrometry fragmentation from you and for allowing me to use GC-MS in your laboratory. Also, thanks for the times you stopped by the laboratory and asked me how I was doing or just reminded me to use safety goggles. I appreciate it.

Thank you, Dr. DeRuiter, for instilling the logic for studying medicinal chemistry in my first semester. Your patience and time given for teaching made the transition easier for me to learn other chemistry subjects.

Thank you, Dr. Goodwin, for allowing me to learn how to purify *MtSK* enzyme in your laboratory in my second semester and for the valuable advice for my future career during the committee meeting.

Also, I sincerely appreciate Dr. Michael W. Greene (College of Human Science) for agreeing to be my Ph.D. dissertation university reader.

I want to express my appreciation to the research collaborators, past mentors, and unofficial mentors who have helped me in my research or shaped my research skills and the friends I met during my graduate school life.

Thank you to Dr. Angela I. Calderon for the opportunity to learn mass spectrometry, perform research in her laboratory, and the mentoring during my first two years in the US.

Thanks to Dr. Thankhoe Rantso'e (a classmate and former lab mate in Dr. Angela I Calderon) for mentoring me for the first time on how to use LC-MS and for simply being a good friend.

Thank you to Dr. Rene Fuanta Ngolui (former student in Dr. Douglas Goodwin's laboratory) for training me on purifying the *MtSK* enzyme for the first time. Also, for being a best friend and a brother until today, always; thank you, Rene!

Thank you to Dr. Victor T. Adebomi (a classmate and former student in Dr. Monika Raj's laboratory) for teaching me how to synthesize peptides using SPPS and purify peptides. I appreciate it!

Thanks to Dr. Ana M. Dmytrejchuk and Dr. Mansour Alturki for teaching me how to perform column chromatography, sharing tips in synthesis, and for the discussion.

Thanks to Dr. Wu Chen (a classmate and former student in Dr. Feng Li Laboratory) for allowing me to borrow some stuff for the SPSS, the discussion, and the friendship.

Thank you to the former visiting scholar in our Lab (late) Dr. Andrey V. Maksimenko, for teaching me to conduct cancer cell culture and colony-forming assays. I'm also acknowledging Sriraja's help in performing colony-forming assays. It will never go unnoticed!

Thank you to Dr. Elizabeth A. Lipke and Dr. Nicole Habbit (Department of Chemical Engineering) for conducting the 3D-growth cell culture model in my projects.

Thank you to Dr. Amit K. Mitra and Dr. Suman Mazumder for helping re-analyze the total RNA-sequencing data and conducting exome-sequencing in my projects. Also, for his valuable insight, Thank you!

Thank you, Dr. Robert D. Rusty Arnold, for allowing me to use HPLC in his laboratory. Also, I thank his students, Shanese (Lani) Jasper and Chu Zhang, for their help.

Thanks to Melissa Boersma, Director of Mass Spectrometry Lab, Department of Chemistry and Biochemistry, for conducting LC/MS and HRMS in my project.

Thanks to NMR facility – COSAM and LC Science for total RNA-sequencing.

Thanks to Dr. Robel Ghebreab for discussing the transfection protocol in 2018.

Thanks to all my lab mates: Olasubomi Akintola, Sriraja Srinivasa, Tyler Jenkins, and Mitchell Patterson, for the research discussion, support from the lab mate, and their friendship moments in the Lab. Special shout out to Ola! Thank you!

Thanks to Addison, the undergraduate working on the DDI2 project.

Thanks always to Coffee and Spotify throughout my graduate school years and the writing process of this dissertation. To Scifinder™, I don't think I could come up with things in my oral Ph.D. qualifying exam without you. I am also acknowledging Grammarly.

My utmost gratitude to my family and friends for their emotional support. To all my siblings (the Jazuli's), thank you for always believing in me regardless of the choices I made in my life. This is for us!

My never-ending appreciation to my extended family in Alabama, Susie and Mark Grantham (Mom and Dad!), a thousand times thank you will not be enough for your continuous support and kindness during my nearly 7 years studying in Auburn, for always welcoming me as part of your family and making me feel at home. I'm grateful to know both of you.

To Becky and Roger Robertson, thank you for helping me in my first year at Auburn. Also, to Dr. Nida Mckee, Dr. Michael Mckee, Jo, and Tim Frinak for being good friends in Auburn. Thank you for all the good memories.

Thanks to Auburn University, Harrison College of Pharmacy (HCOP) - Department Drug Discovery and Development (DDD) administration: Jenny, Kaleia, Kaylen, Chris, Alexis, the HCOP OIT *et al.* Also, to the faculty member In the 3<sup>rd</sup> floor Pharmacy Research Building, Dr. Amin Raj and Dr. Amal Kaddoumi. And to all the professors who have taught me valuable knowledge since my first semester until now in Pharmacy School (HCOP), Chemistry Building, or a webinar class that I cannot name individually in my dissertation, but I do wholeheartedly appreciate you!

Finally, I sincerely appreciate the Indonesia Endowment Funds for Education Ministry of Finance of Republic Indonesia (LPDP Keuangan RI) for providing me with scholarship support since my M.Sc. degree and the full four-year Ph.D. scholarship.

At last, I close this acknowledgment with a favorite quote that somehow motivated me during my Ph.D. graduate studies. "However difficult life may seem, there is always something you can do and succeed at. It matters that you don't just give up" – Stephen Hawking. I hope this quote also motivates other HCOP graduate students never to give up–Ibtisam'ibby'.

## Table of Contents

<b>Abstract</b> .....	<b>2</b>
<b>Acknowledgement</b> .....	<b>3</b>
<b>Table of Contents</b> .....	<b>6</b>
<b>List of Abbreviations</b> .....	<b>9</b>
<b>List of Tables</b> .....	<b>12</b>
<b>List of Figures</b> .....	<b>13</b>
<b>List of Schemes</b> .....	<b>16</b>
<b>Chapter 1 Introduction</b> .....	<b>17</b>
1.1 Overview of Ubiquitin-proteasome system.....	<b>19</b>
1.2 Proteasome 20S CP.....	<b>22</b>
1.3 The 19S Regulatory Particles.....	<b>23</b>
1.4 The Rationale for Targeting Proteasomes for the Treatment of Cancer.....	<b>24</b>
1.5 Proteasome Inhibitors.....	<b>26</b>
1.5.1 Bortezomib.....	<b>27</b>
1.5.2 Carfilzomib.....	<b>28</b>
1.5.3 Ixazomib.....	<b>29</b>
1.6 Regulation of proteasome expression.....	<b>30</b>
1.6.1 Regulation of proteasome genes at the transcriptional level.....	<b>31</b>
1.6.2 Regulation of proteasome genes under basal conditions.....	<b>31</b>
1.6.3 Regulation of proteasome genes in response to proteasome inhibition.....	<b>37</b>
1.7 DDI2 and its role in the proteasome activity recovery.....	<b>40</b>
1.7.1 Structure of hDDI2 and its function.....	<b>41</b>
1.7.2 The RVP domain determines the Ubiquitin endoprotease properties of DDI2.....	<b>42</b>
1.7.3 The UBL – UIMS domains contribute to hDDI2 function as protein shuttle.....	<b>46</b>
<b>Chapter 2 DDI2 is not Required for the Proteasome Activity Recovery</b> .....	<b>51</b>
2.1 Introduction.....	<b>51</b>
2.2 Results.....	<b>52</b>
2.2.1 Knockout and knockdown of DDI2 do not impair proteasome activity recovery in inhibitor-treated cells.....	<b>52</b>

2.2.2 The recovery of proteasome activity depends on the synthesis of new proteasome; however, the early onset of activity recovers before the upregulation of proteasome genes .....	57
2.2.3 The recovery of proteasome activity is mainly due to increases in the 26S/30S activity .....	64
2.3 Discussion .....	70
2.4 Material and methods .....	76
2.4.1 Chemicals (inhibitors, substrates, and other reagents).....	76
2.4.2 Cell lines and cell culture.....	77
2.4.3 siRNA experiments.....	78
2.4.4 Cell lysis and protein quantification .....	79
2.4.5 Activity Assays.....	80
2.4.6 Native gel and immunoblotting .....	83
2.4.7 RNA isolation and qPCR experiments.....	85
2.4.8 Polysome profiling .....	85
2.4.9 Statistical analysis.....	88
2.5 Supporting Information.....	89
<b>Chapter 3 CRISPR KO Screen Coupled with Transcriptome Analysis Approach to Study the Possible Activated Pathway by DDI2 .....</b>	<b>95</b>
3.1 Introduction.....	95
3.2 Results .....	96
3.2.1 DDI2 involves in the antigen presentation, xenobiotic metabolism, and cancer signaling pathway.....	99
3.2.2 DDI2 knockout does not alter the transcriptional expression of proteasome subunits but reduces the transcriptional level and activity of immunoproteasome .....	111
3.3 Discussion.....	115
3.4 Material and methods .....	120
3.4.1 Cell lines and cell culture.....	120
3.4.2 RNA isolation.....	120
3.4.3 Total RNA Sequencing (RNA-seq).....	120
3.4.4 Immunoblotting .....	122
3.4.5 Immunoproteasome activity assays .....	123
3.4.6 Statistical analysis.....	124
3.4.7 Ingenuity Pathway Analysis.....	124
3.5 Supporting Information.....	125

<b>Chapter 4 Development of Biochemical Assay to Target Aspartic Protease DDI2</b>	<b>130</b>
4.1 Introduction	130
4.2 Results and Discussion	132
4.2.1 Design, <i>in silico</i> molecular docking and synthesis of DDI2 probe	132
4.2.2 Biochemical assay of azido peptide 2	149
4.3 Material and methods	152
4.3.1 Chemicals	152
4.3.2 Instrumentation	155
4.3.3 Synthetic methods	158
4.3.4 Cell lysis, protein quantification, and bioconjugation using azido peptide 2	164
4.4 Supporting Information	166
<b>Chapter 5 Concluding Remarks and Future Directions</b>	<b>174</b>
<b>Bibliography</b>	<b>180</b>



## List of Abbreviations

19S RP	19S Regulatory Particles
20S CP	20S Core Particles
ABP	Activity-based Probe
AMC	Adenosine 5'-monophosphate, Adenosine 5'-(dihydrogen phosphate)
ATP	Adenosine 5'-triphosphate, Adenosine 5'-(tetrahydrogen triphosphate)
ARE	Anti-oxidant response element
$\beta$ -TrCP	Beta-transducin repeat containing E3 ligase
Btz	Bortezomib
bZIP	basic-leucine zipper
Cfz	Carfilzomib
CnC	Cap 'n' Collar
CPEB3	Cytoplasmic polyadenylation element 3
CTD	C-terminal domain
CRISPR	Clustered regularly interspaced short palindromic repeats
DDI2	DNA Damage Inducible 1 homolog II
DEGs	Differentially expressed genes
DEPC	Diethyl Pyrocarbonate
DPCs	Drug-induced DNA-protein crosslinks
DTT	Dithiothreitol
DUB	Deubiquitinase
ECM	Extracellular matrix protein
EDTA	Ethylenediaminetetraacetic acid
ERAD	ER-associated degradation
FBXW7	F-box protein Fbw7 E3 ligase
FOXO4	Forkhead box protein O4
GLUT1	glucose transporter
HDD	$\alpha$ -helical domain of DDI
Herpud1	Homocysteine inducible ER protein with ubiquitin-like domain 1

HSF	Heat Shock Factor
IGF	Insulin-like growth factor
IPA	Ingenuity Pathway Analysis
ISR	Integrated stress response
KEAP-1	Kelch-like ECH-associated protein 1
MCL	Mantle cell lymphoma
MES	2-(N-Morpholino)ethanesulfonic acid, C <sub>6</sub> H <sub>13</sub> NO <sub>4</sub> S
MM	Multiple myeloma
NFκB	Nuclear Factor-κB
NF-Y	Nuclear transcription factor Y
Nrf	Nuclear factor erythroid-derived 2-related factor, NFE2L
NHL	non-Hodgkin lymphoma
NTD	N-terminal domain
OGT	O-GlcNAc transferase
PAGE	Polyacrylamide gel electrophoresis
POMP	Proteasome maturation protein
PPi	Inorganic pyrophosphate
PVDF	polyvinylidene fluoride
RTF2	Replication termination factor/ c20orf43
RVP	Retroviral protease domain
SCF	Skp1-Cul1-F-box
SDS	Sodium dodecyl sulfate
sMaf	small-musculoaponeurotic fibrosarcoma
SKN-1	Skinhead 1
TBSt	Tris-Buffered Saline with 0.1% Tween®20
TCFA	β-catenin/ T-cell factor 4
TCGA	The Cancer Genome Atlas
TFs	Transcription factors
TNBC	Triple-negative breast cancer

TNF $\alpha$  Tumor necrosis factor  $\alpha$   
Top1cc Topoisomerase 1 covalent cleavage complex  
UBL Ubiquitin-like domain  
UHMK1 U2AF serine/threonine protein kinase 1 (KIS1)  
UIM Ubiquitin-interacting motifs  
UPS Ubiquitin proteasome system  
UPR Unfolded protein response  
VCP Valosin Containing Protein  
VEGF Vascular endothelial growth factor

## List of Tables

<b>Table 2.1</b> DDI2 CRISPR KO clones and their gRNA sequences.	<b>77</b>
<b>Table 2.2</b> List of primary and secondary antibodies.	<b>84</b>
<b>Table 3.1</b> Ingenuity Canonical Pathways Analysis and their Molecules.	<b>101</b>
<b>Table 3.2</b> Knockout of DDI2 showed the downregulation of MHC class I and its activators in the antigen presentation pathways.	<b>103</b>
<b>Table 3.3</b> Knockout of DDI2 showed the downregulation of several genes that may inhibit Pancreatic Adenocarcinoma signaling pathways.	<b>107</b>
<b>Table 3.4</b> Knockout of DDI2 showed downregulation of EGFR and Interleukin 6 (IL-6) that may inhibit SPINK general cancer signaling pathways.	<b>108</b>
<b>Table 3.5</b> Knockout of DDI2 showed genes' downregulation in xenobiotic metabolism and transporter.	<b>109</b>
<b>Table 3.6</b> Knockout of DDI2 does not alter the transcriptional expression of proteasome subunits except the immunoproteasome PSMB8 and PSMB9.	<b>112</b>
<b>Table 3.7</b> List of primary and secondary antibodies.	<b>123</b>
<b>Table 4.1</b> Covalent docking analysis for three different modified diazo-reactive functional groups	<b>136</b>
<b>Table 4.2</b> List of chemicals used for the peptide synthesis.	<b>153</b>
<b>Table 4.3</b> Analytical LC separation mobile phase.	<b>157</b>
<b>Table 4.4</b> Semi-preparative LC separation mobile phase.	<b>158</b>
<b>Table 4.5</b> List of primary and secondary antibodies.	<b>165</b>
<b>Table SI 2.1</b> List of qPCR primers used in Figure 2.4 and Figure 2.6	<b>89</b>
<b>Table SI 3.1</b> Normalized RNA count values of highly variable differentially expressed genes (DEGS) in Figure 3.3	<b>125</b>
<b>Table SI 4.1</b> The prime MMGBSA dG Bind (NS) energy components of designed probes	<b>166</b>

## List of Figures

<b>Figure 1.1</b> The Ubiquitin-proteasome system (UPS) degradation of protein substrates scheme.	<b>21</b>
<b>Figure 1.2</b> The 26S proteasome and proteasome catalytic active site.	<b>22</b>
<b>Figure 1.3</b> Increased proteasome load in cancer cells disrupts proteostasis balance.	<b>25</b>
<b>Figure 1.4</b> Structure of three FDA-Approved proteasome inhibitors.	<b>27</b>
<b>Figure 1.5</b> Mechanism of Bortezomib covalent inhibition with the N-term threonine of the proteasome $\beta 5$ catalytic active site.	<b>27</b>
<b>Figure 1.6</b> Mechanism of Carfilzomib irreversible inhibition with the N-term threonine of the proteasome $\beta 5$ catalytic active site and forming a stable morpholino adduct.	<b>29</b>
<b>Figure 1.7</b> The regions of transcription factors involved in regulating proteasome genes belong to the Cap and Collar families.	<b>32</b>
<b>Figure 1.8</b> Regulation of Nrf1 transcription factor in the ER, cytosol, and nucleus.	<b>38</b>
<b>Figure 1.9</b> Domain structure of hDDI2 and yDdi1. Human DDI2 differs from yeast Ddi1 in the c-term domain.	<b>42</b>
<b>Figure 2.1</b> Knockout of DDI2 does not impair the proteasome activity recovery.	<b>54</b>
<b>Figure 2.2</b> Knockout of DDI2 does not affect the recovery activity even though DDI2 knockout cells cannot process Nrf1.	<b>55</b>
<b>Figure 2.3</b> Knockdown of DDI2 does not affect the recovery of the proteasome in inhibitor-treated cells.	<b>56</b>
<b>Figure 2.4</b> The recovery of proteasome activity preceded the induction of proteasome mRNA.	<b>58</b>
<b>Figure 2.5</b> Treatment with transcription inhibitors Flavopiridol (FlvP) or Actinomycin D (ActD) activates caspase 3/7 activity and induces cell apoptosis HAP1 wt cells.	<b>60</b>
<b>Figure 2.6</b> The recovery of proteasome activity required the synthesis of new proteasomes.	<b>61</b>
<b>Figure 2.7</b> The translation efficiency of proteasome subunits mRNA.	<b>63</b>

<b>Figure 2.8</b> Increased 26S proteasome activity predominantly contributed to HAP1 cell's proteasome activity recovery upon proteasome impairment.	<b>67</b>
<b>Figure 3.1</b> Global RNA expression of from total RNA-seq of DDI2 CRISPR KO cells (clones c006, c023, and c010) vs. HAP1 wt.	<b>98</b>
<b>Figure 3.2</b> QIAGEN - Ingenuity Pathway Analysis (IPA) from differentially expressed genes (DEGs) of DDI2 CRISPR KO (clone c006, c023, and c010) vs. HAP1 wt.	<b>102</b>
<b>Figure 3.3</b> Pancreatic cancer patients with low DDI2 RNA levels have a significant probability of survival compared to those with high DDI2 RNA levels (p-value 0.0010).	<b>106</b>
<b>Figure 3.4</b> Normalized RNA count of proteasome subunits, proteasome chaperone, and proteasome activators from RNA-seq data of DDI2 KO c010 (gRNA2), DDI2 KO c023, c006 (gRNA1) vs. HAP1 wt.	<b>113</b>
<b>Figure 3.5</b> Knockout of DDI2 lowers the basal level of immunoproteasome $\beta 5i$ activity.	<b>114</b>
<b>Figure 4.1</b> The crystal Structure of DDI2 retroviral-like protease domain (PDB: 4RGH).	<b>131</b>
<b>Figure 4.2</b> List of diazo warhead functional groups reported for aspartic proteases.	<b>133</b>
<b>Figure 4.3</b> Ligand-to-protein receptor binding interactions of Probe 7 in 4RGH receptor.	<b>137</b>
<b>Figure 4.4</b> The LC-MS Spectra and HRMS-ESI of N- $\alpha$ -Fmoc-4(methyl)phenyl glycine compound (7)	<b>143</b>
<b>Figure 4.5</b> The $^1\text{H}$ NMR and $^{13}\text{C}$ NMR chemical shift ( $\delta$ , ppm) of N- $\alpha$ -Fmoc-4(methyl)phenyl glycine in DMSO- $d_6$ .	<b>144</b>
<b>Figure 4.6</b> The $^1\text{H}$ NMR chemical shift ( $\delta$ , ppm) of azido peptide 2 in DMSO- $d_6$ .	<b>147</b>
<b>Figure 4.7</b> LC-MS and HRMS-ESI of purified azido peptide 2.	<b>148</b>
<b>Figure 4.8</b> Bioconjugation of azido peptide 2 in HAP1 cells.	<b>151</b>
<b>Figure 5.1</b> Conversion of Azide into Diazo compounds via Azide deimidogenation mediated by several phopshinoesters	<b>179</b>

<b>Figure SI 2.1</b> Complete data of the translation efficiency of proteasome subunits mRNA.	<b>91</b>
<b>Figure SI 2.2</b> The recovery of proteasome activity is initiated by an increase in 26S/30S proteasome activity.	<b>94</b>
<b>Figure SI 3.1</b> Data processing from total RNA-seq to Differentially Expressed Genes (DEGs) of data set DDI2 CRISPR KO cells clone (c006, c023, c010) vs. HAP1 wt cells.	<b>126</b>
<b>Figure SI 3.2</b> QIAGEN - Ingenuity Pathway Analysis (IPA) from differentially expressed genes (DEGs) of DDI2 CRISPR KO (clone c006, c023, and c010) vs. HAP1 wt.	<b>127</b>
<b>Figure SI 3.3</b> SPINK general activation pathway in cancer.	<b>128</b>
<b>Figure SI 3.4</b> The RNA Integrity Number (RIN) values and RNA quality profiles of RNA samples for total-RNA sequencing.	<b>129</b>
<b>Figure SI 4.1</b> HPLC-MS Spectra of undesirable products	<b>167</b>
<b>Figure SI 4.2</b> HRMS-ESI of 4(methyl) phenyl glycine azide calculated for $C_9H_9N_3O_2$ m/z $[M-H]^-$ 190.0614.	<b>167</b>
<b>Figure SI 4.3</b> Predicted $^1H$ NMR and $^{13}C$ NMR chemical shift (ppm) of N- $\alpha$ -Fmoc-4(methyl)phenyl glycine in $CDCl_3$ .	<b>168</b>
<b>Figure SI 4.4</b> HRMS-ESI of compounds 6, 5, and 4 sampled from crude SPPS reaction.	<b>169</b>
<b>Figure SI 4.5</b> LC-MS chromatography profile of azido peptide 2 before purification showed chromatogram peak at RT between $\pm 6.2$ min for exact mass 692.3358.	<b>170</b>
<b>Figure SI 4.6</b> Chromatography profile of azido peptide 2 and its purification using semi-preparative LC.	<b>171</b>
<b>Figure SI 4.7</b> LC-MS Spectra profile of chromatogram peak 1 (P1) and peak 2 (P2) in Figure SI 4.6.	<b>172</b>
<b>Figure SI 4.8</b> Overlays of LC-MS chromatogram profile of purified azido peptide 2 and blank	<b>173</b>

## List of Schemes

- Scheme 4.1** The Disconnection approach (A) and synthetic route (B) of Probe No.7. **139**
- Scheme 4.2** Possible explanation of the formation of undesirable  $m/z$   $[M+H]^+$  598.28 or  $m/z$   $[M+Na]^+$  620.26 may be due to the partial removal of Fmoc-N- $\alpha$ -lysine following Alloc deprotection. **141**
- Scheme 4.3** Synthetic route from 9 to 8 and 7 **142**



## Chapter 1 Introduction

Protein homeostasis (Proteostasis) is critical to all eukaryotic cells' survival, stability, and normal function<sup>1-3</sup>. The cell maintains a balance of protein synthesis<sup>1</sup>, chaperone-assisted protein folding<sup>1-2</sup> and subunits assembly<sup>4</sup>, and protein degradation<sup>1</sup>. Imbalances of these processes can disrupt protein homeostasis and manifest in many disease pathogenesis (loss-or gain-of-function diseases)<sup>1</sup>, including cancer.

Cancer is characterized by genetic changes (aneuploidy and increased genetic heterogeneity), further promoting uncontrolled cell growth and tumorigenesis<sup>5-8</sup>. As a result, cancer cells produce more unassembled protein subunits<sup>8</sup> and become susceptible to rapid accumulation of incorrect protein conformations or abnormal proteins<sup>5</sup>. Consequently, to avoid the accumulation of misfolded proteins, they have become highly dependent on protein folding and degradation mechanisms<sup>5</sup>. The ubiquitin-proteasome system (UPS) is the primary intracellular protein degradation<sup>9</sup> and turnover machinery in eukaryotes<sup>10</sup>. Hijacking the UPS system using proteasome inhibitors negatively affects the survival of cancer cells more than normal cells<sup>5</sup>.

Nevertheless, a growing body of evidence suggests cases of intrinsic resistance and unresponsiveness to proteasome inhibitors due to the recovery of proteasome activity<sup>5, 11-13</sup>. The latter was reported as the result of the induction of proteasome genes through ER-resident Nrf1 activation and processing<sup>14</sup>. Nrf1 is processed by protease DNA Damage Inducible 1 Homolog II (DDI2)<sup>15</sup>.

DDI2 is a novel aspartic protease whose biological function has not been widely studied and characterized<sup>16</sup>. What is interesting about this protein is that DDI2

consists of a retroviral protease domain with a distinctive motif similar to a well-characterized HIV protease<sup>16</sup>. So far, the only known substrate of DDI2 is the Nrf1 transcription factor involved in proteasome degradation<sup>17</sup>. Moreover, DDI2 is also part of a ubiquitin-binding protein that only recognizes polyubiquitylated substrates similar to the proteasome. However, Unlike proteasomes that cleave the ubiquitylated substrates into smaller peptides, DDI2 performs a singly precise peptide bond cleavage on Nrf1<sup>14</sup>.

In this chapter, first, we will outline the role of the ubiquitin-proteasome (UPS) system in cancer. Then, we will review the transcription regulation of proteasome, focusing on the ER-resident Nrf1 transcription factor. Next, we will discuss the importance of DDI2 in the UPS systems.

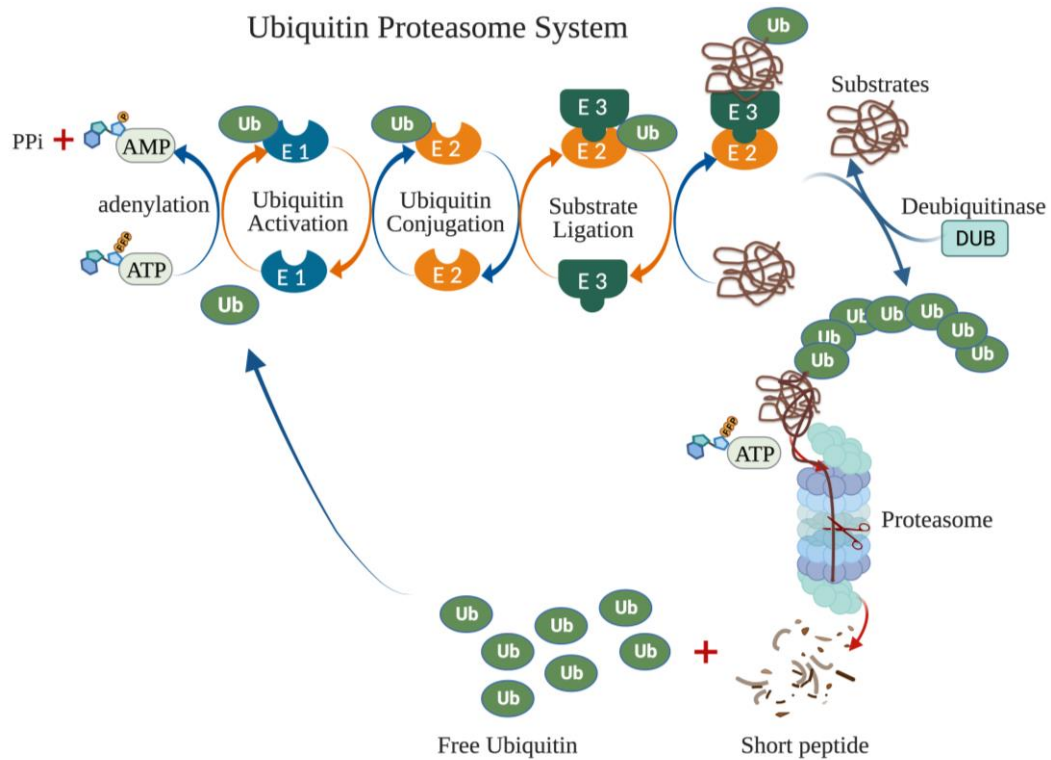
## 1.1 Overview of the ubiquitin-proteasome system

Eukaryotes utilize the Ubiquitin proteasome system (UPS) to maintain intracellular proteins' turnover, balance regulatory proteins and keep them functioning correctly<sup>10</sup>. The UPS plays an essential role in cell cycle progression<sup>10, 18-19</sup>, the unfolded protein response (UPR) and the integrated stress response (ISR)<sup>20</sup>, transcriptional regulation and signal transduction<sup>10, 18-20</sup>, DNA repair and genome integrity<sup>10, 18, 21-23</sup>, programmed cell death<sup>10, 18-19</sup>, inflammation<sup>10, 20</sup>, immune responses<sup>10, 19-20</sup>, changes in brain neural networks<sup>10</sup> and aging<sup>20</sup>, metabolism<sup>19-20</sup>, and pathogenesis of many human diseases.

UPS is an ATP-dependent proteolysis process that selectively degrades substrate proteins in the cytosol and nucleus into peptides fragment (3–22 amino acids)<sup>24</sup>. The peptide fragments are further broken down into amino acids by endopeptidases, and 80% of these amino acids were re-used to synthesize nascent peptides<sup>24</sup>. The substrate proteins degraded by the UPS are soluble intracellular proteins<sup>10</sup>. At the same time, the lysosome degrades the remaining insoluble and long-lived intracellular proteins via autophagy<sup>10</sup>. These lysosomes are also responsible for degrading the extracellular proteins delivered through the endocytic pathway<sup>9-10, 18, 25-26</sup>.

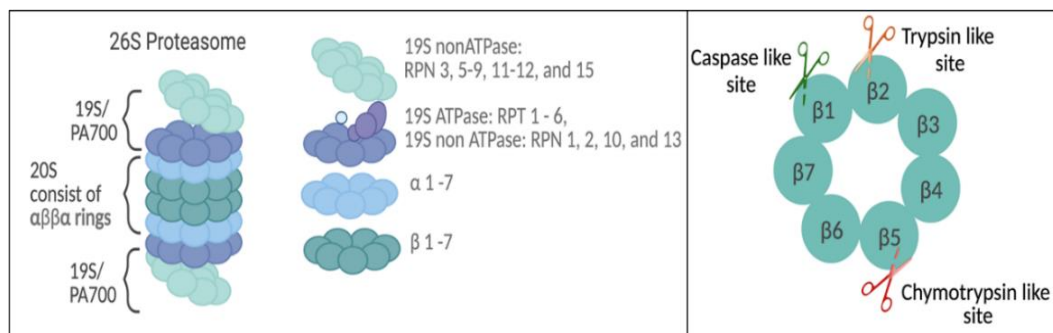
The UPS-dependent proteolysis requires the conjugation of a polyubiquitin chain for substrate recognition<sup>25</sup>. This process is mediated by a cascade involving three enzymatic processes (**Figure 1.1**)<sup>9, 18, 25</sup>. The first step is ubiquitin activation by the ubiquitin-activating enzyme (E1), which requires ATP to adenylate the Glycine residue (Gly76) at the C-terminus of ubiquitin, forming an AMP-Ubiquitin adduct<sup>9, 18</sup>. The ubiquitin activation is essential to initiate a high-energy thioester bond

between an active-site cysteine (Cys) of ubiquitin-activating enzyme (UBA1) E1 and the activated ubiquitin<sup>9, 18</sup>. The activated ubiquitin is transferred to the Ubiquitin-conjugating enzyme (E2), forms an E2-ubiquitin intermediate, and is subsequently conjugated to the substrate-bound E3 ligase<sup>9, 18</sup>. The E3 ubiquitin ligase identifies the substrates and bridges the E2-ubiquitin intermediary with the substrate through a covalent isopeptide bond formation between the C-terminal glycine residue of activated ubiquitin to the lysine residue of the substrates<sup>9, 18</sup>. Following the attachment of the first ubiquitin (monoubiquitylation), additional ubiquitin molecules are attached to one of the seven lysine residues of ubiquitin. Thus formed polyubiquitin chains such as K11, K11/K48, and K48 are recognized by 26S proteasome<sup>9, 18</sup>. The polyubiquitylated substrates are further broken down into short peptides by proteasomes<sup>9, 27</sup>. Meanwhile, the disassembled polyubiquitin chains are recycled to tag other protein substrates<sup>19, 25</sup>.



**Figure 1.1** The Ubiquitin-proteasome system (UPS) degradation of protein substrates scheme. The UPS-dependent proteolysis requires the conjugation of a polyubiquitin chain for substrate recognition and is mediated by a cascade involving three enzymatic processes. Modified from Goldberg, AL. (2012)<sup>28</sup>; Ciechanover, A. and Kwon, Y.T. (2015)<sup>9</sup>. Created with BioRender.com.

The Proteasome is the primary intracellular protein degradation that selectively degrades the polyubiquitylated substrates. The 26S proteasome is a large 2.5 MDa barrel-shaped ATP-dependent protease complex consisting of a catalytic 20S core particle (20S CP) sandwiched between two 19S regulatory particles (19S RP)<sup>9-10, 18-19, 29</sup>. The 20S CP itself consists of four stacked rings arranged to form a cylindrical tube of  $\alpha$ - $\beta$ - $\beta$ - $\alpha$  rings (Figure 1.2, left panel)<sup>10, 19, 29-30</sup>. Each of these rings comprises seven subunits ( $\alpha_{1-7}$ ,  $\beta_{1-7}$ )<sup>29</sup>, where the N-terminus of the  $\alpha$ -subunit serves as an open-close gate to  $\beta$ -ring<sup>10, 30</sup>. The  $\alpha$ -subunit is in a closed gate ring and will re-open its gate following the substrate recognition<sup>19</sup>.



**Figure 1.2** *The 26S proteasome and proteasome catalytic active site. (left panel) The structure of the 26S Proteasome consisted of two rings of 19S regulatory particles and  $\alpha$ - $\beta$ - $\alpha$  rings of 20S core particles. (Right panel) The  $\beta$  rings of the core particle 20S Proteasome consisted of three main active sites: Caspase-like sites, Trypsin-like sites, and Chymotrypsin-like sites—Figure generated through BioRender™.*

## 1.2 Proteasome 20S CP

The proteasome is a multi-catalytic enzyme in which the active site bearing several N-terminal threonine residues is located on the  $\beta$ -ring of the 20S CP human proteasome ( $\sim 750$  kDa)<sup>10, 19</sup>. This  $\beta$ -ring consists of seven subunits ( $\beta_1 - \beta_7$ ) with six proteolytically functional  $\beta$  subunits categorized into three types of proteolytic activities as follows: two caspase-like sites, two trypsin-like sites, and two chymotrypsin-like sites (**Figure 1.2, right panel**)<sup>10, 31</sup>. Each of these catalytic subunits has preferential substrate specificity, such as the caspase-like site or peptidyl glutamyl-peptide hydrolyzing ( $\beta_1$  subunits), cleaving the amide bonds adjacent to the acidic amino acid residue. In contrast, the trypsin-like site ( $\beta_2$  subunits) hydrolyzes the amide bond after basic amino acid residues. The chymotrypsin-like site ( $\beta_5$  subunits), known as the rate-limiting catalytic active site for proteasomal degradation, prefers to cleave after aromatic hydrophobic amino acid residue<sup>10, 19, 31</sup>. As a multi-catalytic enzyme, the catalytic subunits interact synchronously with their neighboring  $\beta$  subunits before exerting their proteolytic activity<sup>10</sup>.

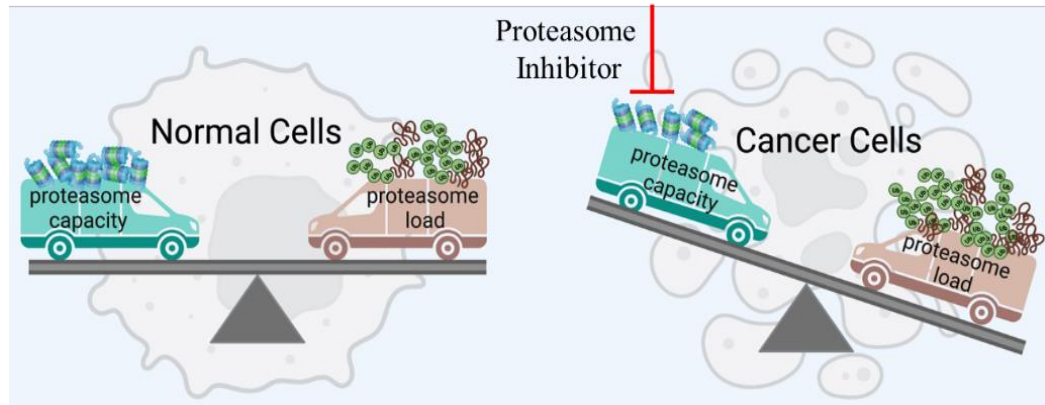
### 1.3 The 19S Regulatory Particles

The 19S Regulatory Particles (RP), also known as PA700 (700 kDa), is composed of 19 subunits which comprise two large subcomplexes, the lid and the base (**Figure 1.2, left panel**)<sup>19, 29</sup>. It consists of ATPase (Rpt 1 - 6) and non-ATPase (Rpn 1 - 3, 5 - 13, and 15) with the primary function as deubiquitinase<sup>19, 29, 31</sup>. The 19S RP has several functions in the UPS. First, the role of the 19S RP (Rpn 1, 10, and 13) is to recognize the ubiquitylated substrates<sup>10, 29, 31</sup>. The 19S RP will further unfold the identified substrate to the nascent peptide, allowing its translocation to 20S CP<sup>10, 19, 29</sup>. Second, the 19S RP (PA700) acts as a proteasome activator<sup>19</sup> through binding with latent 20S CP<sup>10, 30</sup>. This will trigger the gate opening of the  $\alpha$ -ring of the 20S CP chamber and allosterically activates its peptidase activity<sup>19, 30</sup>. Three 19S RP subunits (Rpt2/PSMC1, Rpt3/PSMC4, and Rpt5/PSMC3) function in gate opening upon ATP binding. These three 19S RP Subunits are characterized by a C-terminal hydrophobic-tyrosine-X (HbYX) motif. Active proteasome mainly exists as a double-capped 30S (19S-20S-19S) or a single-capped 26S (19S-20S); however, 20S CP proteasome can bind with other proteasome activators and form alternative proteasome complex<sup>19, 32</sup>. Besides 19S RP, the endogenous 11S regulator hREG PA28 $\alpha\beta$  (PA28 $\alpha$ /PSME1, PA28 $\beta$ /PSME2), PA28 $\gamma$  (PSME3) and yBln10 (PA200, PSME4) can also activate the 20S CP proteasome<sup>19, 32</sup>. Thus, alternative proteasome complexes can exist in various forms, such as PA200–20S–PA200, PA200–20S, 11S–20S–11S; alternatively, they can also exist as hybrid complexes such as PA28 $\alpha\beta$ –20S–19S, PA28 $\gamma$ –CP–PA28 $\gamma$ <sup>19</sup>. Although, to be noted, due to its conformational changes, the latent 20S CP alone still has a limited peptidase activity<sup>10</sup>.

#### **1.4 The Rationale for Targeting Proteasomes for the Treatment of Cancer**

Cancer is a significant public health problem worldwide<sup>33</sup>. Despite the advancement in therapeutic modalities for cancer, the number of cancer-related deaths is still not significantly reduced. Cancer remains the second primary source of death in the United States<sup>33</sup>, with the predicted number of new cases by 2022 and cancer deaths rising to 1,918,030 and 609,360, respectively<sup>33</sup>. Cancer cells are characterized by growth dysregulation or uncontrolled cell division, which leads to malignant proliferation and excessive protein synthesis<sup>5</sup>. However, this increased rate of protein synthesis might not be followed by a similar protein folding or assembly rate<sup>5</sup>. As a result, cancer cells produce a higher rate of misfolded proteins. Due to these excessive productions of the misfolded proteins that must be degraded or refolded, cancer cells highly rely on the protein quality control mechanisms (the UPS and autophagy, molecular chaperone) to achieve proteostasis balance<sup>5</sup>. Therefore, any condition impeding the protein quality control and its pathway can generate more proteotoxic stress on cancer cells than on normal cells<sup>5</sup>. For instance, as the number of misfolded proteins in cancer cells increases while the capacity to degrade them remains the same, inhibiting the ubiquitin-proteasome system reasonably should also cause more proteotoxic stress to cancer cells than normal cells (**Figure 1.3**)<sup>5</sup>.





**Figure 1.3 Increased proteasome load in cancer cells disrupts proteostasis balance.** Cancer cells experienced higher proteotoxic stress due to a higher load of misfolded proteins. Inhibiting proteasome activity leads to selective apoptosis of cancer cells than normal cells. Modified from Deshaies, R.J. (2014)<sup>5</sup>. This Figure is created with BioRender.com.

Through this rationale, several proteasome inhibitors have been developed to target proteasome for cancer treatment; some are FDA-approved as the first-line therapy for Multiple myeloma (MM) and Mantle cell lymphoma (MCL).

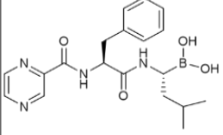
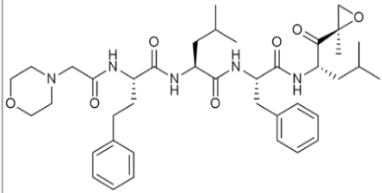
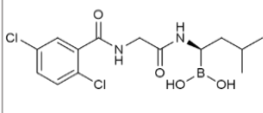
Multiple myeloma (MM) is a cancer of plasma cells, a terminally-differentiated antibody-producing B-cells, accounting for 10% of all blood types of cancer and 1% of all cancer in the US<sup>34</sup>. This tumor proliferates at numerous bone marrow (BM) sites, including extracellular matrix proteins and antigen-presenting cells<sup>35-36</sup>. The attachment of MM cells to the antigen-presenting cells and the extracellular matrix protein (ECM) will induce the expression of growth stimulators for the survival of tumorigenic cells, such as insulin-like growth factor (IGFs)<sup>35</sup> and vascular endothelial growth factor (VEGF)<sup>35</sup>. IGFs and VEGF are modulated by the Nuclear Factor- $\kappa$ B (NF $\kappa$ B)<sup>37</sup>. NF $\kappa$ B is activated through two major pathways. One of the canonical pathways of NF $\kappa$ B activation is mediated by proteasome degradation of I $\kappa$ B $\alpha$ <sup>28</sup>. Furthermore, the attachment of MM cells to ECM or antigen-presenting cells also increased the secretion of abnormal immunoglobulin (IgG and IgA) proteins in their blood and urine<sup>36</sup>. Hence, it is not surprising that

MM cells accumulate high molecular weight polyubiquitylated substrate at their basal level and in the absence of treatment with proteasome inhibitors<sup>38-39</sup>, which significantly contributed to the higher load of proteasome<sup>12, 28, 37</sup>.

Mantle cell lymphoma (MCL) is a mature B-cell lymphocyte malignancy in the mantle of a lymph node follicle<sup>40</sup>. MCL represents a subtype of non-Hodgkin lymphomas (NHLs), accounting for 5 - 6% of all NHLs cases<sup>41</sup>. When the B-lymphocyte becomes malignant, it will accumulate and causes lymph node enlargement. MCL is characterized by the constitutive expression of cyclin D1 (CCND1), which promotes high mitotic rates<sup>40</sup> and contributes to a higher load of the Proteasome in MCL. Cyclin D1 is dependent on 26S proteasome for its degradation. Thus, using proteasome inhibitors for MCL treatment does cause the deregulation of cyclin D1.

## **1.5 Proteasome Inhibitors**

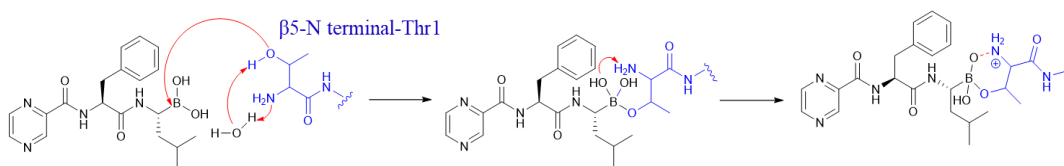
Three FDA-approved proteasome inhibitors currently exist: Bortezomib, Carfilzomib, and Ixazomib (**Figure 1.4**). All these proteasome inhibitors are developed to target the primary catalytic active site of Proteasome  $\beta_5$  subunits.

 <p>Bortezomib</p>	<p>CAS # 179324-69-7</p> <p>Molecular Formula C<sub>19</sub>H<sub>25</sub>BN<sub>4</sub>O<sub>4</sub></p> <p>Molar Mass 384.237 g/mol</p> <p>IUPAC Name [(1R)-3-methyl-1-[[[(2S)-3-phenyl-2-(pyrazine-2-carbonylamino)propanoyl]amino]butyl]boronic acid</p>
 <p>Carfilzomib</p>	<p>CAS # 868540-17-4</p> <p>Molecular Formula C<sub>40</sub>H<sub>57</sub>N<sub>5</sub>O<sub>7</sub></p> <p>Molar Mass 719.9114 g/mol</p> <p>IUPAC Name (2S)-4-methyl-N-[[[(2S)-1-[[[(2S)-4-methyl-1-[(2R)-2-methyloxiran-2-yl]-1-oxopentan-2-yl]amino]-1-oxo-3-phenylpropan-2-yl]-2-[[[(2S)-2-[(2-morpholin-4-ylacetyl)amino]-4-phenylbutanoyl]amino]pentanamide</p>
 <p>Ixazomib</p>	<p>CAS # 1072833-77-2</p> <p>Molecular Formula C<sub>14</sub>H<sub>19</sub>BCl<sub>2</sub>N<sub>2</sub>O<sub>4</sub></p> <p>Molar Mass 361.03 g/mol</p> <p>IUPAC Name [(1R)-1-[[2-[(2,5-dichlorobenzoyl)amino]acetyl]amino]-3-methylbutyl]boronic acid</p>

**Figure 1.4 Structure of three FDA-Approved proteasome inhibitors.** This Figure is generated through Chemdraw™.

### 1.5.1 Bortezomib

Bortezomib (PS-341, Velcade®, Btz, **Figure 1.4, upper panel**) was the first authorized by the U.S. Food and Drug Administration (FDA) to treat relapsed and refractory cases MM<sup>37, 42-43</sup>. Structurally, Btz is a dipeptide boronic acid, a potent small molecule reversible proteasome inhibitor<sup>37, 44</sup>. The boronic acid warhead forms a covalent adduct with the N-terminal threonine nucleophilic hydroxyl group of the 20S proteasome<sup>44</sup>. This binding is further stabilized by the Hydrogen bond between the alpha-amino group and one of the boronates hydroxyl, forming a 6-member ring-like transition state (**Figure 1.5**)<sup>37, 44</sup>.

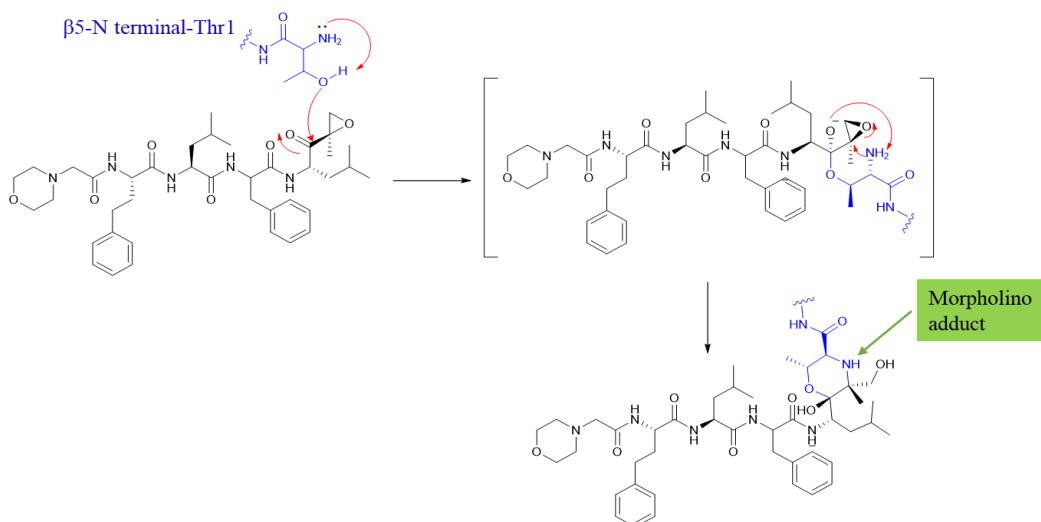


**Figure 1.5 Mechanism of Bortezomib covalent inhibition with the N-term threonine of the proteasome  $\beta 5$  catalytic active site.** Adapted and slightly modified from Kisselev, A.F. et al. (2012). Figure generated through Chemdraw™.

Btz selectively blocks the  $\beta 5$  sites protease at physiologically relevant concentrations, leaving the other proteolytic sites ( $\beta 2$  and  $\beta 1$  sites) to remain active<sup>28, 37, 44-45</sup>. However, at higher concentrations, Btz also co-inhibits the  $\beta 1$ <sup>28, 37, 44-45</sup> but not the  $\beta 2$  site<sup>44</sup>. Clinically, Btz exhibits a biphasic pharmacokinetic profile following a one-hour bolus IV administration. It achieves rapid plasma distribution with a  $t_{1/2}$  of < 10 minutes and maximum chymotrypsin-like ( $\beta 5$ ) inhibition within an hour<sup>46</sup>. Then, the recovery starts at 4h, and activity returns to its original baseline in 72 - 96 hours<sup>46-47</sup>. However, it requires a longer complete elimination time due to the slower rate of Btz dissociation from its proteasome binding site<sup>37</sup>. Hence, Btz cannot be given frequently and must follow a relevant dosing schedule to avoid toxicity or build-up of Btz.

### 1.5.2 Carfilzomib

Carfilzomib (PR-171, Kyprolis®, Cfz, **Figure 1.4, middle panel**) is a selective irreversible proteasome inhibitor developed based on the scaffold of the natural product Epoxomicin<sup>37, 48</sup> and structurally different from Btz<sup>37</sup>. Both Epoxomicin and Cfz bear a reactive electrophilic epoxyketone moiety<sup>37, 48</sup>. Mechanistically, it forms double covalent bonds with the N-terminal threonine of the 20S proteasome<sup>48-49</sup>. The first covalent bond is between the hydroxyl group of the N-terminal threonine with the carbonyl (Thr1O $\gamma$ )<sup>48-49</sup>. The alpha amine of threonine (Thr1N) opens up the epoxide from the most substitute carbon and forms a second covalent bond<sup>48-49</sup>. Subsequently, these two covalent bonds form a stable morpholino ring (**Figure 1.6**)<sup>37, 48</sup> and contribute to irreversible inhibition to block the activity of  $\beta 5$  sites Proteasome<sup>37, 50</sup>. In addition, it shows little off-target activity against other proteases and little toxicity compared to Btz<sup>37, 50-52</sup>.



**Figure 1.6 Mechanism of Carfilzomib irreversible inhibition with the N-term threonine of the proteasome  $\beta 5$  catalytic active site and forming a stable morpholino adduct.** Adapted and slightly modified from Kim, K.B., and Crews, C.M. (2013). Figure generated through Chemdraw™.

Cfz single-agent is FDA-approved for treating relapsed and refractory cases of MM<sup>37, 50-52</sup>. Clinically, Cfz is given as a bolus IV (30 min) where its half-life ( $t_{1/2}$  0.5 - 1.2h) and maximum  $\beta 5$  inhibition (83.1% - 97.7%) depends on the administered dose<sup>11, 53</sup>. Cfz showed complete elimination from plasma within 3 hours of its administration, resulting in full recovery of the proteasome activity a week after. Considering that Cfz is an irreversible inhibitor, it is unlikely that the proteasome activity recovery is due to the dissociation of the inhibitors from the active catalytic site, and other mechanisms involved in making new proteasomes must have taken place.

### 1.5.3 Ixazomib

Ixazomib (MLN-2238, Ninlaro®, **Figure 1.4, bottom panel**) is the first FDA-approved orally bioavailable proteasome inhibitor approved for treating relapsed and refractory MM<sup>54</sup>. Ninlaro is marketed as a prodrug capsule of Ixazomib citrate

(MLN-9708). Ixazomib citrate will undergo rapid hydrolysis in blood to deliver its biologically active metabolite form, Ixazomib<sup>55-56</sup>, and render its inhibitory activity.

Ixazomib is the second-generation boronate-containing peptide proteasome inhibitor, an N-capped dipeptide boronic acid<sup>55-57</sup>. Ixazomib, on the one hand, is chemically similar to Bortezomib in which the boronic acid warhead forms a covalent adduct with the N-terminal threonine hydroxyl group of the 20S proteasome (see **Figure 1.5** for the mechanism of inhibition)<sup>44</sup>. Therefore, it also selectively blocks the  $\beta$ 5-proteasome at pharmacologically relevant concentrations, yet, it will bind to all sites at higher concentrations<sup>55-57</sup>. On the other hand, Ixazomib shows different physicochemical properties, such as a faster dissociation rate of its complex with proteasome than Bortezomib (six-fold faster)<sup>55, 57</sup>, greater tissue penetration and target engagement, and a higher percentage of proteasome recovery in the cells<sup>57</sup>. Furthermore, a preclinical study of ixazomib shows inhibition of activity against Bortezomib-resistant cells<sup>54, 56</sup>.

## **1.6 Regulation of proteasome expression**

Given the importance of the UPS in degrading intracellular protein and protein turnover, understanding how proteasomes are regulated becomes necessary<sup>31</sup>. Previous studies reported that the regulation of proteasomes activity occurs at different levels<sup>31</sup>. The 26S proteasome is controlled at an early stage of biogenesis: 1) from transcriptional to synthesizing 33 subunits, 2) subunits assembly into mature proteasome, and the subsequent post-translational modification<sup>29, 31</sup>. Here, we focused on the regulation of proteasome activity by biosynthesis.

### 1.6.1 Regulation of proteasome genes at the transcriptional level

Upregulation of mRNA of proteasome genes and increasing synthesis of nascent polypeptides may lead to a lesser response to therapeutic proteasome inhibitors<sup>14</sup>. Numerous studies reported that several transcription factors such as Rpn4 (SON1, UFD5) in yeast<sup>29, 31, 58-60</sup>, SKN-1<sup>29, 60-61</sup> in *C. elegans* and its mammalian orthologous Nrf1/Nrf2/Nrf3<sup>15, 29, 60</sup>, NF-Y<sup>29, 62</sup>, and FOXO4<sup>63-64</sup> mediate the transcription of proteasome genes<sup>29, 31</sup>.

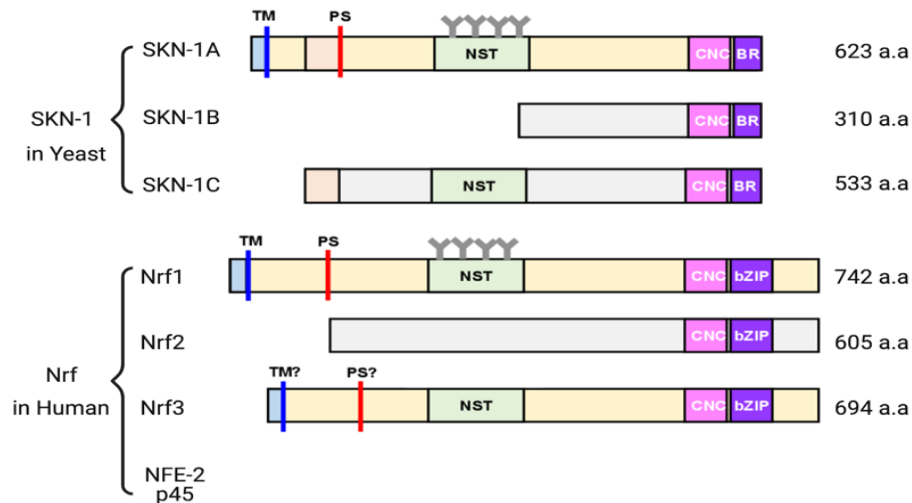
### 1.6.2 Regulation of proteasome genes under basal conditions

In *Saccharomyces cerevisiae* (yeast), the expression of constitutive proteasomes genes is coordinated by the Rpn4 transcription factor (TF) that binds to the promoter region of the proteasome-associated control element (PACE) in every proteasome subunit, following its translocation to the nucleus<sup>29, 31, 59-60</sup>. Rpn4 is a substrate for ubiquitylation and is quickly degraded by the proteasome (Rpn4  $t_{1/2}$  ~ 2 min) under normal conditions<sup>29, 31, 58, 60</sup>. Knockout of Rpn4 or PACE in one of the proteasome genes confirmed the importance of Rpn4 and PACE in proteasome regulation<sup>29, 31, 58, 60</sup>. When the cells lack Rpn4 or PACE, the proteasome activity is attenuated<sup>58, 60</sup>, the cell cycle is impaired<sup>58</sup>, and the cells become sensitized to stress factors<sup>29, 31, 60</sup>. RPN4 itself is regulated through many stress-induced TFs, such as genes related to oxidative stress and heat-shock<sup>31, 60</sup>.

Meanwhile, Skinhead 1 (SKN-1) in *Caenorhabditis elegans* (a nematode, *C. elegans*) plays a similar function to yeast Rpn4<sup>31</sup>. SKN-1 exists in 3 isoforms (SKN-1A, SKN-1B, and SKN-1C) that belong to the Cap 'n' Collar (CnC) family with basic leucine regions (BR) at their C-term sequence (**Figure 1.7**)<sup>29, 31, 60-61</sup>. In normal

conditions, SKN-1 is a substrate for ubiquitylation and proteasome degradation<sup>31</sup>,

61



**Figure 1.7** The regions of transcription factors involved in regulating proteasome genes belong to the Cap and Collar families. SKN-1 (SKN-1A, SKN-1B, SKN-1C) in yeast and Nrf (Nrf1, Nrf2, Nrf3) in human. CnC and bZIP or BR are in the C-terminal domain (CTD). At the same time, the N-terminal domain (NTD) consists of the Transmembrane domain (TM), Processing Domain (PS), and NST (Asn-X-Ser/Thr) glycosylation domain. In addition, Nrf TFs have an acidic domain responsible for transcription factor activation. Adapted and slightly modified from Koizumi, S. et. al. (2018)<sup>29</sup>; Hamazaki, J. and Murata, S. (2020)<sup>60</sup> with copyright permission.

Like *C. elegans* and *S. cerevisiae*, mammals also maintain their regulatory mechanisms to compensate for the hampered proteasome function through multiple TFs regulation, where each of these TFs controls a different subset of proteasome genes depending on the cell type and its cellular conditions. Some of these TFs are the nuclear transcription factor Y (NF-YA, NF-YB, and NF-YC subunits); the nuclear factor erythroid-derived 2-related factor (Nrf1, Nrf2, and Nrf3); and the Forkhead box protein O4 (FOXO4)<sup>29, 31, 62</sup>.

The NF-Y TF partially regulates the basal level expression of CCAAT box-containing proteasome genes (namely: PSMA2, PSMA3, PSMA5, PSMB3, and PSMC2) in human breast cancer (MDA-MB-231) and normal breast epithelial cells



(MCF-1 10A), yet not for the proteasome genes that do not contain any CCAAT box<sup>29, 62</sup>.

FOXO family (consisting of FOXO1, FOXO3, FOXO4, and FOXO6) was also reported to orchestrate gene expression needed for clearance of abnormal proteins by autophagy or the UPS in mammals<sup>63</sup>. FOXO4 (but not FOXO1 or FOXO3) is necessary for upregulating the mRNA level of PSMD11, particularly in human embryonic stem cells (hESCs)<sup>29, 63</sup>, but not in BJ Fibroblast or HEK293T dividing cells<sup>63</sup>. The role of FOXO4/ Daf-16 is also conserved in *C. elegans*<sup>64</sup>. Silencing of RPN6 in *C. elegans* reversibly decreases the proteasome activity and induces the mRNA level of other proteasome subunits as a negative feedback loop of reduced proteasome activity<sup>64</sup>. At the same time, the deletion of Daf-16 only partially decreases the mRNA level of RPN6 and not to the level of control, further implying that Daf-16 is not the only regulator of RPN6<sup>64</sup>.

The mammalian Nuclear factor erythroid-derived 2-related family consists of four ubiquitously expressed TFs (**Figure 1.7**), namely Nuclear factor erythroid 2 (NFE2) p45, Nrf1 (NFE2L1), Nrf2 (NFE2L2), and Nrf3 (NFE2L3)<sup>29, 60</sup>. Nrf1, Nrf2, and Nrf3 play roles in many cellular processes, such as protecting cells against oxidative stress or inflammatory responses<sup>65-70</sup>, controlling protein degradation through proteasome gene regulation<sup>14, 65, 69, 71-74</sup>, and being involved in glucose-glutamine metabolism<sup>66, 69</sup>, carcinogenesis<sup>14, 67-70, 73-78</sup>, embryo growth<sup>66, 69, 79-80</sup>, development and tissue differentiation<sup>67, 69, 80</sup>. These TFs belong to the CnC family and have conserved basic leucine zipper (bZIP) region at their C-terminus domain (CTD), comparable to the SKN-1 of yeast<sup>29, 60, 65</sup>.

The bZIP region heterodimerizes with small-musculoaponeurotic fibrosarcoma (sMaf) proteins (MafF, MafG, or MafK). It forms a heterodimer complex of bZIP - sMaf that allows a subsequent binding to the Anti-oxidant response element (ARE) of proteasome genes<sup>29, 60, 65</sup>. ARE and Nrf1 have a similar binding site with a core sequence 5'-RTGACnnnGC-3' (R = A or G, nnn = TCA in Nrf1) in the promoter/enhancer of all proteasome subunits<sup>29</sup>, including the proteasome assembly chaperones<sup>60</sup>. Furthermore, the N-terminus domain (NTD) of mammalian Nrf1 consisted of the trans-membrane domain (TM), the processing domain (PS), and the seven N-X-S/T (Asparagine-X-Serine/Threonine) rich regions<sup>60</sup> (**Figure 1.7**). The NTD-TM domain of Nrf1 serves as a recognition site for a targeted reticulum endoplasm (ER)-ribosome bound Nrf1 synthesis; therefore, even though the NTD of Nrf1 facing outwards to the cytosol, it is still co-translationally tethered to the ER membrane<sup>29, 60, 65, 72, 81-82</sup>. Meanwhile, the remaining Nrf1 polypeptide, including their CnC-bZIP of the CTD, is facing inwards to ER lumen, enabling Asparagine N-Glycosylation (Asn) on its NST regions<sup>29, 60, 65, 82</sup>. More precisely, Nrf1 glycosylation can occur at eight different sites in the NST region<sup>82</sup>. This N-glycosylated Nrf1 is further ubiquitylated by complex Hrd3-Hrd1/synoviolin (ER-resident E3 ligase)<sup>29, 60, 65, 67</sup>. Then, rapid retro-translocation occurs to the cytosol through homo-hexameric AAA<sup>+</sup> ATPase segregate p97/VCP (valosin-containing protein)<sup>29, 60, 65, 72, 83</sup>. Once in the cytosol, the N-term PUB of the NGLY1 N-glycanase (human PNGase) domain recognizes the protein bound to p97/VCP, and the C-terminal PAW domain of NGLY1 can identify the ERAD-mediated glycosylated protein<sup>82</sup>. It will catalyze the hydrolysis of the N-acetylglucosamine (GlcNAc) from the Nrf1 glycogen amide bond and convert Nrf1 asparagine (-N-x-

S/T-) to aspartate residue (-D-x-S/T-)<sup>82</sup>. This process again reveals the Nrf1 full-length form (p140, p120 kDa inactive isoforms)<sup>60, 83</sup>. Under normal protein homeostasis, this full-length Nrf1 will be rapidly degraded by the ER-associated proteasome with a half-life ( $t_{1/2} < 30$  minutes)<sup>29, 60, 65, 72</sup>. See **Figure 1.8** to illustrate the regulation of proteasome genes under normal conditions.

In addition to cytosolic ERAD degradation, the processed Nrf1 can be degraded, stabilized, and activated once in the nucleus<sup>29, 60, 84</sup>. Serine phosphorylation at LFSPEVE binding site motif of Fbw7 E3 ligase (FBXW7) or DSGLS binding site motif of Beta-transducin repeat containing E3 ligase ( $\beta$ -TrCP) contributes to Nrf1 degradation in the nucleus<sup>29, 60, 66, 84</sup>. On the contrary, increasing O-GlcNAcylation on the serine of  $\beta$ -TrCP further promotes Nrf1 nuclear stabilization and transactivation<sup>66, 69, 84</sup>. Similarly, Deubiquitinase (DUB) USP15 and Kelch-like ECH-associated protein 1 (KEAP1) contribute to the nuclear - Nrf1 stabilization<sup>29, 60, 66</sup>. Additionally, RUVBL1 and TIP60 were reported to modulate the proteasome genes via complex binding with Nrf1<sup>85</sup>. Yet, these two genes were incapable of directly processing Nrf1<sup>85</sup>.

Alongside Nrf1 and Nrf2, TCGA and atlas of human cancer genome data show that Nrf3 expression is abundant in various cancer cells<sup>67-68, 73-78</sup>. Nrf3 is a short-lived protein ( $t_{1/2} \sim 20 - 40$ min)<sup>68</sup>, with the closest structural homology to Nrf1<sup>75</sup>. The binding region of Nrf3 to ARE is similar to that of Nrf1 and retains the TM, PS, NST, CnC, and bZIP domains<sup>73, 75</sup>. Nrf3 is rapidly degraded in the cytosol by ERAD following ubiquitylation by Hrd1 and retrotranslocation by p97/VCP<sup>67, 73, 75</sup>. While in the nucleus, Nrf3 is degraded upon serine phosphorylation to  $\beta$ -TrCP or FBXW7

binding site<sup>67-68</sup>. Studies also reported multiple pathways of Nrf3 regulation<sup>67-68, 73-75, 78</sup>, including the  $\beta$ -catenin/ T-cell factor 4 (TCF4) complex and tumor necrosis factor  $\alpha$  (TNF $\alpha$ ) NF- $\kappa$ B<sup>76-77</sup>.

Although overexpression of Nrf3 alone did not alter the mRNA or protein levels of most proteasome subunits, in a separate study, Waku *et al.* demonstrated that Nrf3 could increase the mRNA level of proteasome maturation proteins (human POMP, yeast ortholog UMP1)<sup>70, 73, 75</sup>. POMP serves as a chaperone to assemble the initial alpha subunits ring formation of the 20S CP<sup>73, 75</sup>. Concurrently, they also found that overexpressing Nrf3 enhances the activity of the 20S proteasome but not the 26S proteasome<sup>73, 75</sup>. Nrf3 may enhance the 20S proteasome activity by promoting 20S proteasome assembly via inducing POMP expression but not 20S proteasome subunits<sup>73, 75</sup>. Increased 20S proteasome activity further reduced the sensitivity of cells to proteasome inhibitors<sup>73, 75</sup>. At the same time, inducing Nrf3 activates homological motifs U2AF serine/threonine protein kinase 1 (UHMK1 or KIS1), increases cell proliferation, elevates glucose transporter (GLUT1) or Vascular Endothelial Growth Factor A (VEGFA) expression, and lowers tumor suppression Rb and p53 protein level<sup>67, 70, 73-76</sup>. Both of these contributed to tumorigenesis<sup>70, 73</sup>. Silencing Nrf3 inhibits the cell's growth and increases apoptotic cells<sup>73, 77</sup>.

Moreover, since Nrf3 is the closest homolog to Nrf1, their functions could overlap and complement each other. Recently, a study found that Nrf1 and Nrf3 worked in synchronization to maintain the basal level of the proteasome in cancer cells<sup>70, 75</sup>. Overexpression of Nrf3 decreases the Nrf1 protein level but does not alter the Nrf1 mRNA level<sup>70, 75</sup>. Nrf3 upregulates Cytoplasmic polyadenylation element 3

(CPEB3) and further suppresses Nrf1 polysomes mRNA translation<sup>70, 75</sup>. Meanwhile, overexpression of Nrf3 alone does not alter the mRNA or protein level of most proteasome subunits<sup>73, 75</sup>. When Nrf3 protein level is decreased, in turn, the Nrf1 protein level increases and further induces the mRNA proteasome genes<sup>75</sup>. Furthermore, knocking down both Nrf1 and Nrf3 can reduce the mRNA basal level of the proteasome subunits (PSMB3, PSMB7, PSMC2, and PSMD3) and the proteasome assembly chaperone (PSMG2, PSMG3, and POMP)<sup>75</sup>.

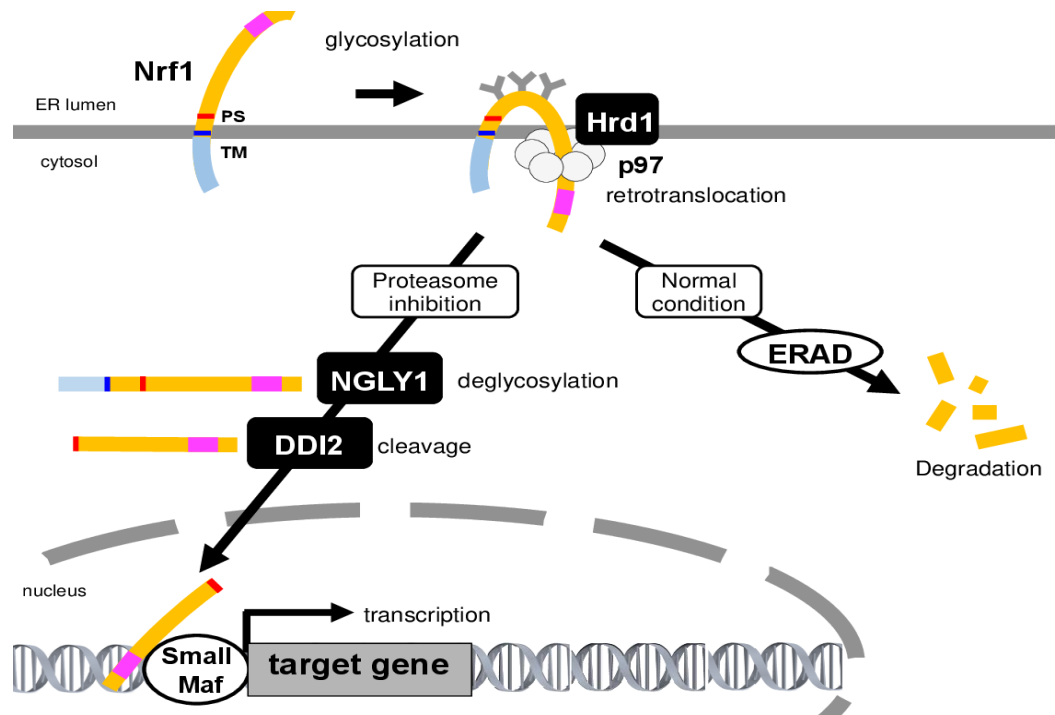
Nrf1 KO mice resulted in late gestational embryonic lethality<sup>66, 69, 79-80, 84</sup>, while deletion of Nrf3 in mice showed no visible abnormalities<sup>70, 74</sup>. Although Nrf3 does not appear significant in embryonic growth, Nrf3 became a hotspot mutation in various cancers and correlated with poor cancer prognosis based on the human cancer genome TCGA data and Kaplan Meier curves<sup>70, 74</sup>.

### **1.6.3 Regulation of proteasome genes in response to proteasome inhibition**

In yeast, inhibiting the proteasome stabilizes Rpn4 and subsequently enhances the synthesis of proteasome subunits<sup>29, 31, 58, 60</sup>. Meanwhile, the SKN-1 of *C. elegans* coordinates a protective response when protein degradation is impaired<sup>61</sup>. For instance, SKN-1A upregulates the mRNA proteasome genes when the proteasome activity is blocked<sup>29, 31, 61</sup>. However, when the translation initiation and elongation are perturbed, SKN-1 will not upregulate proteasome genes, resulting in a reduction in the activity of the proteasome<sup>61</sup>.

In mammals, silencing the nuclear transcription factor Y (NF-YA) reduces cellular proteasome activity (~50%) and cell viability (~30%) upon treatment with a proteasome inhibitor<sup>62</sup>. Likewise, several studies reported Nrf1 and TCF11 (a

longer isoform of Nrf1) in modulating the induction of mRNA proteasome subunits when the activity was partially inhibited<sup>15, 29, 60, 65, 72, 86</sup> and further accelerated the recovery of proteasome activity (the bounce-back response) via *de novo* proteasome synthesis<sup>60, 72</sup>.



**Figure 1.8 Regulation of Nrf1 transcription factor in the ER, cytosol, and nucleus.** This Figure is reproduced from Hamazaki, J. and Murata, S. (2020)<sup>60</sup> with copyright permission.

Under a condition when the protein homeostasis is disrupted through proteasome inhibition or increased proteotoxic stress, the full-length Nrf1 will be processed (p110, p95, p85, p65 kDa active isoforms) and translocated to nucleus<sup>29, 60, 83</sup>. See **Figure 1.8** to illustrate the regulation of proteasome genes when the proteasome function is hampered. Nrf1 processing occurred between W<sub>103</sub> and L<sub>104</sub> of the PS regions (see **Figure 1.7**). The bZIP domain of processed Nrf1 was dimerized with sMaf and formed stable bounds with ARE of proteasome genes<sup>29, 60, 65, 72, 86</sup>. This, in turn, contributes to mRNA upregulation of proteasome genes<sup>29, 60, 72, 86</sup>.

The Nrf1 processing and translocation to the nucleus occurred upon partial or total proteasome inhibition<sup>71, 87</sup>, but the amount of proteins in the soluble fraction varies. At first instance, treatment of cells with proteasome inhibitors at a concentration  $\leq 0.5$   $\mu\text{M}$  and duration of inhibition  $< 4$  h leads to an accumulation of processed Nrf1 in the soluble fraction<sup>87</sup>. Meanwhile, when the cells are treated with high concentrations ( $\geq 1$   $\mu\text{M}$ ) of proteasome inhibitors and the duration of treatment is longer ( $\geq 4$  h), it leads to more potent inhibition of proteasome and the accumulation of processed Nrf1 in the cytosol or nucleus as aggregates rather than in soluble fraction<sup>87</sup>.

Moreover, Vangala *et al.* found that the Nrf1 processing and mRNA proteasome genes induction still occur when the protein synthesis is inhibited after complete proteasome inhibition suggesting that Nrf1 can also be activated independently of proteasome activity<sup>71</sup>. When proteasome is inhibited, the Nrf1 processing may be carried out by an endoprotease<sup>87</sup>. Conversely, when Nrf1 is deleted, cells cannot induce mRNA proteasome genes<sup>71</sup>, and the mRNA level of proteasome genes is reduced by approximately 20% in mice<sup>14</sup>. Hence, deletion or depletion of Nrf1 also reduces *de novo* proteasome synthesis in cells and further increases the sensitivity to proteasome inhibitors<sup>14, 84</sup> and the accumulation of polyubiquitylated proteins<sup>84</sup> due to the inability of cells to restore proteasome activity.

In addition to cytosolic ERAD degradation, the processed Nrf1 can be degraded, stabilized, and activated once in the nucleus<sup>29, 60, 84</sup>. For instance, under inhibited proteasome activity, silencing RUVBL1/ RUVBL2 or TIP60 results in the inability

to induce the mRNA level of several proteasome genes (such as PSMA7, PSMB7, PSMC4, and PSMD12)<sup>85</sup>.

To summarize, several studies found that activation of the ER-resident Nrf1 has triggered a 'bounce back response' in cancer cells and increased proteasome activity after proteasome inhibition<sup>14, 29, 60</sup>. This contributes to a brief positive response in patients treated with single-agent proteasome inhibitors or in cancer recurrence cases, implying the need for effective therapy. Hence, controlling the level of Nrf1 may be necessary to avoid proteasome bounce-back responses. Inhibiting p97/ VCP decreases the Nrf1 retrotranslocation to the cytosol. Nrf1 can also be inhibited by inhibiting Nrf1 deglycosylation by NGLY1 or inhibiting Nrf1 processing and its activation. The latter has a direct effect on proteasome gene upregulation. Recently, a few studies also reported the role of DNA Damage Inducible 1 homolog II (DDI2, a twin homolog of DDI1) in the processing and activation of Nrf1 and Nrf3.

### **1.7 DDI2 and its role in the proteasome activity recovery**

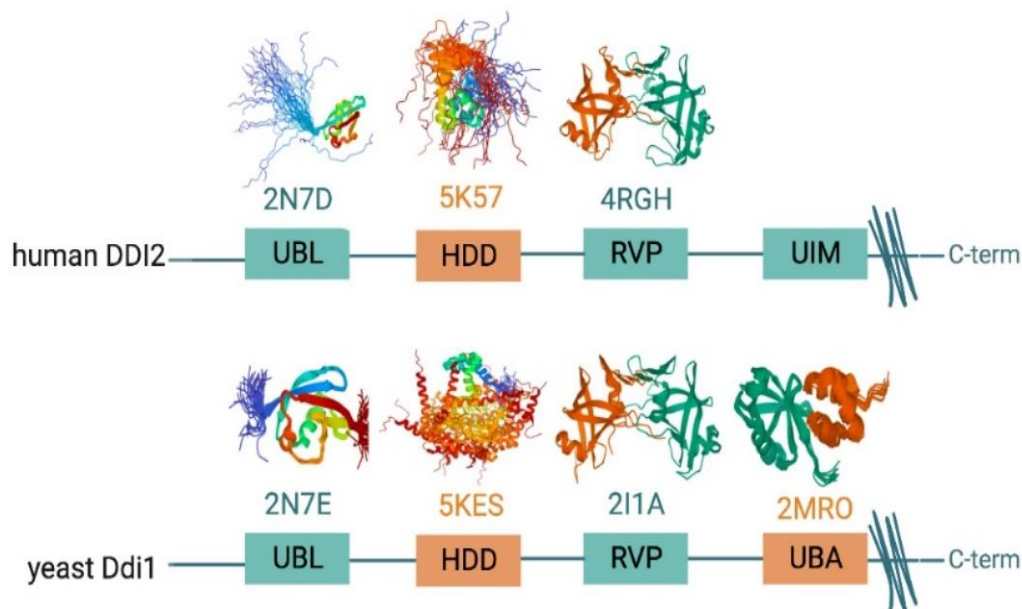
Human DNA Damage Inducible 1 homolog II (hDDI2) is a homolog of hDDI1 and belongs to the Ddi1-like protein family<sup>16</sup>. However, unlike DDI2, the function of the Ddi1 protein has been extensively studied in yeast<sup>22-23, 88-91</sup>. yDdi1 acts as a ubiquitin-binding shuttling protein, along with hRAD23 (yRad23) and Ubiquilin (yDsk2)<sup>22, 89-90</sup>. All of these shuttling proteins can carry ubiquitylated proteins to the regulatory part of 26S proteasome<sup>22, 89-90</sup>. Likewise, hDDI2 has also been reported to function as a shuttling protein<sup>16-17</sup>. The unusual feature of hDDI2 and yDd11 is that it is the only other 'protease' beyond the proteasome that binds the polyubiquitylated proteins<sup>17, 89</sup>. In fact, the requirement of substrate ubiquitylation



appears to be much stronger in DDI2 than in proteasome. Unlike proteasomes, DDI2 cannot cleave peptide substrates and non-ubiquitylated proteins<sup>24</sup>. Furthermore, the proteasome degrades the ubiquitylated proteins into small peptides<sup>24</sup>. Meanwhile, DDI2 can only precisely cleave single peptide bonds of polyubiquitylated substrates<sup>17, 89</sup>. In this section, we outline the structure and known function of hDDI2 and then emphasize the importance of hDDI2 in the UPS systems.

### **1.7.1 Structure of hDDI2 and its function**

Structurally, hDDI2 is a dimer comprised of two identical multi-domain polypeptides<sup>16</sup>. It consists of an N-terminal ubiquitin-like domain (UBL) and a C-terminal ubiquitin-interacting motifs domain (UIMs) resembling the ubiquitin-associated domain (UBA) in Ddi1<sup>16, 20</sup> (**Figure 1.9**). These two domains mediate binding with ubiquitin and contribute to its function as a protein shuttle<sup>16-17</sup>. It also has a conserved  $\alpha$ -helical domain of DDI (HDD) and a retroviral protease domain (RVP)<sup>16-17, 20</sup>. These domains are responsible for Nrf1 processing<sup>17</sup>.



**Figure 1.9 Domain structure of hDDI2 and yDdi1.** Human DD12 differs from yeast Ddi1 in the c-term domain. hDDI2 has UIM domain, while yDdi1 has UBA domain. The 3D structure PDB of the available domains is included. UBL is the ubiquitin-like domain, HDD is the  $\alpha$ -helical domain of DDI, RVP is the Retroviral protease domain, UIM is the ubiquitin interacting motif domain, and UBA is ubiquitin associated domain. This Figure is created with BioRender.com.

### 1.7.2 The RVP domain determines the Ub-endoprotease properties of DD12

The proteolytic activity of hDDI2 is largely due to its retroviral protease domain, which distinguishes Ddi1-like proteins (yDdi1, hDDI2) from other shuttling proteins. Both hDDI2 and yDdi1 share an 81% similarity (contains two  $\alpha$ -helices, six  $\beta$ -barrel lobes, and three  $\beta$ -sheets)<sup>16</sup>. One crucial difference lies in the signature motifs at their active catalytic sites where hDDI2 has serine (D<sub>252</sub>-Ser<sub>253</sub>-Gly<sub>254</sub>-Ala<sub>255</sub>) instead of threonine (D<sub>220</sub>-Thr<sub>221</sub>-Gly<sub>222</sub>-Ala<sub>223</sub>)<sup>16</sup>. These two Thr or Ser are necessary to form several hydrogen bonds to the amide backbone of its dimer resembling 'The Fireman's grip' shape<sup>16, 88, 90</sup>. The signature motif forming a homodimerization between the RVP domain, a water molecule, and a hydrophobic  $\beta$ -sheet platform<sup>16, 88</sup>. The water molecule acts as a nucleophile assisting the hydrolysis of the polypeptide bond after hydrophobic residues<sup>16, 88</sup>. The hydrophobic  $\beta$ -sheet platform function brings both RVP domains in proximity,

particularly the catalytic Asp residues<sup>88</sup>. This signature (D-T/S-G-A) motif is also preserved in several Ddi1-like proteins such as Ddi1 of *S. cerevisiae*<sup>90</sup>, Ddi1 of *Leishmania sp.* (*L. major*, *L. infatum*, *L. mexicana*, *L. braziliensis*)<sup>16, 92-93</sup>, DDI1 in *C. elegans*<sup>16</sup>, Rngo in *Drosophila melanogaster*<sup>16</sup>, *Mus musculus* and *Rattus norvegicus* DDI1 and DDI2<sup>16, 90</sup>, hDDI1 and hDDI2<sup>16, 88, 90</sup>.

Interestingly, the signature motif of the hDDI2 RVP domain is also identical to the HIV protease catalytic active site<sup>16, 88</sup>. Hence, Ddi1-like protein is thought to have similar proteolytic properties to HIV aspartic proteases. Unfortunately, only two studies successfully reported the proteolytic activity of Ddi1, both of which are in *L. major*<sup>16, 91-93</sup>. They showed that recombinant Ddi1-like protease could cleave HIV protease substrate RE-(EDANS)-SQNYPIVQK-(DABCYL)-R and Cathepsin D substrate Bz-RGFFL/P-4MBNA at acidic pH (range between 4 - 6.5, and optimum at pH 5.5)<sup>16, 92-93</sup>. They also showed that the recombinant Ddi1-like protease of *L. major* could be inhibited by HIV protease inhibitor Nelfinavir (NFV, 20uM), a broad class-specific inhibitor of aspartic proteases Pepstatin (15uM), and a covalent inhibitor of aspartic proteases Diazoacetyl-DL-norleucine methyl ester (DAN, 500uM) with a percentage of inhibition 60%, 70%, and 95%, respectively<sup>16, 92-93</sup>.

Contrary to the observed proteolytic activity in *L. major*, Sivá *et al.* reported no proteolytic cleavage of Bovine/ Human serum albumin,  $\beta$ -casein, and HIV protease substrate (VSFSFPQITL, ac-KARVLAEM-NH<sub>2</sub>) by hDDI2 at pH 5 and 7. It can be argued that the preferred pH difference for RVP proteolytic activity is highly dependent on substrate localization, and RVP tends to cleave cytoplasmic

substrates<sup>92</sup>. For example, a study reported that the RVP protease yDdi1 suppresses the cell-secreted proteins *in vivo* by inhibiting SNARE protein<sup>91</sup>.

Moreover, Siva *et al.* did not observe any inhibition of hDDI2 with various HIV protease inhibitors such as amprenavir, atazanavir, brexanavir, darunavir, indinavir, nelfinavir, ritonavir, saquinavir, phosphonate-containing GS-8374, and acetyl-pepstatin<sup>16</sup>. The inability of HIV protease inhibitors to bind with the hDDI2 RVP domain may be due to differences in the size of the cavity and the opening of hDDI2 compared to HIV protease<sup>16, 88</sup>. The cavity and the opening of the RVP domain of yDdi1 and hDDI2 are larger (~22Å) and consist of hydrophobic patches, implying that both yDdi1 and hDDI2 have a larger substrate binding groove than HIV proteases allowing bulkier substrates into their cavities, with substrate preferences for hydrophobic interactions at P1 position<sup>16, 88</sup>. At the same time, the P2 residue can be an amino acid that is able to form a Hydrogen bond donor or acceptor<sup>88</sup>. Of note, yDdi1 and hDDI2 may have different substrates due to differences in their binding affinity to ubiquitin determined by their UBA/UIM and UBL binding and the overall surrounding charges inside their cavity.

Nevertheless, the yDdi1 RVP catalytic domain was not active in the absence of HDD and UBL domains<sup>89</sup>, in contrast to that observed in *L. major*. The same is true for hDDI2 RVP domain proteolytic activity<sup>39</sup>. Both proteolytic activities of yDdi1 and hDDI2 require both RVP and HDD domains<sup>39, 89</sup>. At the same time, aspartic acid at the signature motif is essential because mutation from aspartic acid to asparagine in yDdi1 (D220N) or hDDI2 (D252N) deprives its proteolytic activity in cells<sup>17, 89</sup>.

Recently, two studies demonstrated that **yDdi1 and hDDI2 function as ubiquitin-dependent endoproteases**<sup>17, 20, 89</sup> and cleave the long-chain polyubiquitylated substrates to balance the cells from accumulating highly-polyubiquitylated proteins<sup>17, 89</sup>. This ubiquitylation binding is needed for substrate recognition of Ddi1-like protein, implying a crucial role of the UBL domain<sup>89</sup>. yDdi1 tends to cleave the substrate characterized by a minimum of eight ubiquitin long chains and prefers chains of 15 - 25 ubiquitin long<sup>89</sup>. This will allow the substrate to be in proximity to the large cavity of Ddi1, and the same is true for DDI2. In contrast to the functional Ddi1-like proteins, inactive hDDI2 protease and KO hDDI2 led to a significant accumulation of high MW polyubiquitylated proteins in mice and endothelial cells<sup>20</sup>.

To date, Nrf1 is the only characterized DDI2 substrate. hDDI2 processed the cytosolic deglycosylated Nrf1 into soluble fractions when proteasome function is impaired, allowing their translocation to the nucleus<sup>14-15, 20, 29, 38, 60, 65, 72, 86, 94</sup>. As a result, Nrf1 can actively bind with sMaf ARE to induce proteasome gene expression<sup>14-15, 29, 38, 60, 65, 72, 86, 94</sup>. A study by Yip *et al.* confirmed the findings previously reported by Radhakrishnan *et al.*, stating that the Nrf1 putative cleavage site is between W<sub>103</sub> and L<sub>104</sub> (of NAW<sub>103</sub>-L<sub>104</sub>VH peptide), and W<sub>103</sub> is essential for protease recognition<sup>65, 89</sup>. Purified yDdi1 also cleaved the engineered polyubiquitylated Nrf1 between W<sub>103</sub> and L<sub>104</sub>, proving the direct protease activity of Ddi1-like protein on Nrf1<sup>65, 89</sup>. At the same time, Dirac-Svejstrup *et al.* confirmed that Nrf1 is processed between W<sub>103</sub> and L<sub>104</sub> by hDDI2 in a ubiquitylation-dependent fashion<sup>17</sup>.

### **1.7.3 The UBL – UIMs domains contribute to hDDI2 function as protein shuttle**

The ubiquitin-binding protein shuttle is characterized by its capacity to bind with polyubiquitin and proteasome. The prominent character of this ubiquitin-binding shuttling protein is primarily contributed by the UBL domain and UBA/UIMs domains. As a shuttle or an adaptor, it interacts with the ubiquitylated substrate via the UBA/UIMs domain and with the 26S proteasomes via the UBL domain.

Likewise, Ddi1-like proteins bind the polyubiquitylated substrates via the UBA or UIM domains and to the proteasome via the UBL domain<sup>16-17, 95</sup>. These two domains bind ubiquitin with different binding affinity<sup>16, 20</sup>. The hDDI2 UIMs domain binds to ubiquitin with a much weaker affinity than the yDdi1 UBA domain<sup>16, 20</sup>. The yDdi1 UBA domain has an extra function for mediating the heterodimerization with yRad23 and homodimerization of yRad23 but is unable to mediate the homodimerization of yDdi1 itself<sup>96</sup>. The capacity of yDdi1 to promote heterodimerization with yRad23 implies that these two shuttling proteins may work in synchronization<sup>90</sup>.

Meanwhile, the UBL domain of hDDI2 binds to ubiquitin with a weaker affinity than the yDdi1 UBL domain<sup>16, 90</sup>. This may arise due to their sequence differences, where they only bear ~46% in common<sup>16</sup>. The yDdi1 UBL domain binds to ubiquitin via hydrophobic surface patches and salt bridges<sup>90</sup>. The surrounding negatively charged  $\beta$ -sheet surfaces help stabilize this binding<sup>90</sup>. The UBL-ubiquitin complex stabilization via a negatively charged  $\beta$ -sheet surface is a feature of yDdi1 and is not preserved in yRad23 or yDsk2<sup>90</sup>. Therefore, the yDdi1 UBL domain can bind with high affinity and deliver the long-chains/bulky

polyubiquitylated substrate to the proteasome<sup>89</sup>. In contrast to yDdi1, the  $\beta$ -sheet surface of the hDDI2 UBL domain consisted of smaller hydrophobic patches surrounded by moderate negative charges<sup>16</sup>. Despite these differences, both UBL and UBA/UIMs domains have played a part in determining the function of yDdi1 and hDDI2 as ubiquitin shuttling factor proteins.

In addition, several studies reported the practical advantage of Ddi1-like protein function as ubiquitin shutting factors in which **Ddi1-like proteins play a role in maintaining genome integrity in response to DNA replication stress and drug-induced DNA-protein crosslinks (DPCs)**. yDdi1 mediates the regulation of the cell cycle progression and prevents premature mitosis through its UBA - UBL domain<sup>96-97</sup>. At the same time, yDdi1 protease serves as a complement repair pathway to metalloprotease Wss1 (hSPRTN) in response to DNA-replication stress by cleaning out DPCs-induced Topoisomerase 1 (Top1) covalent cleavage complex (Top1cc) from the replication fork and removing chromatin-associated proteins that covalently crosslinked to DNA<sup>22-23</sup>. yDdi1 also promotes an efficient renewal of stalled DNA-directed RNA polymerase II subunit (Rpb1) from the replication fork<sup>22-23, 96</sup>.

Like yDdi1, hDDI1 and hDDI2 protease interchangeably play a role in DNA-damage response by removing the C20orf43/ replication termination factor 2 (RTF2) from the stalled fork. Removing the RTF2 is needed to restart the replication fork and maintain an effective DNA replication rate. This, in turn, promotes genome integrity<sup>21</sup>. RTF2 on the stalled fork itself negatively impacts the fork restart by continuously activating the DNA damage signal causing

chromosomal instability<sup>21</sup>. Cells accumulate ssDNA when constantly exposed to DNA damage signals<sup>21</sup>. Therefore, Knockout of DDI2 increases the cell sensitivity toward DPCs agents (Formaldehyde, UV or ion radiation, protein crosslinker, Hydroxyurea, ROS, and metal ions)<sup>21-23</sup>. Also, it increases cell sensitivity to gyrase/topoisomerase inhibitors, cyclin-dependent kinase inhibitors, and DNA/ RNA polymerase inhibitors<sup>21-23</sup>.

**The Nrf1 processing by hDDI2 proteases has raised questions about the biological importance of hDDI2 itself.** As described in sections **1.6.2** and **1.6.3**, on the one hand, Nrf1 knockdown lowers the basal level of the proteasome genes. It represses the upregulation of the proteasome subunits under impaired proteasome function. Induction of proteasome subunits genes requires Nrf1 translocation and activation. At the same time, hDDI2 processes the cytosolic Nrf1 and translocates it to the nucleus to induce proteasome gene expression<sup>14-15, 20, 29, 38, 60, 65, 72, 86, 94</sup>. KO of DDI2 or inactivation of protease DDI2 decreases the chymotrypsin-like activity and lowers proteasome 26S/30S level after Btz treatment<sup>20</sup>.

Moreover, Knockout of Nrf1 or hDDI2 resulted in mid-late-stage embryonic lethality, growth retardation, and developmental defect in mice<sup>20, 66, 69, 79-80, 84</sup>. Mutating the hDDI2 active site also interferes with embryonic development and shows an obvious sign of growth retardation<sup>20</sup>.

On the other hand, increases in the mRNA level of proteasome genes further induce *de novo* synthesis of proteasome subunits and the recovery of proteasome activity. This, in turn, contributes to reducing proteasome inhibitor therapeutic efficacy, particularly in proteasome-dependent cancer. Following this, some studies reported



the dependency of Multiple myeloma (MM), a cancer of plasma cells, on hDDI2-Nrf1 mediated response to proteasome inhibition<sup>38-39</sup>. Both of these studies found that the basal level of Nrf1 is consistently high and even strongly expressed across Multiple myeloma and AL amyloidosis cells<sup>38-39</sup>. These cells accumulated high molecular weight polyubiquitin without treatment with proteasome inhibitors<sup>38-39</sup>. Therefore, it is not surprising that KO of hDDI2 alone is cytotoxic to myeloma cells and aggravated under impaired proteasome function<sup>38-39</sup> due to increased proteotoxic stress.

Interestingly, they found reconstituting hDDI1 in hDDI2 KO cells restores the cell's ability to process Nrf1 to a similar level as hDDI2, suggesting the possible complementary functions of hDDI1 and hDDI2<sup>39</sup>. Several UPS pathways were activated to compensate for impaired hDDI2 function *in-vitro* or *in-vivo*<sup>20</sup>. These include induction mRNA of Rad23a - Rad23b - Ubiquilin shuttling factors, NGLY1, Nrf1, and Nrf2, and immune-proteasome<sup>20</sup>. It can also elevate the mRNA and protein level of the UPR marker (ATF4, Chop, GADD34, pPERK) involved in global translation<sup>20</sup>. Moreover, it can upregulate Herpud1 (Homocysteine inducible ER protein with ubiquitin-like domain 1) and types I interferon signaling (IFN- $\beta$  and STAT3). Meanwhile, the loss of hDDI2 function can also lead to forming of alternative proteasome complexes, which increase chymotrypsin-like 20S activity<sup>20</sup>.

To our knowledge, the function of hDDI2 in the ubiquitin-proteasome system results in only slight differences in proteasome recovery between hDDI2 KO and parental cells (Triple-negative breast cancer<sup>94</sup> and colorectal cancer<sup>15</sup>), raising a

question on the significance of hDDI2 in the recovery of proteasome activity. If hDDI2 is essential for Nrf1 processing and Nrf1 activation is a primary determinant for the induction of proteasome genes when the proteasome is inhibited, disabling hDDI2 should result in a >50% reduction in the recovery of proteasome activity.

**In our study, we sought to ask whether the deletion of hDDI2 will directly affect the recovery of proteasome activity in a time-dependent manner using relevant subtoxic doses in which the cells are not under continuous proteotoxic stress.**

## Chapter 2 DDI2 is not Required for the Proteasome Activity Recovery

### 2.1 Introduction

Cancer cells experience higher proteotoxic stress than normal cells due to the accumulation of abnormal proteins that further increase their dependency on the Ubiquitin-proteasome system (UPS) clearance<sup>5</sup>. Hijacking the UPS system using proteasome inhibitors negatively affects the survival of cancer cells more than normal cells<sup>5</sup>. To date, proteasome inhibitors are clinically approved and have shown greater efficacy in treating immunoglobulins-expressing cancers, Multiple myeloma (MM), and Mantle Cell Lymphoma (MCL); yet, only limited efficacy in other cancers<sup>12, 28</sup>. In addition, a growing body of evidence suggests intrinsic resistance or relapse cases to proteasome inhibitors due to the induction of proteasome genes/ubiquitin pathway components (p97 and its cofactor)<sup>86, 98-99</sup> and subsequently generate a proteasome 'bounce-back response.' It establishes the need to understand and target the recovery pathway of proteasome activity.

As discussed in the previous chapter, DDI2 is involved in the up-regulation of proteasome genes via transcriptional regulation by Nrf1<sup>14</sup> and is implicated in resistance to proteasome inhibitors. However, it has not been shown that this upregulation leads to the recovery of proteasome activity.

This has raised questions about the direct biological function of DDI2 protease in proteasome activity recovery, which we will address in this study. We will use reverse genetic approaches DDI2 RNAi and DDI2 CRISPR knockout cells to answer our questions.

## 2.2 Results

### 2.2.1 Knockout and knockdown of DDI2 do not impair proteasome activity recovery in inhibitor-treated cells

To answer whether DDI2 plays a substantial role in proteasome activity recovery, we first asked whether the deletion of DDI2 affects the recovery of activity after pulse treatment with proteasome inhibitors. Here, we utilized DDI2 CRISPR knockout clones of HAP1 cells (**Table 2.1**). HAP1 cells are near-Haploid cells derived from chronic myelogenous leukemia (CML) KBM-7 cell lines<sup>100-102</sup>. Despite its origin, this cell line is an adherent-type cell with a fibroblast-like morphology that behaves like solid tumors<sup>100-102</sup>. This cell line is often used in CRISPR experiments because a highly efficient knockout level can be achieved by modifying only one allele<sup>100</sup>. We obtained three different DDI2 CRISPR KO clones from two different guide RNAs (gRNA) to exclude the off-target effect of CRISPR. Before conducting the experiments, we confirmed that DDI2 is not expressed in all knockout cells (**Figure 2.1 A**).

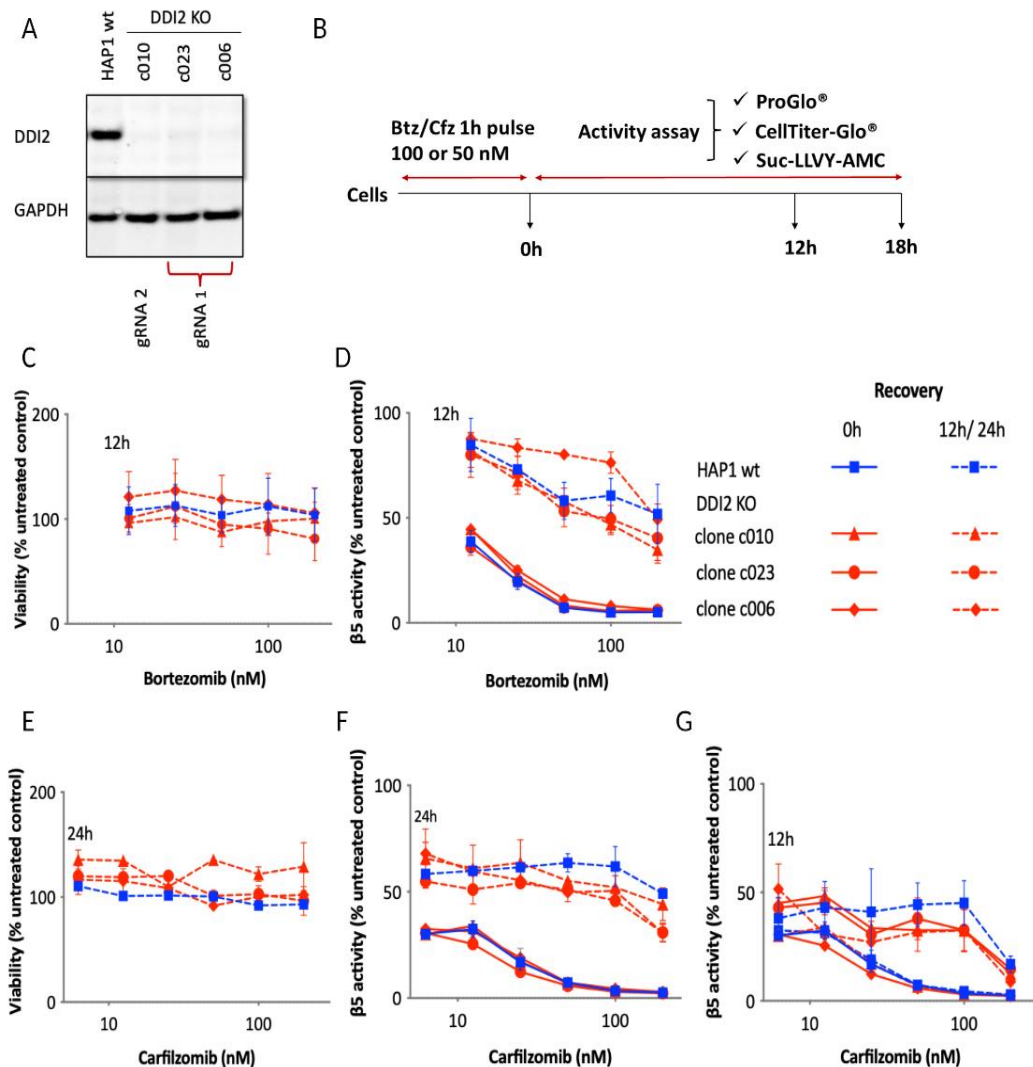
CRISPR KO gene editing works by employing Cas9 endonuclease and guide RNA<sup>103</sup>. A guide RNA (gRNA) recognizes the target DNA region of interest and ensures the specificity of gene editing by directing the Cas9 endonuclease to break the DNA double-strand at specific sites<sup>103</sup>. These Double-strand DNA breaks (DSBs) can be repaired by homology-directed repair (HDR) or nonhomologous end joining (NHEJ), which often generates indel mutations that often cause frameshifts<sup>103</sup>. Frameshift deletion usually leads to the disruption of gene functions and generates off-target effects<sup>103</sup>.

Next, we selected a treatment setting for our *in-vitro* study according to previously reported PK and PD studies on clinical doses of Bortezomib (Btz) or Carfilzomib (Cfz). Based on these studies, the maximum proteasome inhibition was achieved within an hour (reaching more than 90% inhibition) after bolus drug administration<sup>11, 46-47</sup>. Therefore, we chose a one-hour treatment in this *in-vitro* study. After treatment with inhibitors, cells were allowed to recover in a drug-free medium, and the proteasome activity was measured at different times (**Figure 2.1 B**).

As described in **Chapter 1, Sections 1.6.1 and 1.6.2**, treatment with Btz or Cfz at clinically achievable concentrations inhibits the 20S proteasome<sup>47, 50</sup>. Hence, we treated cells with Btz and Cfz at concentrations inhibiting 20S proteasome. Similar to the reported clinical study<sup>47, 50</sup>, we only measured the  $\beta 5$  proteasome activity using luminescence-based Suc-LLVY-aminoluciferin and fluorescent-based Suc-LLVY-7-amido-4-methylcoumarin substrates (Suc-LLVY-AMC). To ensure that the reduction of proteasome activity is not due to the cell losing its viability, we measured the ATP content in the cells (using CellTiter-Glo™, **Figure 2.1 C and E**). We measured ATP simultaneously with the  $\beta 5$  proteasome activity measurement (using Proteasome-Glo™, **Figure 2.1 D, F, and G**).

Treatment with Btz or Cfz ( $\geq 50 - 200$  nM) resulted in more than 90%  $\beta 5$  inhibition (**Figure 2.1 D, F, and G**). Proteasome activity recovery was measured 12 – 24 hours after treatment, showing  $\geq 50\%$  of  $\beta 5$  activity recovery. The recovery of activity was similar in cells treated with Cfz (an irreversible inhibitor) and Btz (a reversible inhibitor with a slow-off rate), indicating that the activity recovery is not due to the

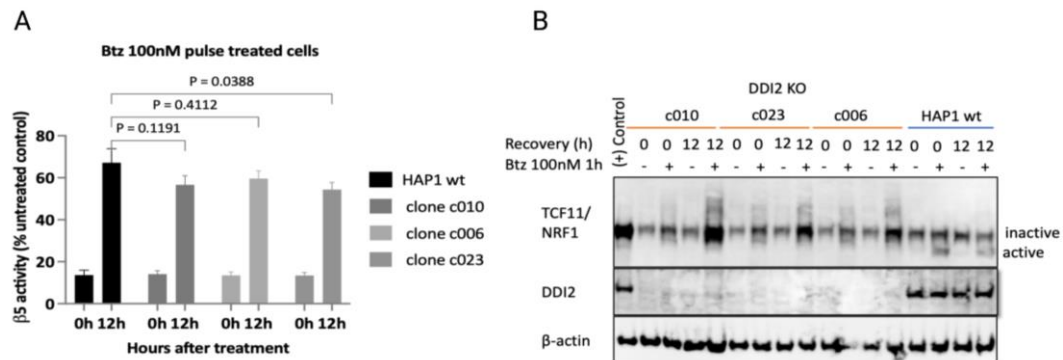
dissociation from the inhibitor (**Figure 2.1 D** vs. **Figure 2.1 F** and **G**). Most importantly, the activity recovery was the same in the wild-type cells and cells lacking DDI2 expression. Thus, after pulse treatment with proteasome inhibitors, DDI2 does not recover the proteasome activity in HAP1 cells (**Figure 2.1 D, F, and G**).



**Figure 2.1 Knockout of DDI2 does not impair the proteasome activity recovery.** **A.**) western blot showing basal expression of DDI2 in HAP1 wt and DDI2 KO (c010, c023, and c006) cell lines. **B.**) shows the experimental setting for **Figure 2.1 – Figure 2.3**. Viable cell (% of untreated control) was measured using CellTiter-Glo™ after the cells were pre-treated for one h with Btz or Cfz and graphed as **C.**) % of viable cells, 12h after Btz treatment, and **E.**) % of viable cells, 24h after Cfz treatment. Meanwhile, **D.**), **F.**), and **G.**) show the percentage of  $\beta 5$  activity assayed using Proteasome-Glo™ at 0, 12, and 24h. All **C.**) and **E.**) Dose – cell viability and **D.**), **F.**), and **G.**) Dose –  $\beta 5$  activity curves are generated by plotting the averages  $\pm$  S.E.M. of two to five biological replicates.

We confirmed these results using fluorogenic substrates Suc-LLVY-AMC. For this purpose, we pulse-treated the cells with Btz (100nM) and recovered the cells in a drug-free medium for 12h. We showed that recovery of the  $\beta 5$  activity occurred in the absence of DDI2 in cells (**Figure 2.2 A**).

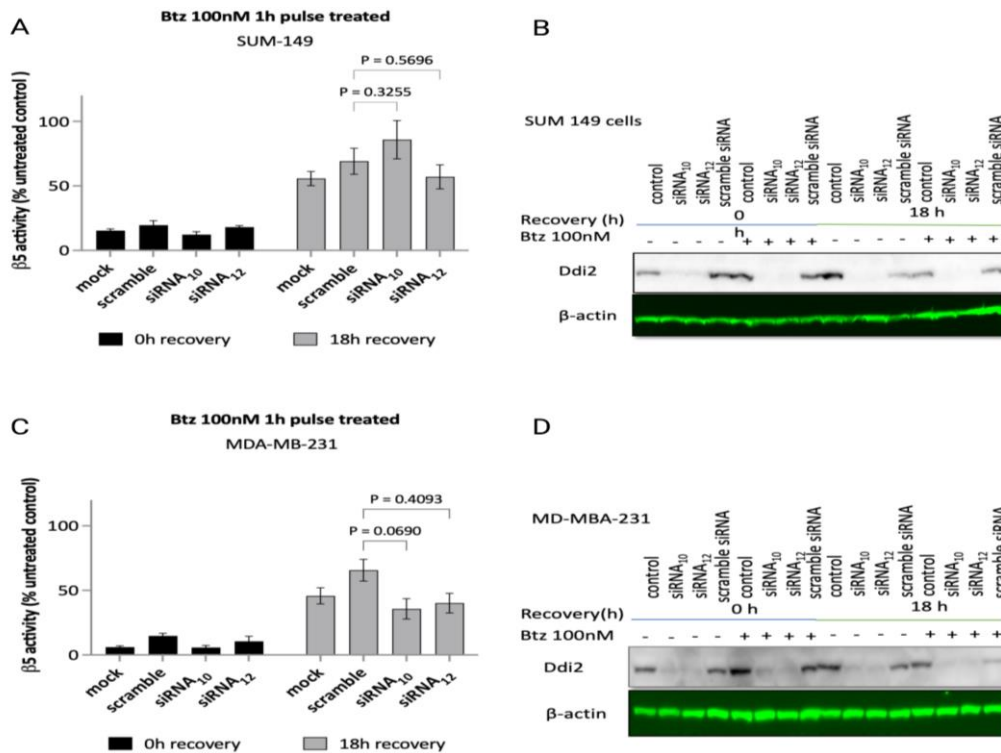
As described in **Chapter 1 Section 1.7.2**, cleavages by DDI2 activate transcription factor Nrf1, which upregulates the transcription of proteasome genes. We also found that Knockout of DDI2 eliminates the ability of cells to process Nrf1 transcription factors upon partial proteasome inhibition (**Figure 2.2 B**), as reported by previous studies, yet the proteasome activity is still recovered. We will further discuss this in **Section 2.3**



**Figure 2.2 Knockout of DDI2 does not affect the recovery activity even though DDI2 knockout cells cannot process Nrf1.** **A.)** shows the percentage of  $\beta 5$  activity in HAP1 wt and DDI2 KO (c010, c023, and c006) cells after being pre-treated with Btz (100nM) and assayed using fluorogenic substrates Suc-LLVY-AMC at 0 and 12h. A Two-way ANOVA multiple comparison statistical analysis was performed by comparing the average of nine biological replicates of each HAP1 wt vs. DDI2 KO cells at 0 and 12h time points. **B.)** the cell lysates in **A.)** were further fractionated using gel electrophoresis and probed with TCF11/NRF1 igG. Western blot showing DDI2 KO cells loses their ability to process and activate Nrf1. Nrf1 exists predominantly in its glycosylated form in DDI2 KO cells. A positive control is HAP1 cells pulse-treated with Btz followed by 11h of CB-5083 p97 inhibitors.

We then tested whether silencing DDI2 by siRNA affects the recovery of proteasome activity in Triple-negative breast cancer (TNBC) cells, SUM-149 and MDA-MB-231, which are sensitive to proteasome inhibitors<sup>12</sup>. After knocking

down DDI2 with two different DDI2 siRNA (siRNA<sub>10</sub> and siRNA<sub>12</sub>) for 72 hours, we pulse-treated the cells with Btz (100nM). We measured the  $\beta 5$  activity immediately after treatment and after 18h recovery. The 18h timepoint was chosen because these cells take 18h to reach optimum recovery of  $\beta 5$  activity. Again, we found that silencing DDI2 in both TNBC cells does not affect the cells' ability to restore their  $\beta 5$  activity (**Figure 2.3 A and C**). We confirmed that the DDI2 protein remained silenced During the experiments (**Figure 2.3 B and D**).



**Figure 2.3 Knockdown of DDI2 does not affect the recovery of the proteasome in inhibitor-treated cells.** A.) MDA-MB-231 and C.) SUM 149 cells were treated with DDI2 siRNA for 72h, followed by Btz (100nM) for 1 hr, then cultured in the drug-free medium for 18h. The  $\beta 5$  proteasome activity was measured using Suc-LLVY-AMC immediately (0h) and 18h after treatment. A Mixed-effect multiple comparison statistical analysis was performed by comparing the average of two to five biological replicates of each DDI2 siRNA vs. scramble (non-targeted pool siRNA) at 0 and 18h after treatment. DDI2 expression in extracts from experiments A.) and C.) were analyzed by western blotting. DDI2 expression in B) MDA-MB-231 and D) SUM 149 remains suppressed during the experiment.

In summary, our findings suggest that knockout or knockdown of DDI2 does not impair proteasome activity recovery in the inhibitor-treated cells.



### **2.2.2 The recovery of proteasome activity depends on the synthesis of new proteasome; however, the early onset of activity recovers before transcriptional upregulation of proteasome genes**

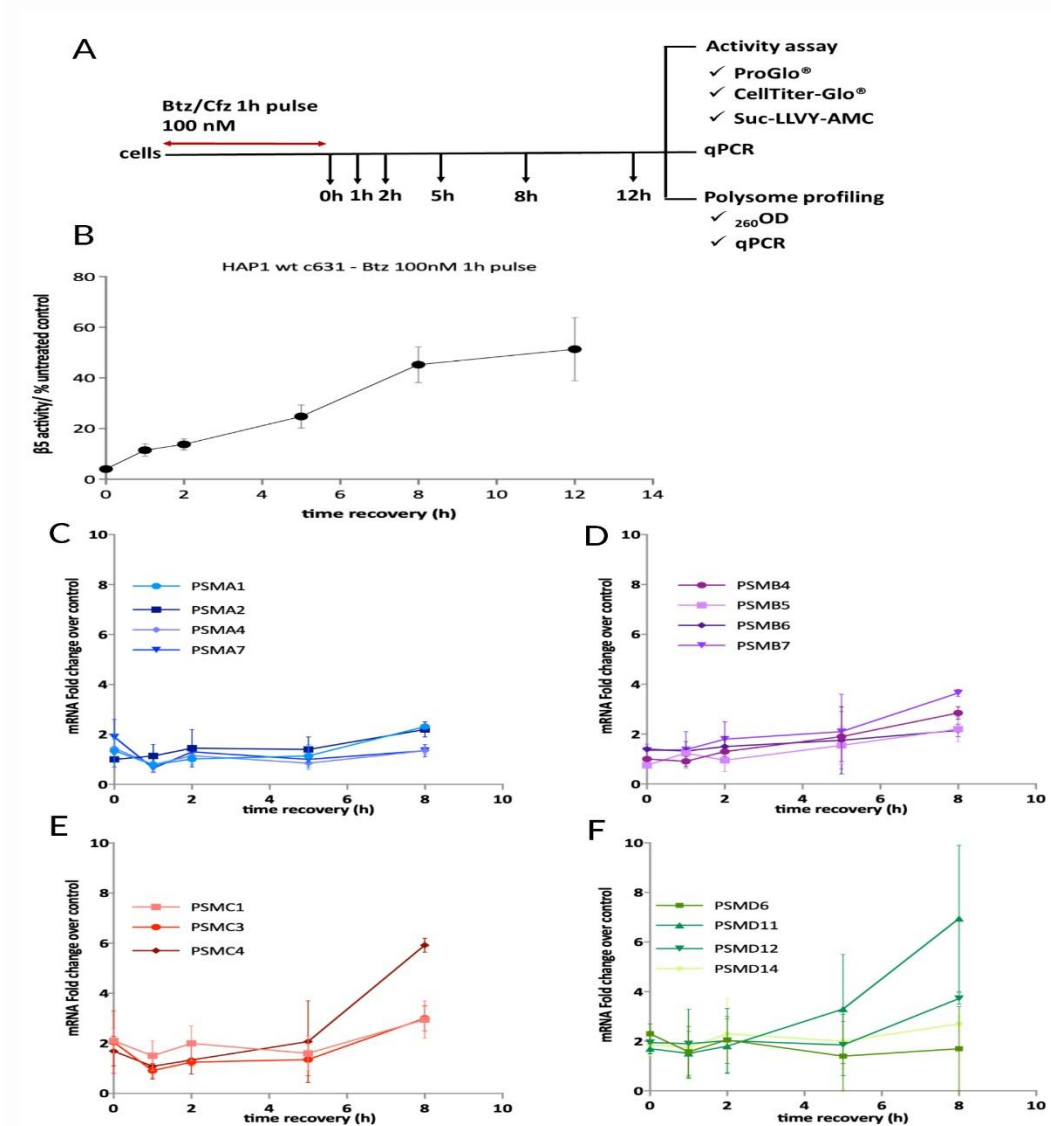
In the previous section, we showed that DDI2 does not play a role in proteasome activity recovery, and Nrf1 activation is DDI2 dependent. At the same time, some studies reported the activated Nrf1 further upregulates proteasome genes allowing the cells to recover their activity. This suggests that the proteasome activity recovery might be independent of DDI2 – Nrf1 pathway. To prove this notion, we must first show that the activity recovery is not dependent on the upregulation of proteasome genes.

In scenarios where the upregulation of proteasome genes plays a role in activity recovery, increased levels of proteasome subunits mRNA must occur prior to proteasome subunit synthesis and proteasome activity recovery. Hence, we wondered whether the upregulation of proteasome genes occurs before the activity recovery. For this purpose, we conducted a time-dependent experiment utilizing HAP1 wt that can process Nrf1 (**Figure 2.4 A**).

To this end, we subjected the cells to an hour of Btz pulse treatment. Then, we let them recover in a drug-free medium before testing them for  $\beta 5$  activity and used the Cell-titer Glo viability assay performed in parallel for normalization; thus, it will consider the potential slow growth of inhibitor-treated cells. In separate similar experiments, we isolated the RNA at each time point and performed qPCR for genes encoding proteasome subunits (**Figure 2.4 A**).

Surprisingly, we found sharp increases in activity recovery, mainly between 5 – 8h (**Figure 2.4 B**). At the same time, the qPCR data from the proteasome subunits also

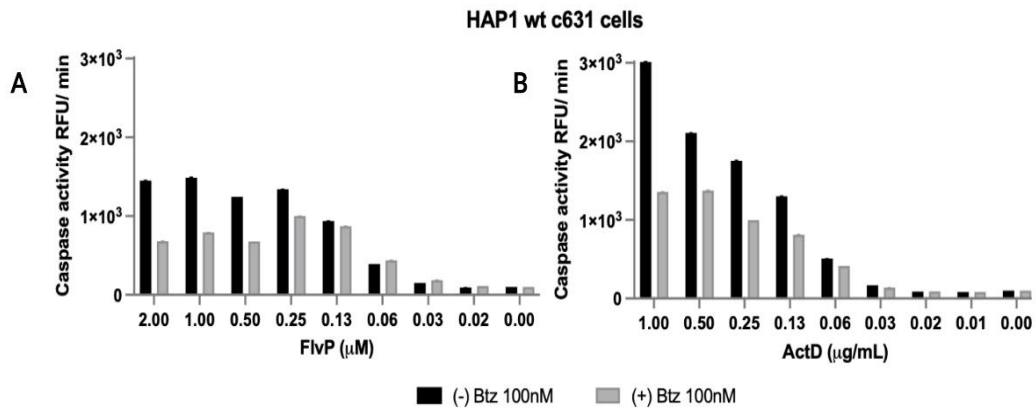
showed a similar pattern of increased mRNA production from 5 – 8h time points and decreased after 8 h (Figure 2.4 C – F). This tells us that the recovery of proteasome activity might precede the upregulation of proteasome subunits, and the recovery of activity might not be transcriptionally dependent.



**Figure 2.4** The recovery of proteasome activity preceded the induction of proteasome mRNA. HAP1 wt was pulse-treated with Btz (100 nM) and then cultured in their drug-free medium. Samples were collected at each time point after the end of the treatment as described in B.) and then assayed for proteasome  $\beta 5$  activity using ProGlo™ substrates as shown in A.). The percentage of  $\beta 5$  proteasome activity is obtained after normalization to cell viability at each time point (CellTiter-Glo™). Similar to the treatment in A.), RNA was isolated from the samples collected at 0 – 1 – 2 – 5 – and 8h. C.) to F.) exhibits a pattern of proteasome mRNA induction in inhibitor-treated cells. The  $\beta 5$  activity and mRNA proteasome induction were graphed in PRISM by plotting the averages  $\pm$  S.E.M. of two to five biological replicates. No statistical analyses were performed on these graphs.

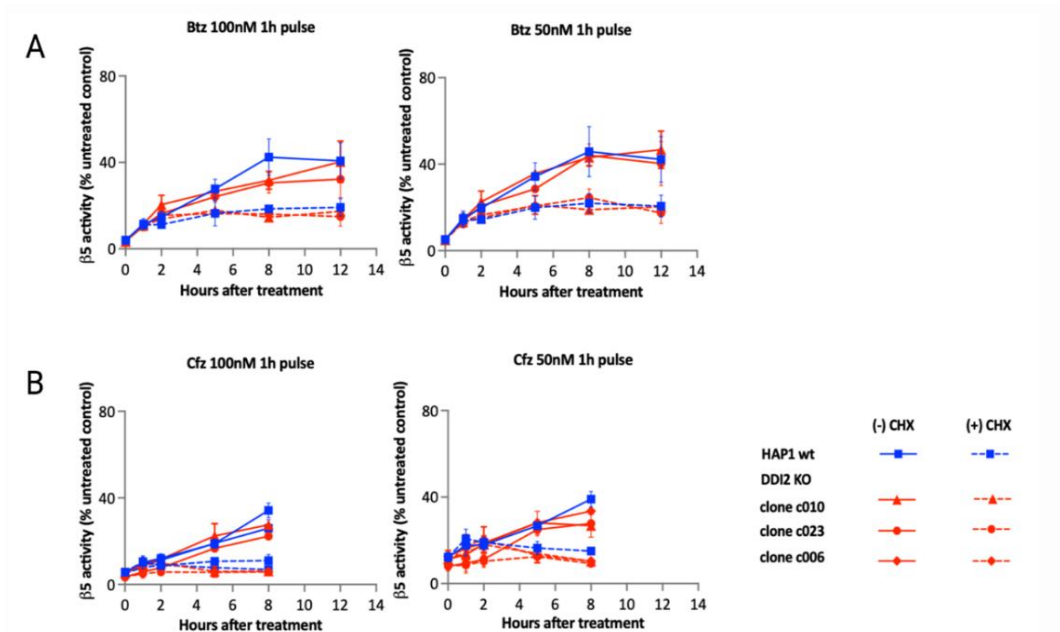
Additionally, to confirm whether the recovery of proteasome activity is transcriptionally dependent, transcription inhibitors can be used following treatment with a proteasome inhibitor. Here, we experimented using transcription inhibitors Flavopiridol (FLA) and Actinomycin D (ActD). As a brief description, FLA inhibits transcription by inhibiting CDK7 and CDK9, kinases in charge of Serine-phosphorylation of the CTD of RNA Polymerase II, which are responsible for transcription initiation and elongation<sup>104-105</sup>. Meanwhile, ActD causes RNA polymerase inhibition and decreases transcription<sup>106</sup>. This DNA intercalator binds to a DNA double helix and stops the activity of DNA-dependent RNA polymerase<sup>106</sup>.

However, some studies reported that transcription inhibitors induced cell apoptosis<sup>104, 107-108</sup>; hence, we first assessed the visibility of our experiment by pulse-treating the cells with Btz followed by ActD or FLA for 5h and measured caspase activity. The 5h treatment was chosen because the maximum upregulation of proteasome genes occurred between 5 – 8h. Unfortunately, our attempted experiments using FLA or ActD caused HAP1 cell apoptosis. Five hours of HAP1 wt cell treatment with these transcription inhibitors (without pretreatment with Btz) activates 3/7 caspase 4-5x higher compared to untreated cells at low doses >60 nM and >60 ng/mL, respectively (**Figure 2.5**). Pretreatment with Btz slightly reduces the apoptotic effect of FLA and ActD, as previously reported<sup>107, 109</sup>. Furthermore, studies reported that higher concentrations of FLA or ActD are needed to inhibit transcription factors globally<sup>105, 110</sup>. Thus, we cannot use this method to determine whether proteasome activity recovery is transcriptional-dependent.



**Figure 2.5 Treatment with transcription inhibitors Flavopiridol (FlvP) or Actinomycin D (ActD) activates caspase 3/7 activity and induces cell apoptosis HAP1 wt cells.** Caspase activity (RFU/min) in HAP1 wt cells treated for 5h with a serial dilution of A.) FlvP from 0.02 - 2  $\mu\text{M}$ , B) ActD from 0.01 - 1  $\mu\text{g/mL}$ . The experiment was only performed 1x.

Next, we further ask whether recovery of proteasome activity involves biosynthesis of new proteasomes. Here, we employed Cycloheximide (CHX) to inhibit protein synthesis by immobilizing ribosomes on RNA, thereby blocking translation elongation. To this end, we treated the cells with a proteasome inhibitor for one hour at two different concentrations (50 and 100 nM), followed by a medium containing 100  $\mu\text{g/mL}$  CHX or CHX-free medium at different time points. We then used these treated cells in luminescent-based assays to determine  $\beta 5$  activity (Proteasome-Glo<sup>TM</sup>) and cell viability (CellTiter-Glo<sup>TM</sup>).



**Figure 2.6** *The recovery of proteasome activity required the synthesis of new proteasomes.* HAP1 wt and DDI2 KO (c010, c023, and c006) cell lines were pulse-treated with 100nM or 50nM of A.) Btz or B.) Cfz and cultured in a drug-free medium (solid line) or CHX-containing medium (dashed line) at times indicated and assayed for  $\beta 5$  proteasome activity remaining using Proteasome-Glo™. The 0h time point indicates a point when Btz was removed. The percentage of  $\beta 5$  proteasome activity was obtained after normalization to cell viability at each time point (CellTiter-Glo™). All  $\beta 5$  activity graphs were generated via PRISM by plotting the averages  $\pm$  S.E.M of 2 - 5 biological replicates.

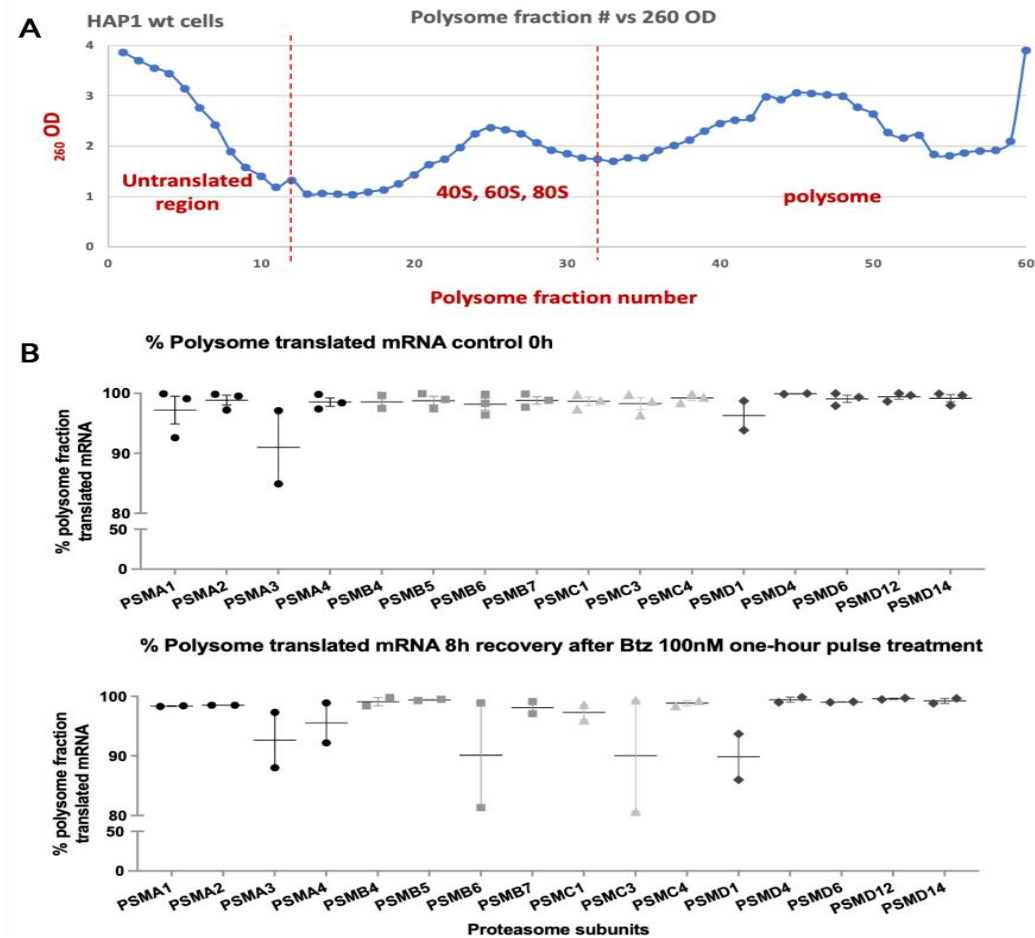
There are two main observations we can see in **Figure 2.6** above. Although we blocked the synthesis of the new proteasome in the CHX-treated cells, we can still see that the cells exhibited a sharp activity recovery from 0 – 5h after pulse treatment with Btz (**Figure 2.6 A**) or 0 – 2h after pulse treatment with Cfz (**Figure 2.6 B**). These results suggest that the early onset of proteasome recovery is independent of newly synthesized proteasomes. There are two possibilities to explain the above hand notion. First, the recovery may be facilitated by the availability of a non-inhibited proteasome. Second, the recovery is due to the rapid assembly of the remaining proteasome subunits within the cells. We will cover this further in **Section 2.3**.

More importantly, we show that when the synthesis of new proteasomes is blocked by CHX (dashed line), we cannot see a recovery of proteasome activity from 5 – 12h in Btz-treated cells and 2 – 8h in Cfz-treated cells. This is particularly evident in **Figure 2.6 B**, where the recovery of proteasome activity in Cfz-treated cells should only be possible through synthesizing new proteasomes<sup>5, 99</sup>. Thus, we can say that the late onset of proteasome activity recovery is due to the newly synthesized proteasome in inhibitor-treated cells.

Following our previous results, we further asked whether the efficiency of mRNA translation from the proteasome genes increases when the proteasome is inhibited. Therefore, we conducted a polysome profiling experiment to investigate this question using the previously described protocol<sup>111-114</sup>. Polysome profiling is a standard method for assessing the efficiency of global mRNA translational levels or subsets of mRNA, which are actively translated in cells<sup>113</sup>. It separates the actively translated polysome-bound mRNA from untranslated free mRNA using a sucrose density gradient fractionation<sup>113</sup>. The untranslated mRNAs remain on the top of the gradient. On the other hand, the mRNA bound to polyribosomes (polysomes) is heavier than mRNAs bound only to monoribosome (the 80S or monosomes); hence, the polysome will be in the higher sucrose density gradient (**Figure 2.7 A**). Then, the mRNA content of untranslated mRNA, the 80S, and the polysome bound mRNAs is detected by qPCR.

Overall, the qPCR data of proteasome genes in **Figure 2.7 B** demonstrate an efficient mRNA translation (>90%) of untreated controls and samples harvested 8h after

one-hour treatment with Btz (100nM). Complete data is included in the **Supporting Information Figure SI 2.1**.



**Figure 2.7** The translation efficiency of proteasome subunits mRNA. HAP1 wt cells were pulse-treated with Btz (100nM) and recovered in a drug-free medium at the time indicated. **A.**) shows an example of manual polysome fractions differentiation based on the  $OD_{260}$  in HAP1 wt cells. Fractions 1 – 12 contain the untranslated mRNA, fractions 13 – 32 contain mRNA bound to mono-ribosome (the 80S), while fraction > 30 contains the mRNA bound to polysome. The polysome fractions in **A.**) were combined and divided into untranslated and translated fractions, then the RNA was isolated from these two combined fractions. Subsequent qPCR of proteasome subunits was performed. **B.**) shows the percentage of translated mRNA of proteasome subunits of control 0h and 8h proteasome recovery (collected after treatment). The percentage of translated mRNA was calculated as a fraction of the total. **A.**) was generated in Microsoft Excel, while **B.**) was generated through PRISM by plotting the averages of 2 – 3 biological replicates.

In summary, our findings suggest that the late onset of proteasome activity recovery is indeed dependent on the synthesis of the new proteasome. The proteasome mRNA is efficiently translated during the activity recovery suggesting that the early onset of the activity recovery may not be transcriptional-dependent.

### 2.2.3 The recovery of proteasome activity is mainly due to increases in the 26S/30S activity

Earlier in **Figure 2.4 E and F**, we showed that proteasome genes encoding subunits of the 19S regulatory particles (PSMCs and PSMDs) are induced stronger than genes encoding the 20S proteolytic core (PSMAs and PSMBs). We also showed in **Figure 2.7** that proteasome activity recovery depends on synthesizing new proteasomes, except for very early recovery. Hence, to confirm the results in **Figure 2.7** and verify whether proteasome activity recovery is due to the increased activity of the 26S proteasome, we used Native gel to separate the 20S and 26S proteasomes. The native gel separation technique coupled with Suc-LLVY-AMC activity assay has been widely used to study the structure and composition of proteasome complexes which can separate the 30S, 26S, 20S proteasomes, and proteasome activators<sup>115</sup>. After the separation, we measured activity by in-gel Suc-LLVY-AMC activity assay. Suc-LLVY-AMC is a substrate for the  $\beta 5$  catalytic active site, yet it can also stimulate  $\beta 1$  activity<sup>116</sup>. Next, we used this gel for immunoblotting, probing the membrane with antibodies recognizing the 20S (PSMA or  $\alpha 6$ ) and 19S (PSMC6 or Rpt4) proteasome subunits. We expect to see 26S proteasome activity in the untreated cells. We also expect to see a disappearance of the 26S band immediately after 1h treatment and a gradual increase in 26S proteasome activity after treating HAP1 cells with a one-hour pulse of Cfz (100nM).

Confirming our results in **Figure 2.7**, we found an increase in the 26S proteasome activity starting at 5h (see the 26S band in-gel activity assay, **Figure 2.8 A**). However, we could not see an active 20S band during Suc-LLVY-AMC activity



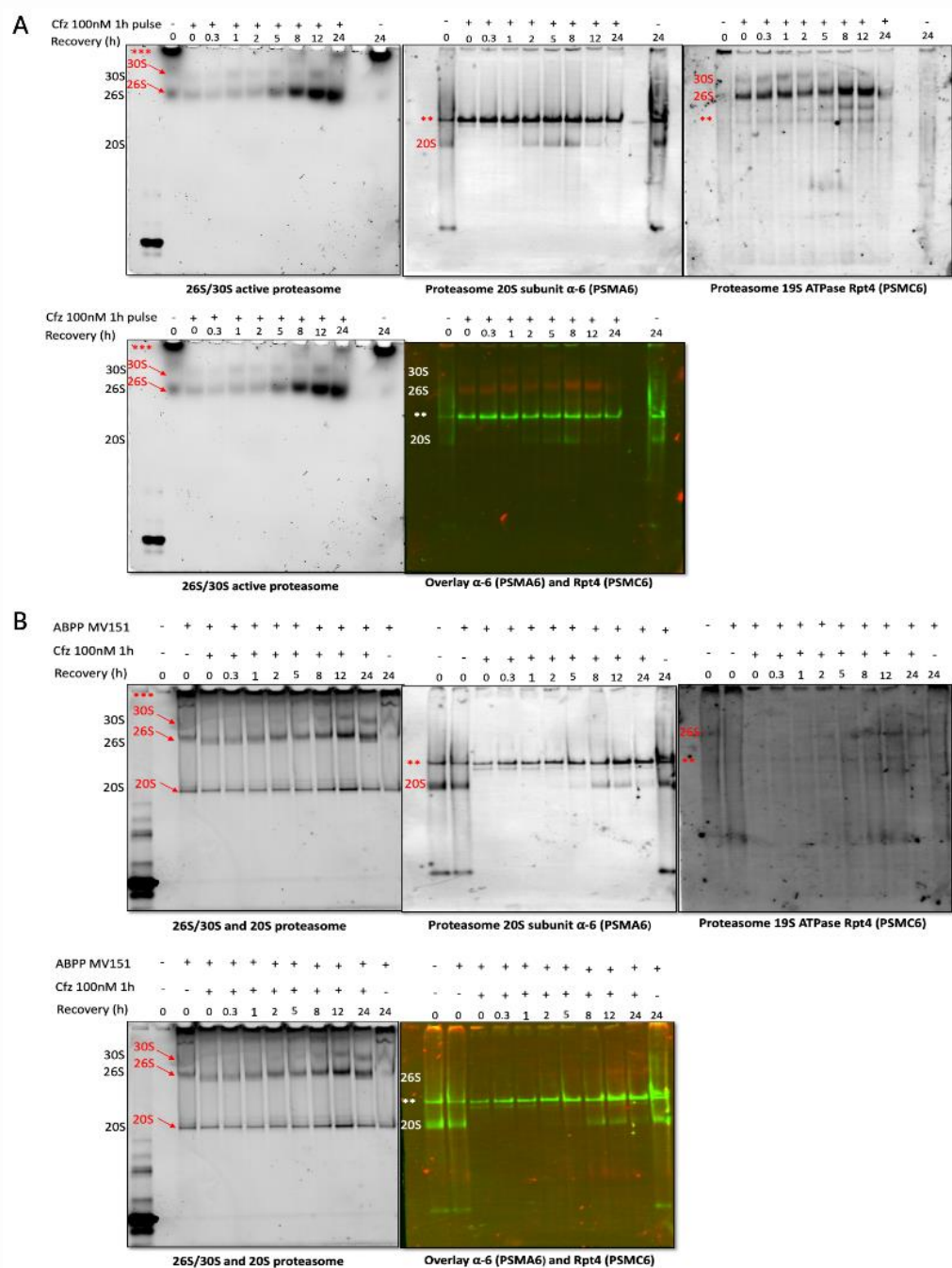
assay. This may be due to the 26S assay buffer containing KCl that can suppress 20S activity<sup>116</sup>.

Investigating using  $\alpha 6$  antibody showed that the slow- and fast-migrating 20S proteasomes were present in the untreated controls. A previous study reported that 20S peptidase activity was primarily due to the presence of a slowly migrating 20S proteasome<sup>115</sup>. However, the slow-migrating 20S proteasome seems to also be present in an unknown complex (**labeled as \*\***). Meanwhile, the fastly migrating 20S proteasome has minimal peptidase activity or exists mainly as an inactive form<sup>115</sup>. The active proteasome should have been inhibited after the treatment with Cfz and still be detected by antibodies. Nonetheless, we do not see any inactive 20S band at 0 – 1 h recovery, suggesting this inhibited 20S band may have degraded independently of the UPS (via autophagy-lysosome pathway)<sup>117</sup>. Furthermore, we also observed an inactive 20S proteasome formed at 2h and increased from 5h (see the 20S band in PSMA6/ $\alpha$ -6 immunoblot **Figure 2.8 A**). This observation is consistent with the literature that free 20S exists in the cells in the latent state<sup>115</sup>.

Moreover, probing with antibody for 19S ATPase subunit Rpt4, we show an increase in 19S protein expression, reconfirming our result in **Figure 2.4 E** and **F**. We also notice that a double-capped 26S proteasome (sometimes referred to as the 30S) is primarily present at the beginning of activity recovery from 0 – 2h, as shown in **Figure 2.8 A** (see the 30S band in PSMC6/Rpt4 immunoblot).

In addition, we observed an unknown band (**labeled as \*\***). These bands were visible when the blot was probed with  $\alpha 6$  or Rpt4 antibodies, indicating that they contain not only 20S but also 19S. These bands lay above the 20S, separated

between the single-capped 26S and 20S proteasome (see the \*\* bands in both PSMA6/ $\alpha$ 6 and PSMC6/Rpt4 immunoblots **Figure 2.8 A**). However, the exact composition of this unknown band is to be determined in the future.



**Figure 2.8 Increased 26S proteasome activity predominantly contributed to HAPI cell's proteasome activity recovery upon proteasome impairment.** HAPI wt cells were pulse-treated with Cfz (100nM) and recovered in the drug-free medium at times indicated. The cells extract was used for Native gel followed by in-gel activity assay and immunoblotting of proteasome subunits the 20S ( $\alpha$ -6) and 19S ATPase (Rpt4) as shown in A.). At the same time, samples in experiment A.) were treated with ABP MV151 for 30 min prior to native gel and immunoblotting of proteasome subunits the 20S ( $\alpha$ -6) and 19S ATPase as shown in B.).

Furthermore, although we probed the 26S proteasome with  $\alpha 6$  (mouse) and Rpt4 (rabbit) in the same blot at the same time, we noted that the 26S proteasome did not react with the  $\alpha 6$  antibodies but reacted strongly with the Rpt4 antibodies. Since the 26S proteasome comprises 19S–20S–19S, the 26S proteasome is supposed to respond with both antibodies. Again, an overlay of these two antibodies showed strong protein expression of 19S, yet undetectable 20S proteasome. Strong expression of 19S may be due to the addition of ATP to 5 mM in the sample's lysis buffer and 100  $\mu$ M in the Tris-Glycine buffer<sup>118</sup>, where ATP and  $Mg^{2+}$  are required for 26S proteasome activity<sup>119</sup>.

We also noted some activity on the top of the native gel (**labeled as \*\*\***) in control 0 and 24h, indicating some sample does not go inside the native gel. Native gel separates samples based on the shape and length of the protein, as well as the mass (size) -to-charge ratio<sup>118</sup>. There are several possibilities to explain this—first, changes in the structure and shape of the proteasome. Under normal conditions, when the proteasome is activated by 19S RP (PA700), the proteasome undergoes conformational changes to allow gate opening of the  $\alpha$ -ring of the 20S CP chamber<sup>30</sup>. In contrast, when the proteasome is inhibited, the proteasome structure is reshaped to allow the cell to adapt to impaired proteasome function, while it may differ for untreated controls. Second, the total surface charge may have changed when the structure is reshaped. This may explain why cells treated with inhibitors can migrate, while not the same for untreated controls. The pH of the Tris-Glycine running buffer can further influence the changes in the total surface charge of the native protein complexes<sup>118</sup>.

Moreover, we used cells from the abovementioned experiments and subjected proteasomes to affinity labeling ABP MV151 before the native gel to confirm the in-gel activity assay result in **Figure 2.8 A**. ABP MV151 is a BODIPY-TMR-containing probe with a reactive vinyl sulfone warhead<sup>120</sup>. This fluorescent probe can bind covalently with all proteasome-active catalytic sites<sup>120</sup>. In our case, ABP MV151 will label all remaining active proteasomes, including 26S and 20S. Here, we expect that we will see notable increases in 26 proteasomes labeled by ABP MV151, showing more 26S formation compared to untreated controls at 0 h. Likewise, we should also see the remaining active 20S proteasome marked with ABP MV151 and show an increase starting from 5h recovery.

As expected, we show an increase in 26S proteasome labeled with ABP MV151 during the proteasome activity recovery compared to untreated control 0h (see the 26S band ABP MV151, **Figure 2.8 B**). During the recovery of proteasome activity, we also see the uninhibited active 20S proteasome labeled by ABP MV151. For the record, ABP MV151 will label all active proteasomes, including  $\beta 1$  and  $\beta 2$  of 20S proteasome active sites. Immunoblotting with PSMA6/ $\alpha 6$  antibody in **Figure 2.8 B** shows the disappearance of inhibited proteasome from 0 – 1h recovery and an increase of inactive 20S proteasome starting 2h after Cfz-pulse treatment (see the 20S band), confirming the result in **Figure 2.8 A**.

Altogether, our findings suggest that the activity recovery is predominantly due to the increased activity of the 26S proteasome in the inhibitor-treated cells. The early onset of recovery of activity may not be due to the induction of mRNA proteasome but rather due to rapid assembly of remaining proteasome subunits.

### 2.3 Discussion

Current treatment using proteasome inhibitors has shown a pattern where the maximum inhibition lasts for a few hours, and the proteasome activity slowly recovers<sup>5, 11</sup>. This temporary proteasome inhibition (at sub-toxic doses) generates a ‘transient’ efficacy and may have contributed to unresponsiveness and resistance to proteasome inhibitors<sup>5, 11-13</sup>. Our study stems from the idea that inhibiting the recovery pathway of proteasome activity will help prolong the proteasome inhibition's duration and retain the anti-neoplastic activity of proteasome inhibitors. We were interested in targeting genes that regulate proteasome activity recovery. One of these genes is the ER-resident Nrf1 transcription factor which is responsible for the induction of proteasome genes after partial proteasome inhibition<sup>14</sup>. Nrf1 is processed by DDI2<sup>15</sup>. We hypothesized that blocking DDI2 would block the activation of Nrf1 and further block the proteasome recovery. However, when we began this work, it had not yet been demonstrated whether the Nrf1-mediated upregulation of proteasomes transcription results in the recovery of proteasome activity after a clinically relevant pulse treatment with proteasome inhibitors.

This study demonstrates that knockout or knockdown of DDI2 does not impair the recovery of proteasome activity (**Figure 2.1, Figure 2.3**). Likewise, recently, other studies found that the proteasome activity recovery is only 24 – 25% less in DDI2 KO cells than in DDI2 WT parental cells (MDA-MB-231 and Multiple Myeloma AMO-1 cell)<sup>38, 94</sup>. Furthermore, we also confirmed that Nrf1 activation is DDI2 dependent (**Figure 2.2**), and proteasome activity recovered in the absence of Nrf1 activation. This raises the question of whether other transcription factors are

involved or whether Nrf1 transcription is responsible at all. So far, our results suggest that recovery of proteasome activity may occur independently of the DDI2 – Nrf1 pathways arguing that we found no differences in the recovery of proteasome activity in DDI2 KO cells and wt HAP1 cells.

To understand how the Nrf1-independent proteasome recovery may occur, we first asked how the proteasome recovery is regulated when Nrf1 is processed. **Suppose the induction of proteasome genes plays a role in proteasome recovery. In that case, the upregulation of the mRNA of the proteasome subunit should occur ahead of the *de novo* synthesis of proteasome subunits and the activity recovery.** However, our time-dependent experiment that measured  $\beta 5$  activity in inhibitor-treated cells in parallel with measuring proteasome subunit mRNA levels showed that activity recovery preceded proteasome subunit mRNA upregulation. Sharp increases in proteasome activity recovery occurred between 5 – 8h, while the increased mRNA level of proteasome genes also happened from 5 – 8h (**Figure 2.4 A vs. Figure 2.4 C – F**). Although, we could not determine whether the proteasome recovery is not transcriptional-dependent because transcription inhibitors induced rapid apoptosis (**Figure 2.5**), as previously reported<sup>104, 107-108</sup>. Our data are consistent with a study by Welk *et al.* also observed a delayed Nrf1-dependent transcriptional regulation of proteasome genes upon catalytic inhibition of  $\beta 5$  active site with Btz<sup>121</sup>.

The recovery of activity before the induction of proteasome genes does not wholly rule out that activity in HAP1 recovers due to the production of the new proteasomes. Our results in **Figure 2.6** suggest that the early onset of proteasome

recovery is independent of newly synthesized proteasomes. There are two possibilities to explain the above hand notion. First, the recovery may be facilitated by the availability of a non-inhibited proteasome. Second, the recovery is due to the rapid assembly of the remaining proteasome subunits within the cells. Since we treated the cells at a concentration that gave >90% inhibition, we think the early onset of proteasome activity recovery is most likely due to the rapid assembly of the remaining proteasome subunits. Supporting this idea, Welk *et al.* also found rapid recruitment of proteasome activators immediately after Btz treatment and at the beginning of the recovery<sup>121</sup>. They also found that proteasome activators were present in untreated cells, albeit at low concentrations<sup>121</sup>, suggesting that not all nascent proteasome polypeptides were rapidly degraded when not assembled. These proteasome activator subunits were recruited to the 20S and 26S in a transcriptionally independent manner upon cellular inhibition of the catalytic  $\beta 5$  site<sup>121</sup>. In our case, it may be proteasome activators and 26S proteasome nascent polypeptides. Moreover, kinetic analysis of the stability of proteasome subunit proteins after treatment with proteasome inhibitor in mammalian cells showed a non-exponential Degradation (NED) profile of these proteins where the protein subunits were stabilized after passing their average  $t_{1/2}$  (~4h)<sup>122</sup>. Although we did not conduct experiments to prove that the early onset of proteasome activity recovery was due to the rapid assembly of nascent polypeptide proteasome, the above two studies strengthen the notion that the stability of unassembled nascent polypeptide proteasome subunits may also play a role in the recovery of proteasome activity<sup>121-122</sup>. While MM cells display rapid degradation of nascent polypeptides<sup>123</sup>, it is not always the same as other cancers<sup>121-122</sup>.



We also found that the newly synthesized proteasome activity increases between 5 – 8h. It means the late onset of proteasome activity recovery in inhibitor-treated cells is due to the newly synthesized proteasome. This is particularly evident in **Figure 2.6 B**, where the recovery of proteasome activity in Cfz-treated cells should only be possible through synthesizing new proteasomes<sup>5, 99</sup>. Several other papers also reported the increased synthesis of all proteasome subunits and new proteasome formation due to the induction of proteasome genes, a classic response to impaired proteasome function<sup>14, 86, 99</sup>. Since activity recovery depends on the synthesis of new proteasomes, we conducted polysome profiling to see the possibility of increased translation efficiency of proteasome mRNA when the proteasome is inhibited. Our results suggest an efficient mRNA translation (>90%) of all proteasome subunits at 1 – 8h recovery of proteasome activity comparable to the untreated control at 0h (**Figure 2.7** and **Figure SI 2.1**). We did not see the activation of proteasome mRNA translation.

All of the results from **Figure 2.4** – **Figure 2.7** has led us to the idea that the proteasome activity recovery in HAP1 cells depends on synthesizing the new proteasome. Yet, the early onset of activity recovery is transcriptionally independent. This further left us with the notion that the activity recovery might begin by rapidly assembling preexisting proteasome subunits or forming alternative proteasome complexes. Both of these occur independent of the induction of proteasome genes. As mentioned above, nascent polypeptides do not always degrade rapidly, including proteasome subunits nascent polypeptides in mammalian cells<sup>122</sup>. Through global pulse-chase experiments, McShane *et al.* have identified ~10% non-exponentially degraded protein subunits (NED) compared to 49%

exponentially degraded protein (Ed) counterpart<sup>122</sup>. These NED protein subunits were part of multiprotein complexes (~70%) and were found to be excessively synthesized<sup>122</sup>. They also found that the NED protein subunits assemble earlier than the ED proteins<sup>122</sup>. Normal proteostasis is disrupted when the proteasome is inhibited, and misfolded proteins accumulate. Cells will try to repair their normal function to survive and avoid apoptosis. Therefore, it is logical that cells would use their energy (ATP) to quickly assemble proteasome subunits that are already available rather than synthesizing new proteasomes.

Additionally, we previously observed in **Figure 2.4** that the induction of proteasome genes mainly occurs between 5 – 8h, and the 19S subunits show stronger increases than the 20S subunits. We also observed the activity recovery rises from 5 – 8h in **Figure 2.4 – Figure 2.6**. We confirmed these findings through in-gel activity assay, ABP MV151, and activity assay from sucrose fractions (**Figure 2.7 – Figure 2.8**). Indeed, we found that the 26S proteasome activity predominantly increases after 5h. We also show increases in the protein level of 19S proteasome subunits, suggesting that 19S mainly contributed to the 26S proteasome activity. Although we could not show active 20S proteasome during the activity assay, we could show active 20S proteasome labeled by ABP MV151. In both activity assay and ABP MV151, we showed that inactive 20S proteasomes started to form at 2h and increased from 5h, while the inhibited proteasomes from 0 – 1h possibly degraded independently of the UPS system. Supporting our results, a pharmacokinetic study of Btz following a one-hour bolus IV administration showed the recovery of proteasome activity also increased from 4h and returned to basal level in 72 – 96h<sup>47</sup>.

To summarize, we initially hypothesized that inhibiting the DDI2 protein would impair Nrf1 processing, thereby blocking the induction of proteasome genes and the recovery of proteasome activity. However, our results in **Figure 2.1– Figure 2.3** indicate that the proteasome activity recovery occurred independent of the DDI2 – Nrf1 pathway. We support this finding through our results in **Figure 2.4 – Figure 2.7**, showing that the late onset of proteasome activity recovery predominantly occurred through increased activation of the 26S proteasome. At the same time, the early start of proteasome activity recovery is independent of the induction of proteasome genes. The practical implication of our finding is that the proteasome activity recovery depends on the cell's physiology and can not be generalized to all cancer cells. We will discuss this further in **Chapter 5**.

## 2.4 Materials and Methods

### 2.4.1 Chemicals (inhibitors, substrates, and other reagents)

Unless stated otherwise, we purchased all additional lysis buffer or assay buffer preparation reagents from VWR Life Sciences.

Bortezomib (Btz) and Carfilzomib (Cfz) were purchased from LC Laboratories. CB-5083 was purchased from Cayman Chemical. Cycloheximide was purchased from Sigma Aldrich. RNAsin Plus™ and PhosSTOP™ were obtained from Promega and Roche, respectively. PhosSTOP™ consisted of Sodium molybdate dihydrate ( $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ ), Sodium vanadium oxide ( $\text{Na}_3\text{VO}_4$ ), Cantharidin ( $\text{C}_{10}\text{H}_{12}\text{O}_4$ ), and Polyethylene glycol (PEG). PhosSTOP™ 1x (v/v) was prepared by dissolving one tablet PhosSTOP™ in ddH<sub>2</sub>O. Flavopiridol was obtained from Selleckchem, while Actinomycin D was obtained from MP Biomedicals. Digitonin was obtained from GoldBio, while CHAPS (3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate) was obtained from ThermoFisher.

The fluorogenic substrate Suc-LLVY-7-amido-4-methylcoumarine (Suc-LLVY-AMC) was obtained from Bachem, Suc-LLVY-Aminoluciferin was included within the ProteasomeGlo™ Cell-Based Assay kit, which was obtained from Promega. The caspase 3/7 fluorogenic substrate (ZDEVD)<sub>2</sub>-R110 and Ac-DEVD-AMC were obtained from Promega and Bachem, respectively.

Activity-based probe (ABP) Bodipy TMR-Ahx3L3VS (MV151), a peptide-based BODIPY-TMR fluorophore probe for labeling all the catalytic active-site of

proteasome subunits, was provided by the Overkleeft laboratory (University of Leiden)<sup>120</sup>.

#### 2.4.2 Cell lines and cell culture

human near-Haploid HAP1 (wt clone 631, DDI2 KO clones 006, 023, and 010 were obtained from Horizon Discovery<sup>TM</sup>, although c006 was initially purchased and provided to us by the Goldberg laboratory (Harvard Medical School). All clones were cultured in Corning® Iscove's medium supplemented with 10% Gibco<sup>TM</sup> Fetal Bovine Serum (FBS) and 100 µg/mL Corning® Penicillin-streptomycin at 37°C in the presence of 5% CO<sub>2</sub>.

**Table 2.1 DDI2 CRISPR KO clones and their gRNA sequences.**

<b>DDI2 KO clones</b>	<b>gRNA sequence</b>
HZGHC000182c <b>006</b> (4bp deletion)	GCTCGAAGTCGGCGTCGACC
HZGHC000182c <b>023</b> (1bp insertion)	GCTCGAAGTCGGCGTCGACC
HZGHC000396c <b>010</b> (41bp deletion)	AATAGCTATGGAAGAGGCTC

MDA-MB-231 cells were purchased from ATCC® (Cat. #HTB-26<sup>TM</sup>) and cultured in Corning® Dulbecco's Modified Eagle's Medium (DMEM)/Hams F-12 50/50 Mix supplemented with FBS 5%, at 37°C with 5% CO<sub>2</sub>. SUM149 cells were obtained from Asterand and cultured in Corning® Dulbecco's Modified Eagle's Medium (DMEM)/Hams F-12 50/50 Mix (supplemented with 5% FBS, 100 µg/mL Penicillin-streptomycin insulin, Glutamine, HEPES, Hydrocortisone) at 37°C with 5% CO<sub>2</sub>. All cells were validated by STR DNA fingerprinting using Promega<sup>TM</sup> GenePrint 10 System (Promega Cat. #B9510) at the Vermont Cancer Center DNA Analysis Facility<sup>12</sup>.

### 2.4.3 siRNA experiments

Two human DDI2 siRNA were chosen from On Target plus a set of four DDI2 siRNA (Dharmacon Horizon Discovery®) based on the efficiency of the knockdown. The target sequences are GGACAUGC UUAACGGCAC (siRNA<sub>10</sub>, Cat. #J-032713-10-0050) and CAAGAAAGGAUUCGUCUGU (siRNA<sub>12</sub>, Cat. #J-032713-12-0050).

Two human NFE2L1 siRNA were chosen from On Target plus a set of four Nrf1 siRNA (Dharmacon Horizon Discovery®) in a similar fashion as above. The target sequences are GGGAUUCGGUGAAGAUUUG (siRNA<sub>8</sub>, Cat. #J-019733-08-20) and CCAACACUCCCAUCAUCA (siRNA<sub>9</sub>, Cat. #J-019733-09-20).

The Non-Targeting Pool siRNA #D-00110-10-20 containing target sequences UGGUUUACAUGUCGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGUUUUCUGA, and UGGUUUACAUGUUUCCUA was used as a negative control.

We optimized the siRNA transfection methods on MDA-MB-231 and SUM 149 cells according to Dharmacon™ DharmaFECT™ 1–4 transfection protocol (<https://horizondiscovery.com/-/media/Files/Horizon/resources/Quick-protocols/dharmafect-rna-transfection-quick-protocol.pdf>) and DharmaFECT™ Transfection Reagents - siRNA transfection protocol (<https://horizondiscovery.com/-/media/Files/Horizon/resources/Protocols/basic-dharmafect-protocol.pdf>).

MDA-MB-231 or SUM 149 cells were seeded in 6-well plates overnight at  $2 \times 10^5$  cells/ well. The next day, the cells were transfected with siRNA at a final concentration of 25nM and DharmaFECT 1 Cat. #T-2001-03 Transfection reagent at a final concentration of 0.3% in Gibco™ Opti-MEM 1X Reduced Serum Medium and Corning® DMEM/Hams F-12 50/50 Mix without antibiotic. Then, 72h after transfection, the cells were pulse treated with Btz or Cfz at a concentration of 50 – 100 nM and collected for time points 0 and 18h. At the same time, the transfected cells were cultured in a drug-free medium (with FBS, without antibiotics) for the control.

#### **2.4.4 Cell lysis and protein quantification**

The inhibitor- and mock-treated cells were washed and harvested with PBS, collected by centrifugation, and then frozen for immunoblotting and activity assays. The frozen cell pellets were lysed either in a whole-cell lysis buffer or in a modified homogenized lysis buffer. The whole-cell lysis buffer contains 0.5% CHAPS, 10% glycerol, 5 mM  $MgCl_2$ , 1 mM EDTA, 1 mM DTT, 100 $\mu$ M ATP, 10% PhosSTOP™ 1x (v/v), and 50 mM Tris-HCl to pH 7.5. Meanwhile, the Modified homogenized lysis buffer contains 25% sucrose, 2 mM EDTA, 100  $\mu$ M ATP, 1 mM DTT, 0.05% digitonin, 10% PhosSTOP™ 1x (v/v), and 50 mM Tris-HCl to pH 7.5. The cells were incubated for 15 min on ice, centrifuged at  $20,000 \times g$  for 15 – 20 min, and the supernatants were used for experiments.

For Native gel, the cell pellets were lysed in a buffer similar to whole cell lysis buffer except that DTT was at 0.5 mM and ATP was at 5 mM as described<sup>118</sup>.

The protein concentration of the cells extract was determined using Pierce™ Coomassie Plus (Bradford) Assay reagent (Cat. #23238) with bovine serum albumin as a standard. The absorbance reading was done at 595 nm in triplicates using Biotek Gen5™ Cytation. For activity assay and native gels, cell pellets were lysed and quantified on the day of the experiments.

#### **2.4.5 Activity assays**

##### **Proteasome recovery activity assays**

The proteasome activity in inhibitor-treated cells was determined by luminescent Proteasome-Glo™ (ProGlo) assay or by using Suc-LLVY-AMC fluorogenic substrate as previously described by Britton *et al.*, 2009<sup>124</sup>. For the luminescent assay (**Figure 2.1**), the cells in 96-well plates were treated with inhibitors for 1h, then cultured in a drug-free medium for times indicated, washed with PBS, and lysed by one cycle of freezing and thawing in 25µL of cold PBS containing Digitonin 0.05%. ProteasomeGlo™ assay buffer (25µL) containing Suc-LLVY-aminoluciferin was added to the lysed cell at RT. The mixture was thoroughly mixed and preincubated on a shaker for ~10 minutes at RT. The remaining proteasome activity was monitored by the release of aminoluciferin following the cleavage of Suc-LLVY-aminoluciferin using Biotek Gen5™ Cytation with luminescence reading. The luminescence reading value of each sample was then normalized to the luminescence reading of the blank value. The proteasome activity was calculated as a percentage of the luminescence reading of the treated sample over the untreated sample (control). Three technical replicates were analyzed for each



sample in which the CellTiter-Glo™ (CTG) Luminescent Cell Viability Assay (see below) was used for normalization.

For the experiment in **Figure 2.6**, the cells in 96-well plates were pulse treated with proteasome inhibitors for one-hour at two different concentrations (50 and 100 nM), followed by a medium containing 100 µg/mL CHX or CHX-free medium at different time points. Subsequently, half of these plates were used in parallel assays for determining β5 activity (ProGlo™) and cell viability (CTG™).

To determine the proteasome activity in cell extracts, 1 µg of cell lysates was spiked into 100 µL of 100 µM of Suc-LLVY-AMC in 26S assay buffer (50 mM Tris HCl pH 7.5, 40 mM KCl, 2 mM EDTA, supplemented with a freshly added with 1 mM DTT, 100 µM ATP), which then preincubated at 37°C for 10 minutes. KCl is used to suppress the activity of the 20S proteasome<sup>116</sup>. An increase in AMC-fluorescence was monitored continuously at 37°C on the Biotek Gen5™ Cytation instrument excitations wavelength of 380 nm and emission of 460 nm. The slope curves of the reaction progress were then determined. Inhibition in the inhibitor-treated samples was calculated as a percentage of the slope of mock-treated controls. Each sample was conducted in three technical replicates.

In addition to the in-gel activity assay, the occupancy of the proteasome 26S and 20S active sites was evaluated using ABP MV151<sup>120</sup> prior to the Native gel run. To this end, the cell lysates were treated with 1 µM of ABP MV-151 for 30 min at 37°C and followed by native gel electrophoresis as described above. Then, the gel was visualized using RGB Cy3 (green channel) c600 azure™ BioSystems imager.

### **Viability assays**

The cell viability was determined using Promega CellTiter-Glo™ Luminescent Cell Viability Assay based on existing ATP measurements. Similar to the Proteasome-Glo™ assay, the live cells in 96-well plates were harvested and washed with PBS, then lysed in 25µL of cold PBS containing Digitonin 0.05%. The samples were resuspended to ensure they were thoroughly lysed and pipetted to assay plates, followed by incubation on the shaker for ~10 – 15 min at RT (without freezing the cells). Then, 25µL of CellTiter-Glo™ assay buffer containing Ultra-Glo™ Recombinant Luciferase was added to the lysed cells, followed by 15 min incubation at RT. The bioluminescence reading was directly proportional to the ATP or corresponded to the viable cells, and it was monitored using Biotek Gen5™ Cytation with luminescence readings (550 – 570nm).

### **Caspase 3/7 assay**

To measure the activity of caspase-3/7 in living cells or the number of cells that undergoes apoptotic in 96 wells plates, a fluorogenic substrate (ZDEVD)2-R110 was employed. To this end, live cells in 96-well plates were harvested, supplemented with 50µL of PBS containing Chaps 0.5%, and preincubated on the shaker at RT for 1 – 1.5h. Afterward, a 50µL of caspase assay buffer (containing 50 mM HEPES, DTT 10 mM, EDTA 4 mM, Chaps 0.1%, and 20µM of (ZDEVD)2-R110 substrate) was added to the lysed cells. Then, the 96-well plates were re-incubated in RT on the shaker for 1.5 hours before fluorescence readings. Cleavage of (ZDEVD)2-R110 by caspases 3/7 produces strong R110 fluorescent that can be

monitored using Biotek Gen5™ Cytation at excitation of 490 nm, emission of 525 nm.

#### **2.4.6 Native Gel and immunoblotting**

##### **Native gel**

The native gel separation was carried out as described by Roelofs *et al.* with slight modifications (see subheading Native PAGE)<sup>118</sup>. During the optimization of the procedure, Tris-Glycine provides better separation of proteasome complexes and sharp band resolution than Tris-Borate.

The native gel was performed using 1 – 1.5 mm thick precast Tris-Acetate gels (Invitrogen™) and 1X cold Tris-Glycine running buffer supplemented with 100µM ATP. The gel electrophoresis was conducted at 100 Volt for 6 – 7h at 4°C. The samples were mixed with one-third of the 5X native loading buffer containing 250 mM Tris-HCl (pH 7.5), 50% glycerol (v/v), and 0.03% xylene cyanol loading dye before loading into a native gel as described<sup>118</sup>.

The in-gel Suc-LLVY-AMC activity assay was performed immediately after the gel run 1 to determine the fractions of 26S and 20S during the recovery of proteasome activity. To this end, the gel was incubated with 100µM Suc-LLVY-AMC in 26S buffer containing 50 mM Tris HCl pH 7.5, 40 mM KCl, 2 mM EDTA, supplemented with 100 µM ATP, and 1 mM DTT for ± 20 – 30 min at 37°C. Then, the gel was imaged using c600 azure™ BioSystems under UV 365 nm exposure for 1 minute. The incubation was repeated with 0.02% SDS added to the substrate to measure

20S proteasome activity. The native gel was subsequently transferred to the membrane and probed with  $\alpha$ -6 and Rpt4 antibodies.

### Western blotting

The gel electrophoresis was conducted using NuPAGE™ Bis-Tris 8% Midi Gel (Invitrogen™, Cat. #WG1003BOX) at 150V, GenScript SurePAGE™ Bis-Tris 8% mini gel (Cat. #M00662) at 200V, and GenScript MES SDS running buffer (1X, containing 0.5M MES, ddH<sub>2</sub>O to pH 6) Cat. #M00677. The protein was transferred on Immobilon-pSQ PVDF membrane 0.2 $\mu$ M pore-diameter (Cat. #ISEQ00010) using Invitrogen™ Power Blotter 1-Step™ Transfer Buffer (1X) Cat. #PB7300. The membrane was blocked with 5% Milk in TBSt.

**Table 2.2 List of primary and secondary antibodies.**

<b>IgG</b>	<b>Type</b>	<b>Host</b>	<b>Dilution</b>	<b>Company</b>	<b>Catalog #</b>
<b>Primary</b>					
TCF11/NRF1 D5B10	mAb	rabbit	1:500	Cell Signaling	8052S
GAPDH D4C6R	mAb	mouse	1:1000	Cell Signaling	97166S
Beta-actin 8H10D10	mAb	mouse	1:1000	Cell Signaling	3700S
DDI2	pAb	rabbit	1:5000	Bethyl@Laboratories	A304-629A
$\alpha$ -6		mouse	1:1000		<i>*see notes</i>
Rpt4		rabbit	1:1000		
<b>Secondary</b>					
Anti-rabbit HRP-linked		Goat	1:1000	Cell Signaling	7074S
Anti-mouse HRP-linked		Goat	1:1000	Cell Signaling	7076P2
Alexa Fluor anti-rabbit 647		Goat	1:3500	Invitrogen™	A32733
Alexa Fluor anti-rabbit 680		Goat	1:3500	Invitrogen™	A20176
IRDye® 800CW anti-Mouse		Goat	1:3500	Li-COR Biosciences	926-32210

**Notes:**  $\alpha$ 6 antibody was purified from the supernatant of Mab2-17 hybridoma cell lines, which was kindly provided by Dr. Keiji Tanaka<sup>125</sup>.

#### 2.4.7 RNA isolation and qPCR experiment

The mRNA was isolated from cells using TRIzol™ Reagent (ThermoFisher Scientific Cat. #15596018) according to the manufacturer's protocol ([https://tools.thermofisher.com/content/sfs/manuals/trizol\\_reagent.pdf](https://tools.thermofisher.com/content/sfs/manuals/trizol_reagent.pdf)). Then, cDNA synthesis was performed using Applied Biosystems™ High-Capacity cDNA Reverse Transcription kit cat. #4368814 ([https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms\\_042557.pdf](https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_042557.pdf)), p. 6 – 8). RNA and cDNA were quantified using NanoDrop2000 (Thermo Scientific™) before the qPCR run. Real-time qPCR was performed using 2x SYBR Green Bimake™ qPCR Master Mix on a Bio-Rad C1000 thermal cycler CFX96™ Real-Time System. The qPCR forward and Reverse primers were listed in **Supporting Information Table SI 2.1**.

#### 2.4.8 Polysome profiling

We performed polysome profiling experiments as described elsewhere<sup>111</sup> to determine the efficiency of mRNA translation. We separated the untranslated mRNA from the 80S monosome and polysome. Here, we calculated the percentage of the translated mRNA fraction (80S monosome, polysome) and compared it to the untranslated mRNA fraction.

The 65% (w/v) sucrose stock solution was prepared in an autoclaved ddH<sub>2</sub>O. The serial dilution of sucrose from 5 – 55% was prepared by dilution of 65% sucrose stock solution in DEPC H<sub>2</sub>O and the 10X sucrose gradient buffer containing 200 mM HEPES-KOH, 1 M KCl, 50 mM MgCl<sub>2</sub>, 100 µg/mL cycloheximide, 1x complete mini protease inhibitor EDTA free, and 100 units/mL RNasin®. The sucrose concentration in each gradient fraction was determined using AicoHse 0 –

80% Brix Meter Refractometer, thus ensuring the gradient sucrose remained consistent throughout all the polysome profiling experiments.

To prepare the sucrose gradient, we pre-chilled the Beckman-Coulter 12mL tubes inside an ice chamber filled with dry ice. The latter was carried out to prevent the diffusion of sucrose and the formation of air bubbles while making sucrose gradients. Subsequently, the sucrose (1 mL) was spiked through the walls of the tubes with the following sequences 55% - 50% - 45% - 40% - 35% - 30% - 25% - 20% - 15% - 10% - and 5%. Each subsequent solution was layered on top of the previous and frozen at -80°C between each layer. Thus, the highest sucrose concentration was at the bottom of the tubes, while the lowest was at the top. The tubes were sealed with parafilm and kept at -80°C. Tubes were incubated overnight at 4°C prior to the polysome profiling experiment to create the gradient.

To isolate the polysome, cells (three 15 cm plates/gradient) were grown to 80% confluency. The overnight cultured cells were pulse-treated with Btz (100 nM) and then allowed to recover in their drug-free medium for times indicated. Then, at each time point, the cells were treated with 100 µg/mL CHX for 5 minutes prior to harvesting to preserve polysomes, which were present at all subsequent steps.

The cells were collected and lysed according to Morita *et al.*<sup>111</sup>. The cell pellets were washed and harvested in ice-cold PBS containing 100 µg/mL CHX before collecting the cells and then resuspended in 500µL hypotonic buffer and briefly vortexed. The hypotonic buffer contains 5 mM Tris-HCl pH 7.5, 2.5 mM MgCl<sub>2</sub>, 1.5 mM KCl, 1x complete mini protease inhibitor EDTA free, 100 µg/mL CHX, 1 mM DTT, and 100 units RNAsin. We then added Triton X-100 and Sodium

deoxycholate to a final concentration of 0.5% to the suspended cells, briefly vortexed, and centrifuged at  $20,000 \times g$  for 15 – 20 min  $4^{\circ}\text{C}$ . The supernatant was transferred to prechilled tubes, and the concentration was determined using ThermoScientific™ NanoDrop2000.

Next, approximately 80% of each sample lysate (depending on its concentration or OD) was slowly loaded on top of the sucrose gradient before placing them into the Beckman SW40Ti rotor. The sucrose gradient was centrifuged at 35,000 rpm for 2h.30 minutes at  $4^{\circ}\text{C}$  using an Ultra-high centrifuge Beckman-Coulter Optima XPN-100 with a deceleration brake at 1 or zero. After centrifugation, the sucrose was manually fractionated into clear 96-well UV plates (200 $\mu\text{L}$  per fraction or  $\pm 58$  fractions in total). We first checked each fraction's  $\text{OD}_{260}$  and  $\beta 5$  activity (without SDS).

The  $\text{OD}_{260}$  of each fraction was determined using Biotek Gen5™ Cytation. Then 5 $\mu\text{L}$  of each fraction was used for the Suc-LLV-AMC proteasome activity assay. At the same time, the remaining fractions were pooled and divided into three parts: the untranslated/ free RNA, translated/ 80S-bound ribosome (including the 40S and 60S), and actively translated/ polysome-bound. We then collected them for RNA isolation, as described above. Here, we were only interested in the percentage of translated mRNA (fraction 13 – 60) compared to this experiment's untranslated mRNA (fraction 1 – 12), as described in **Figure 2.7 A**. We calculated the mRNA upregulation for each subunit as a fraction of the total.

#### **2.4.9 Statistical analysis**

All statistical analyses were carried out in GraphPad PRISM using mixed-effect multiple comparisons or two-way ANOVA comparisons with Bonferroni correction, p-values  $\leq 0.05$  were considered significant.



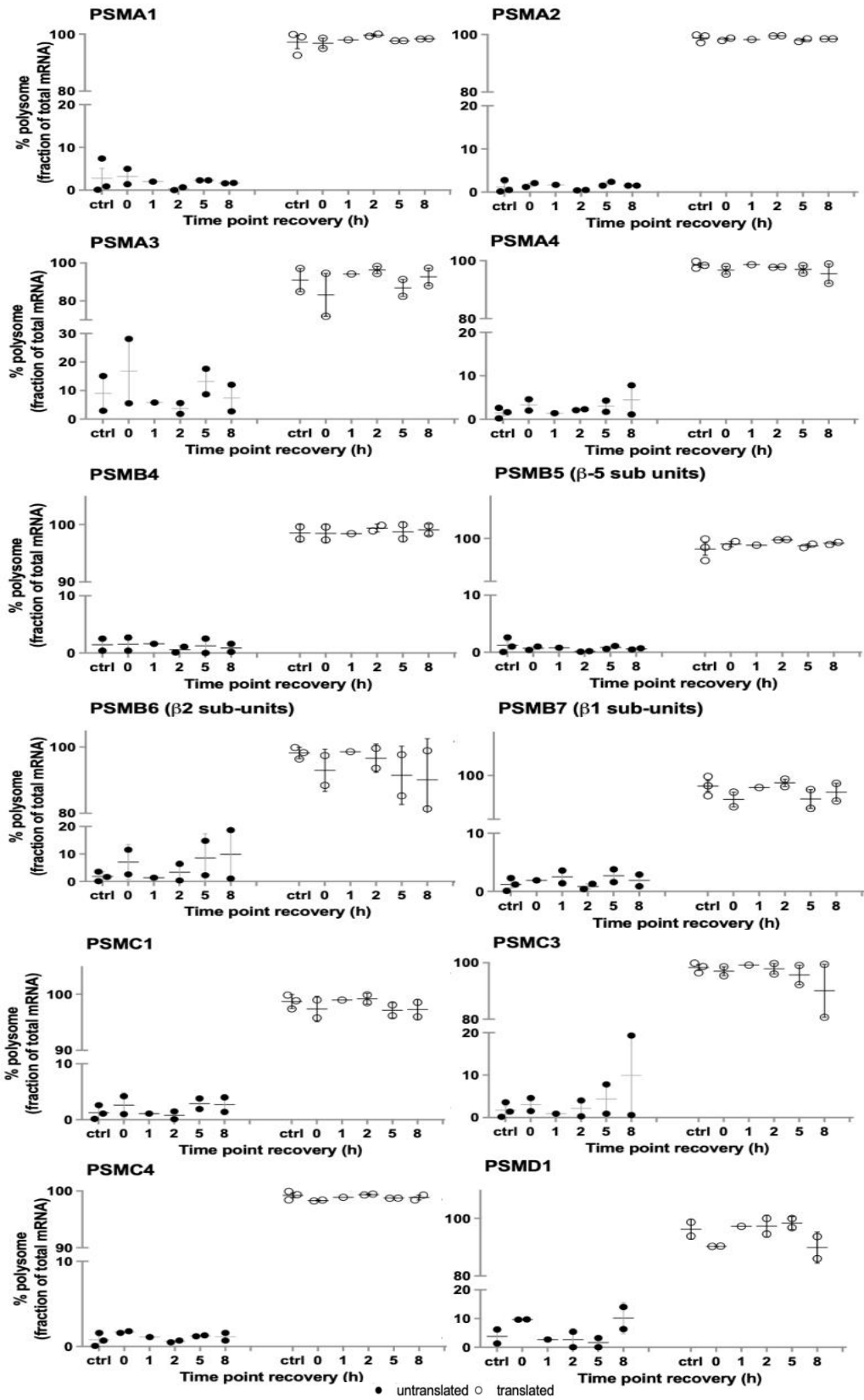
## 2.5 Supporting Information

**Table SI 2.1 List of qPCR primers used in Figure 2.4 and Figure 2.6**

Classification	Subclassification	Genes name	qPCR primers	Primer sequences
20S Proteasome				
Alpha subunits	$\alpha 6$	PSMA1	forward	AGAGCTTGACAGCTCATCAGA
			reverse	CAAGACGAGACACAGGCAGT
	$\alpha 2$	PSMA2	forward	GCCCCGATTACAGAGTGC
			reverse	TGGACGAACACCACCTGA
			forward	GGTGCGCAACTCTACATGATTG
	$\alpha 3$	PSMA4	reverse	TCCTCTGCCTCGTATCAGTTATTAA
			forward	CATTGGCTGGGATAAGCA
beta subunits	$\alpha 4$	PSMA7	reverse	ATGCATGTGGCCTTCCAT
			forward	CTGTGCTTTGGATGACAACG
	$\beta 7$	PSMB4	reverse	CGATGTAGCGGGTGATGTACT
			forward	TCTCGGCCAGATGGTGAT
			reverse	CACATAAACCAGGAAAGCT
	$\beta 5$	PSMB5	forward	GCTTGCCCAACATGGGTGATC
			reverse	ATCATAGGCCTGCTCCACTT
$\beta 2$	PSMB6	forward	CCAAGGAAAGAGTGTCTGCAA	
		reverse	TGCATCAGTACAGGGCATCT	
$\beta 1$	PSMB7	forward	ATTGACCTCTGCCGTATCAG	
		reverse	CTGTTTCTCCAGCACCTCA	

**Table SI 2.1 (Continuation) List of qPCR primers used in Figure 2.4 and Figure 2.6**

Classification	Subclassification	Genes name	qPCR primers	Primer sequences
		PA700 19S Regulatory Particles		
ATPase subunits	Rpt2	PSMC1	forward	GATGACCTCTCTGGTGCTGA
			reverse	CCCTTTCAGGGATTGAGAAA
	Rpt5	PSMC3	forward	CCAGAAATCATGCAGATCCAC
			reverse	GCCTTCCATGTAGTCCTCGT
	Rpt3	PSMC4	forward	CCGCCAGAAGAGATTGATTT
			reverse	ACAATGTAGCGGTTTTACAG
Non-ATPase subunits	Rpn2	PSMD1	forward	GGGGACCTCTTCAATGTCAA
			reverse	TAGGCAGAGCCTCATTGGCT
	Rpn10	PSMD4	forward	AGGAGGAGGCCCGGC
			reverse	TCACCTTCTGTCTTCC
	Rpn7	PSMD6	forward	CAGTCAGCTGCTGGAATCAT
			reverse	TGGTACTGCCAGTTCTTGCT
	Rpn6	PSMD11	forward	GCTGCTCTGGAAACAATTCA
			reverse	ACAGATGCACCAAAATGAGGA
	Rpn5	PSMD12	forward	TTGGTCCACACTTGTGAGG
			reverse	TGAGAGAAAGGCTTCGGACT
	Rpn11	PSMD14	forward	TTGGATGGAAGGTTTGACAC
			reverse	CCACATGTTCTCCTCCAAATGA



**Figure SI 2.1 Complete data of the translation efficiency of proteasome subunits mRNA.** HAP1 wt cells were pulse-treated with Btz (100nM) and recovered in a drug-free medium at the time indicated. A.) shows an example of manual polysome fractions differentiation

*based on the  $OD_{260}$  in HAP1 wt cells. Fractions 1 – 12 contain the untranslated mRNA, fractions 13 – 32 contain mRNA bound to mono-ribosome (the 80S), while fraction >30 contains the mRNA bound to polysome. The polysome fractions in **A.**) were combined and divided into untranslated (black dot) and translated (clear dot) fractions, then the RNA was isolated from these two combined fractions. Subsequent qPCR of proteasome subunits was performed. **B.**) shows the percentage of translated mRNA of proteasome subunits of control 0h and 8h proteasome recovery (collected after treatment). The percentage of translated mRNA was calculated as a fraction of the total. **A.**) was generated in Microsoft Excel, while **B.**) was generated through PRISM by plotting the averages of 2 – 3 biological replicates.*

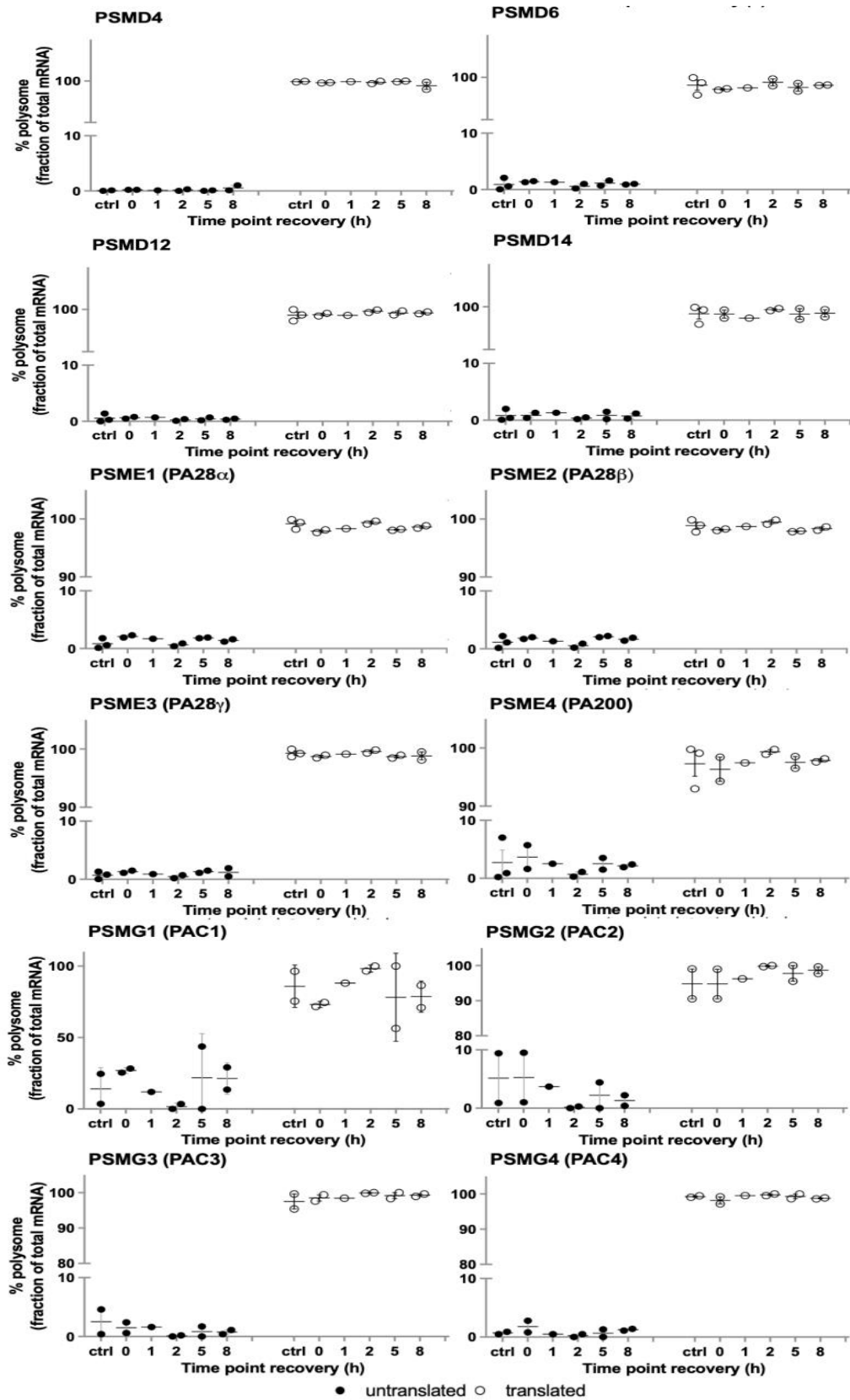
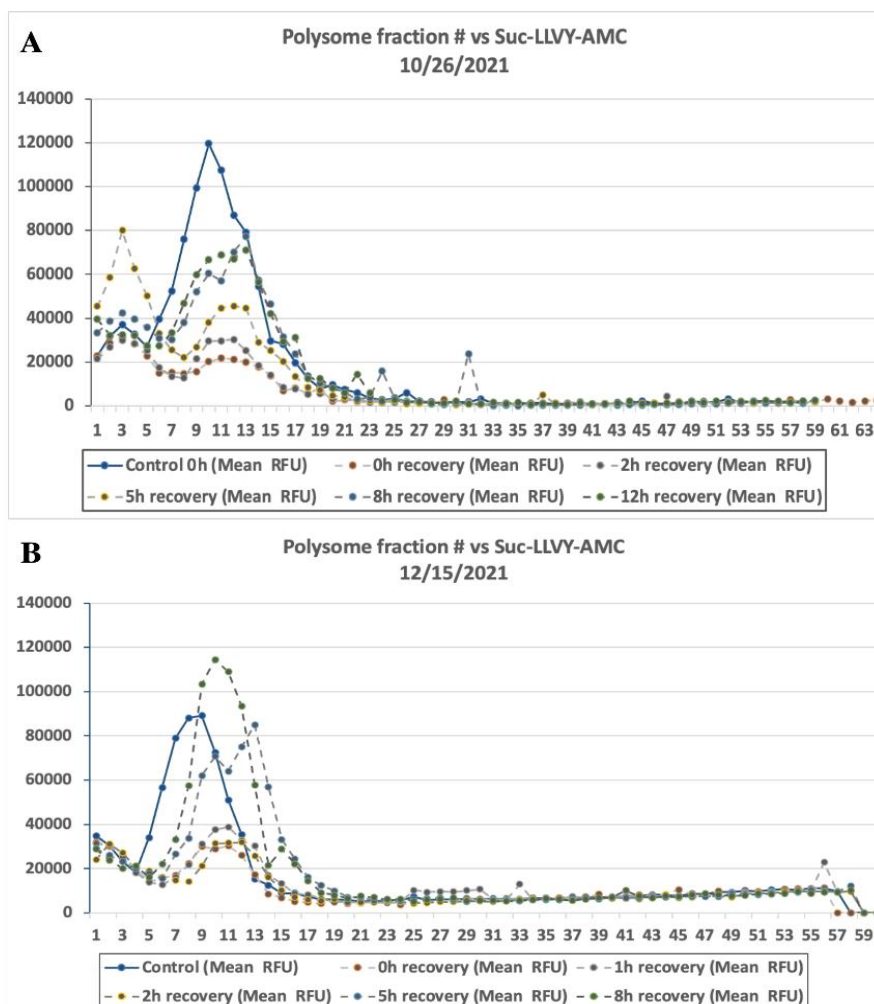


Figure SI 2.1 (Continuation) Complete data of the translation efficiency of proteasome subunits mRNA. HAP1 wt cells were pulse-treated with Btz (100nM) and recovered in a drug-free medium at the time indicated. A.) shows an example of manual polysome

fractions differentiation based on the  $OD_{260}$  in HAP1 wt cells. Fractions 1 – 12 contain the untranslated mRNA, fractions 13 – 32 contain mRNA bound to mono-ribosome (the 80S), while fraction > 30 contains the mRNA bound to polysome. The polysome fractions in A.) were combined and divided into untranslated (black dot) and translated (clear dot) fractions, then the RNA was isolated from these two combined fractions. Subsequent qPCR of proteasome subunits was performed. B.) shows the percentage of translated mRNA of proteasome subunits of control 0h and 8h proteasome recovery (collected after treatment). The percentage of translated mRNA was calculated as a fraction of the total. A.) was generated in Microsoft Excel, while B.) was generated through PRISM by plotting the averages of 2 – 3 biological replicates.



**Figure SI 2.2 The recovery of proteasome activity is initiated by an increase in 26S/30S proteasome activity.** Both plots A.) and B.) show Increased active 26S/30S proteasome during the restoration of proteasome activity. The  $\beta 5$  proteasome activity was measured using fluorogenic substrates Suc-LLVY-AMC in each fraction following the  $OD_{260}$  measurement. Each value is a mean (RFU) plotted against the fraction number and was generated using Microsoft Excel. Plots A.) and B.) are two independent biological replicates. At the same time, each dot represents one technical replicate.

## Chapter 3 CRISPR KO screen coupled with transcriptome analysis approach to study the possible activated pathway by DDI2

### 3.1 Introduction

The human DNA Damage Inducible 1 Homolog II (DDI2) is a multi-domain multi-functional protein<sup>16</sup>. DDI2 is a novel aspartic protease whose biological function has not been widely studied and characterized<sup>15-16</sup>. A recent *in-vivo* study reported deletion of DDI2 causes severe developmental failure and mid-late gestation embryonic lethality<sup>20</sup>, suggesting the importance of DDI2 in normal function and development. However, the extent to which this deletion of DDI2 leads to the disruption of normal physiological homeostasis is still unknown.

CRISPR/Cas9 gene editing has provided a powerful method for determining the effects of genetic manipulation (knockout or knock-in) to understand gene function. We decided to use CRISPR KO clones of HAP1 cells described in the previous chapter to gain insights into DDI2 biology. As mentioned in **Chapter 2**, CRISPR KO may have off-target effects<sup>103</sup>. Therefore, screening using CRISPR KO derived from different guide RNAs (gRNAs) is deemed necessary. Combo approaches using CRISPR KO screening with transcriptomic or proteomic analysis have been widely used to reduce laborious and unnecessary time-consuming experiments.

Whole transcriptome analysis offers global RNA expression levels profile of a biological system consisting of coding RNA to non-coding RNA (small-interfering RNAs, microRNAs, and extracellular RNAs). It may provide insight into transcriptome behavior changes and signaling pathways upon DDI2 KO.

This study began with an effort to answer the challenge of understanding the biological function of hDDI2 in cells using a combination of the CRISPR KO gene-editing screening and a high-throughput whole-transcriptome analysis approach. Assessing the changes in RNA molecules level and integrating it with pathway analysis may provide us with possible affected pathways upon knockout of DDI2.

### 3.2 Results

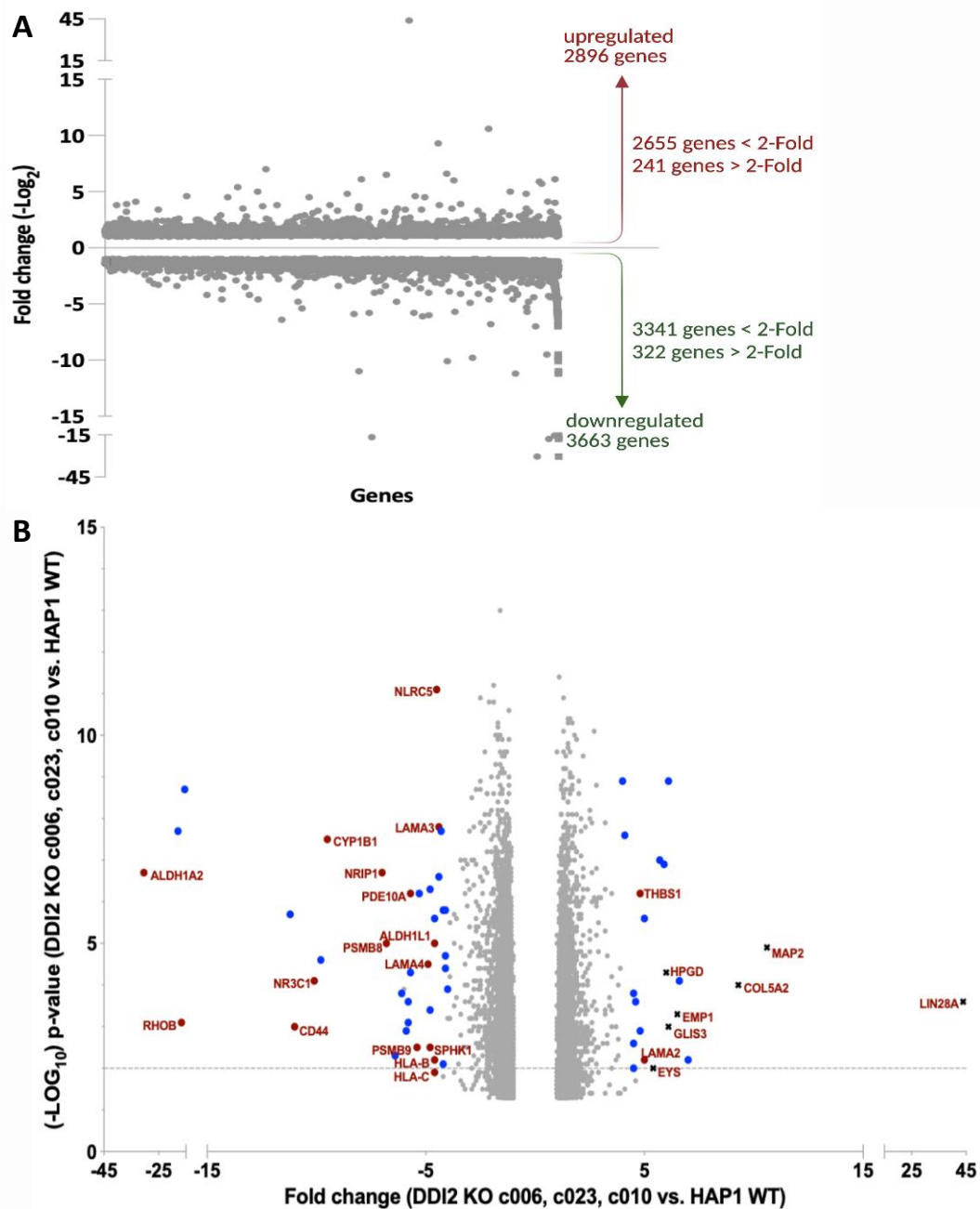
We first looked at the global gene expression analysis to investigate the extent to which transcriptional changes can lead to the disruption of normal physiological homeostasis after the deletion of DDI2 in HAP1 cells. This will provide us with qualitative and quantitative information on whether deletion of DDI2 causes transcriptional amplification or repression.

We performed total RNA-sequencing (RNA-seq) as described in the materials and methods sections, generating a raw RNA count of 23196 genes. These raw RNA count genes were further normalized, representing the relative level of the genes accurately. This normalization resulted in 14371 Normalized RNA count genes. Next, Differentially Expression Genes analysis (DEGs) was conducted to the normalized RNA count of DDI2 CRISPR KO cells (c006, c023, and c010) against HAP1 wt. We found 6559 genes exhibiting statistically different RNA levels between all three CRISPR KO cells against HAP1 wt (p-value <0.05, see **Figure 3.1 A, Figure SI 3.1**). Global RNA expression analysis shows the downregulation of 3663 genes (55.8%) and upregulation of 2896 genes (44.2%) out of 6559 differentially expressed genes. While analyzing the global RNA expression, we also



performed QIAGEN-Ingenuity Pathway Analysis (IPA) on 6559 Differentially expressed genes (DEGs).

When the gene function is perturbed, transcriptome analysis usually focuses on significantly changed RNA. Of these 6559 genes significantly differentially expressed, only 60 showed more than 4-fold changes between HAP1 wt and DDI2 CRISPR KO cells (**Figure 3.1 B**). However, of these 60 genes, only 19 were reported to be involved in the pathways analysis (labeled in the red dots, as shown in **Figure 3.1 B**). These genes are ALDH1A2, RHOB, CD44, NR3C1, CYP1B1, NRIP1, PSMB8, PSMB9, PDE10A, HLA-B, HLA-C, LAMA4, LAMA3, NLRC5, ALDH1L1, LAMA4, LAMA3, SPHK1, LAMA2, THBS1, and COL5A2. Moreover, to avoid bias or overlooking one gene in our pathway analysis, we included all the highly variable expression genes between DDI2 CRISPR KO derived from gRNA1 (clone c006, c023) and gRNA2 (clone c010). These genes are LIN28A, MAP2, COL5A2, EMP1, HPGD, GLIS3, and EYS (labeled as x, as shown in **Figure 3.1, Table SI 3.1**). However, the high variability of these gene expressions should also be considered when interpreting the pathway analysis results.



**Figure 3.1** Global RNA expression of from total RNA-seq of DDI2 CRISPR KO cells (clones c006, c023, and c010) vs. HAP1 wt. Transcriptome analysis shows **A.**) 6559 Differentially expressed genes (DEGs). Of these 6559 genes, 3341 were repressed, and 322 genes showed >2-Fold downregulation. Of these 6559 genes, 2655 were amplified, and 241 genes showed >2-Fold upregulation. **B.**) Volcano plot of the p-value ( $-\text{Log}_{10}$ ) vs. Fold-change of DEGs in DDI2 CRISPR KO cells (clones c006, c023, and c010) vs. HAP1 wt. Of these 60 DEGs, **1.**) 7 genes (labeled as x) show high variability between DDI2 CRISPR KO c023, c006 (gRNA1) vs. DDI2 CRISPR KO c010 (gRNA2). **2.**) Only 19 genes were reported involved in the Qiagen-IPA analysis (labeled as red dots, not including COL5A2). **3.**) The remaining 35 significant DEGs were not reported in the IPA analysis (labeled as blue dots). The volcano plot was generated in Graphpad-PRISM by plotting the p-value ( $-\text{Log}_{10}$ ) and fold-change from 2 biological replicates of HAP1 wt against each of the DDI2 CRISPR KO clones (c006, c023, and c010) with a Fold-change cut-off  $\geq 4$ , p-value  $\leq 0.01$ .

### 3.2.1 DDI2 involves in the antigen presentation, xenobiotic metabolism, and cancer signaling pathway

Based on differentially expressed genes analysis, we asked what signaling pathways were affected by the loss of DDI2 function. Abovementioned, we performed pathway analysis from differentially expressed genes. We obtained 72 significant signaling pathways that might be affected by the loss of function of DDI2 (cut-off  $p$ -value  $<0.05$ ; please refer to **Figure SI 3.2** for a complete list).

We then used  $p$ -value  $<0.05$  and  $-1 < z$ -score  $< +1$  as a cut-off and focused on the 25 most affected signaling pathways (**Figure 3.2**). These pathways were related to antigen presentation, inflammation, xenobiotic metabolism, and cancer signaling. In this study, we will limit the discussion of pathway analysis within the scope of cancer-related signaling pathways.

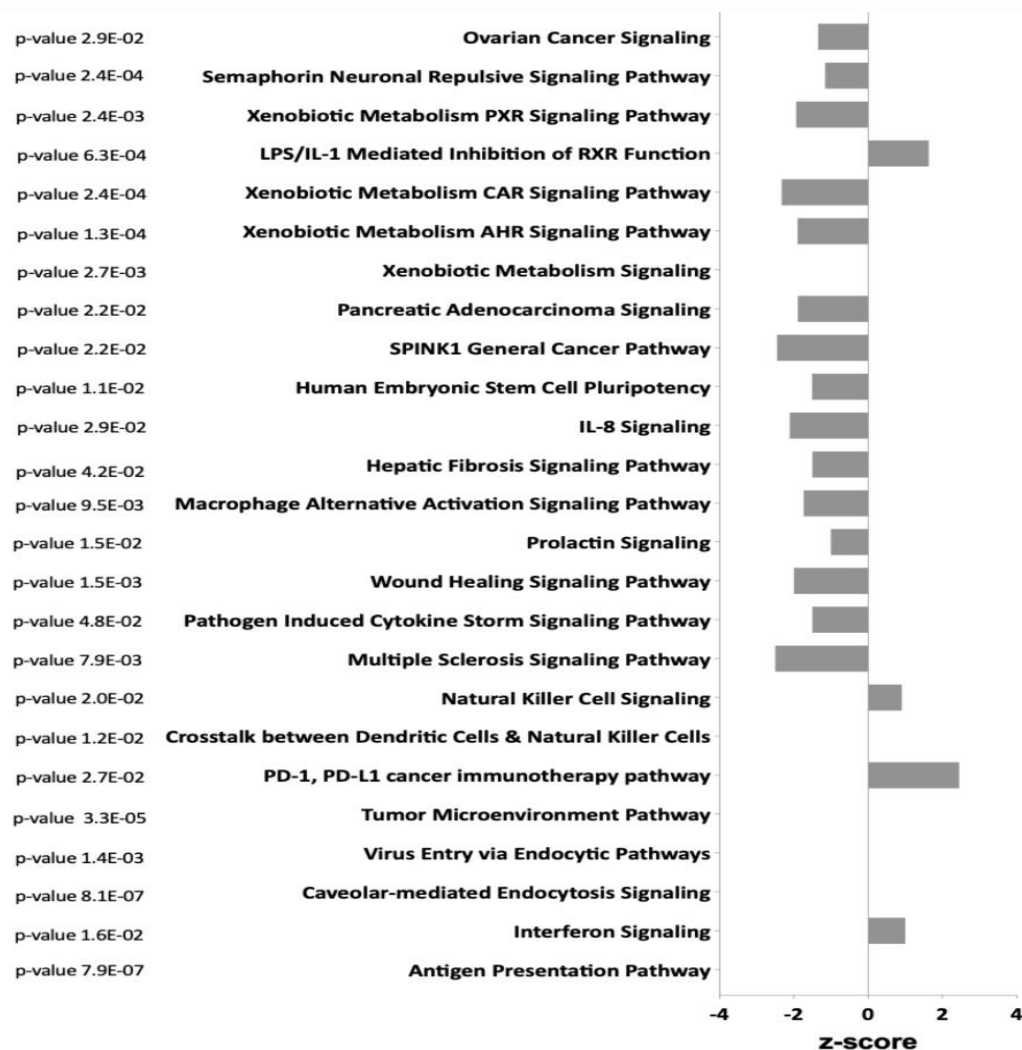
Although antigen presentation shows no direct positive or negative correlation with DDI2 knockout, the downregulation of MHC class I and its activator molecules involved in the antigen presentation pathways, in turn, affects several other significant pathways (**Table 3.1**). These pathways were interferon signaling, caveolar-mediated endocytosis signaling, virus Entry via endocytic pathways, tumor microenvironment pathway, PD-1/PD-L1 cancer immunotherapy pathway, crosstalk of dendritic cells & natural killer cells, natural killer cell signaling, and multiple sclerosis signaling (**Table 3.1, Figure 3.2**). Coincidentally, all these pathways, including NLRC5, HLA-B, HLA-C, PSMB8, and PSMB9 genes, were downregulated more than 4-fold (**Table 3.1, Figure 3.1 B**). As mentioned above, total RNA-seq analysis usually focuses on significantly changed relative RNA; hence, significant impairment of antigen presentation molecules suggests DDI2

may affect these pathways. The downregulation of MHC class I antigen-presenting peptides generally correlates with the worsening prognosis in cancer<sup>126</sup>.

**Table 3.1 Ingenuity Canonical Pathways Analysis and their Molecules.**

Ingenuity Canonical Pathways	-log p-value	z-score	Molecules
Antigen Presentation Pathway	6.1		B2M,HLA-A,HLA-B,HLA-C,HLA-E,MR1,NLRCS,PSMB8,PSMB9
Interferon Signaling	1.8	1.0	PSMB8
Caveolar-mediated Endocytosis Signaling	6.1		B2M,CD55,EGFR,FLNA,FYN,HLA-A,HLA-B,HLA-C,HLA-E,ITGA11,MR1
Virus Entry via Endocytic Pathways	2.8		B2M,CD55,FLNA,FYN,HLA-A,HLA-B,HLA-C,HLA-E,MR1
Tumor Microenvironment Pathway	4.5		CD44,CSPG4,FOXO6,HLA-A,HLA-B,HLA-C,HLA-E,IL6R,LEPR,MIMP14,MMP24,PDGFA,VEGFC
PD-1, PD-L1 cancer immunotherapy pathway	1.6	2.4	B2M,HLA-A,HLA-B,HLA-C,HLA-E,IL12A,MR1
Crosstalk of Dendritic Cells & Natural Killer Cells	1.9		HLA-A,HLA-B,HLA-C,HLA-E,IL12A
Natural Killer Cell Signaling	1.7	1.0	B2M,FYN,HLA-A,HLA-B,HLA-C,HLA-E,HSPA8,IL12A,PAK6,PTK2B
Multiple Sclerosis Signaling	2.1	-2.5	DUSP6,HLA-A,HLA-B,HLA-C,HLA-E,IL12A,TNFSF9
Pathogen Induced Cytokine Storm Signaling	1.3	-1.5	COL4A5,COL5A2,IFIH1,IL12A,IL6R,NLRCS,STXB2,TNFSF9,VEGFC
Wound Healing Signaling	2.8	-2.0	CEBPB,COL4A5,COL5A2,EGFR,IL12A,LAMA2,LAMA3,NFKBID,PDGFA,TNFSF9,VEGFC
Prolactin Signaling	1.8	-1.0	CEBPB,FYN,NR3C1,PRKCH,PLR
Macrophage Alternative Activation Signaling	2.0	-1.7	CEBPB,EIF4EBP1,IL10RB,IL12A,KLF4,LPL,NFKBID,NR3C1,THBS1
Hepatic Fibrosis Signaling	1.4	-1.5	CEBPB,CACNA2D1,ITGA11,LEPR,NFKBID,PDGFA,PRKAR1B,PRKCH,RHOB,RND3,VEGFC,WNT3
IL-8 Signaling	1.5	-2.1	EGFR,EIF4EBP1,GNB3,GPLD1,PLD3,PRKCH,PTK2B,RHOB,RND3,VEGFC
Human Embryonic Stem Cell Pluripotency	1.7	-1.5	BMP4,ID1,KLF4,NOG,NR5A2,PDGFA,PRKCH,SPHK1,WNT3
SPINK1 General Cancer Pathway	1.9	-2.4	EGFR,IL6R,MT1E,MT1M,MT1X,MT2A
Pancreatic Adenocarcinoma Signaling	1.7	-1.9	EGFR,ERBB2,GPLD1,HDAC9,NOTCH1,PLD3,VEGFC
Xenobiotic Metabolism Signaling	2.6		AHR,AHRR,ALDH1A2,ALDH1L1,CHST15,CHST2,CYP1B1,GSTM3,GSTO2,NR113,NRIP1,PPP2R2C
Xenobiotic Metabolism AHR Signaling	3.9	-1.9	ABCG2,AHR,AHRR,ALDH1A2,ALDH1L1,CYP1B1,GSTM3,GSTO2,NRIP1
Xenobiotic Metabolism CAR Signaling	3.6	-2.3	ALDH1A2,ALDH1L1,CHST15,CHST2,EGFR,GSTM3,GSTO2,NR113,NRIP1,PPP2R2C
LPS/IL-1 Mediated Inhibition of RXR Function	3.2	1.6	ABCC4,ACSL6,ALDH1A2,ALDH1L1,CHST15,CHST2,GSTM3,GSTO2,NR113,NR5A2,PLTP
Xenobiotic Metabolism PXR Signaling	2.6	-1.9	ALDH1A2,ALDH1L1,CHST15,CHST2,GSTM3,GSTO2,NRIP1,PRKAR1B
Semaphorin Neuronal Repulsive Signaling	3.6	-1.2	CD44,CSPG4,CSPG5,ERBB2,FYN,ITGA11,NRP2,PAK6,PLXNA3,PRKAR1B,SEMA5A,VCAN
Ovarian Cancer Signaling	1.5	-1.3	CD44,EGFR,HDAC9,PRKAR1B,VEGFC,WNT3

© 2000-2023 QIAGEN. All rights reserved.



**Figure 3.2 QIAGEN - Ingenuity Pathway Analysis (IPA) from differentially expressed genes (DEGs) of DDI2 CRISPR KO (clone c006, c023, and c010) vs. HAP1 wt.** This graph represents the significant pathways with a p-value cut-off  $<0.05$  and  $-1 \leq z\text{-score} \leq 1$ . HAP1 wt represents 2 biological replicates against each biological replicate of DDI2 CRISPR KO.

Under normal conditions, MHC class I can be regulated by NF $\kappa$ B, Interferon regulatory factors (IRFs), and NOD-like receptor caspase recruitment (CARD) domain containing 5 (NLRC5)<sup>127-129</sup>. Interferon- $\gamma$  induces the expression of immunoproteasome subunits (PSMB8, PSMB9) and Proteasome activators PA28 to generate antigen peptides for MHC class I molecules (HLA-A, HLA-B, HLA-C, and  $\beta$ 2M)<sup>127-129</sup>. These nascent MHC class I peptides were then translocated to the

ER by Transporter associated with antigen-presenting (TAP1) and bound within the stabilized complexes of the MHC class I groove or processed by ER-aminopeptidase (ERAP1/2)<sup>128-129</sup>. NLRC5 is an interferon- $\gamma$ -inducible protein that functions as a transcriptional activator of the MHC class I antigen presentation (HLA-A, HLA-B, HLA-C), immunoproteasome (PSMB8, PSMB9),  $\beta$ 2-microglobulin ( $\beta$ 2M), and Transporter TAP1/2<sup>127</sup>. Hence, the Disruption of MHC class I and its activator molecules are one of the immune evasion signature mechanisms by cancer cells in the tumor microenvironment, allowing them to escape immune surveillance that can recognize and kill them<sup>126-127</sup>. Therefore, first, we asked whether the transcriptional expression of all MHC class I genes, the immunoproteasome, B2M, and TAP1/2, were also altered. **Table 3.2** suggests all genes related to cancer cells' immune evasion were downregulated.

**Table 3.2 Knockout of DDI2 showed the downregulation of MHC class I and its activators in the antigen presentation pathways.**

DEGs of DDI2 KO c010, c023, c006 Vs. HAP1 wt			
Genes	P-value (-Log <sub>10</sub> )	FDR step-up (-Log <sub>10</sub> )	Fold-change
B2M	1.27E-04	9.27E-04	<b>-2.5</b>
TAP1	4.03E-02	8.53E-02	<b>-2.0</b>
NLRC5	7.39E-12	2.65E-08	<b>-4.5</b>
HLA-A	1.18E-02	3.24E-02	<b>-3.7</b>
HLA-B	6.98E-03	2.14E-02	<b>-4.6</b>
HLA-C	1.19E-02	3.25E-02	<b>-4.6</b>
HLA-E	3.72E-03	1.30E-02	<b>-2.2</b>

As mentioned, NF $\kappa$ B, interferon regulator factors (IRFs), and NLRC5 regulate MHC class I<sup>128</sup>. Since NLRC5 is downregulated, it is possible that NF $\kappa$ B and IRFs pathways to induce MHC class I were activated. Here, we asked whether genes involved in other MHC class I antigen peptide generation pathways are upregulated

due to NLRC5 downregulation. The MHC class I can be generated by activating the JAK/STAT (STAT 1, 2, and 3)<sup>128</sup>. The Interferon- $\gamma$  may activate STAT1 (IFN-type II) and upregulates IRFs<sup>128</sup>. Alternatively, Interferon- $\alpha$  or Interferon- $\beta$  may also activate JAK1/STAT1 via IFN-Type I by stimulating GAS pathways and upregulating the IRF1<sup>128</sup>. Interferon- $\alpha$  or Interferon- $\beta$  may also trigger JAK1/STAT1 activation by enabling interferon-stimulated response element (ISRE) pathways and directly upregulating MHC Class I<sup>128</sup>. However, based on our transcriptome analysis, upon knockout of DDI2, transcriptional STAT 1, STAT2, and IRF1 were only changed by  $\pm 1$ -fold (p-values and FDR-set up data were not shown), suggesting that all pathways involving MHC class I were either downregulated or remained the same.

Next, since NLRC5 and MHC class I expression was correlated with cytotoxic CD8<sup>+</sup> T-cell and NK cell activation<sup>126-127</sup>; we asked whether downregulation of these genes alters the transcriptional expression of CD8<sup>+</sup> T-cell and NK cells activation biomarkers, such as CD8A, CD56, CD25, perforin (PRF1), and granzyme A (GZMA). CD25, PRF1, and GZMA are highly expressed in the cancer's cell surface of activated CD8<sup>+</sup> T-cells and natural killer (NK) cells<sup>127</sup>. CD8A is a marker for CD8<sup>+</sup> T-cells, while CD56 is a marker for NK cells<sup>127</sup>. Our transcriptome analysis suggests none of the cytotoxic effector cell markers (CD8A, CD56, CD25, PRF1, and GZMA) are expressed. However, other studies have shown CD44 is a feature of activated CD8 T cells and memory<sup>130-131</sup>. Interestingly, transcriptome data indicate that CD44 is also downregulated by 11-fold (**Figure 3.1 B**). CD44 is a transmembrane glycoprotein, a receptor for the extracellular matrix of hyaluronic acid (HA) ligand, known for its function in cell adhesion and

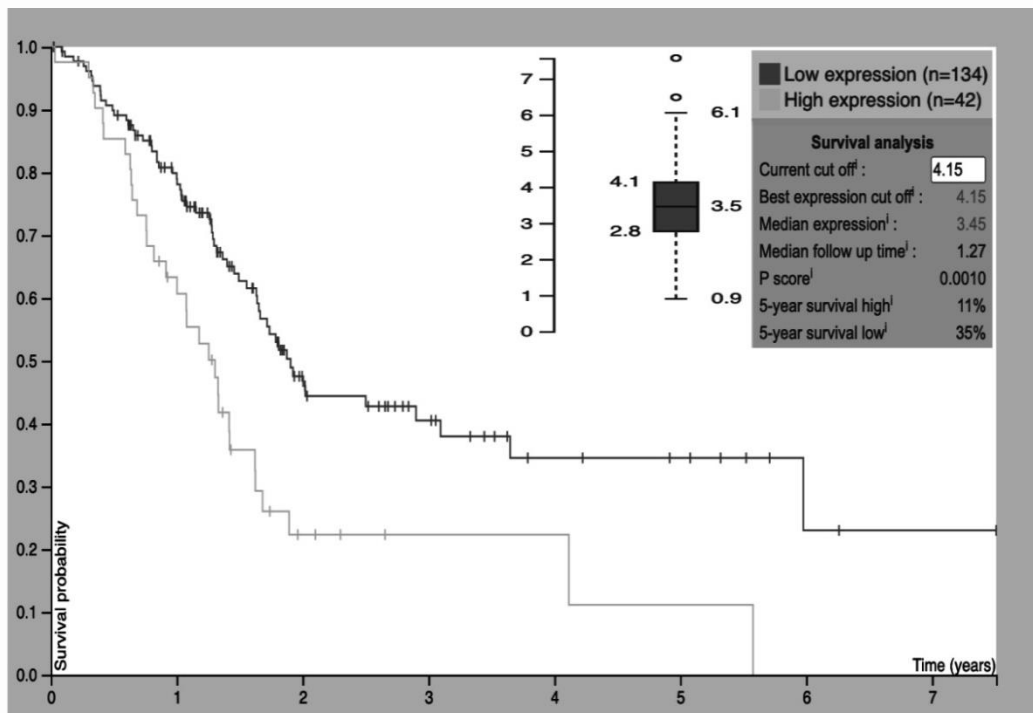


migration<sup>130-131</sup>. Pathway analysis suggest that the downregulation of CD44 and MHC class I expression significantly affects the tumor microenvironment (**Table 3.1, Figure 3.2**). Furthermore, CD44 positively regulates Programmed death-ligand 1 (PD-L1) ligand in Triple-negative breast cancer (TNBC) and non-small cell lung cancer (NSCLC)<sup>132</sup>. In line with this, our pathway analysis also suggests deregulation of antigen presentation, in turn, positively correlates with anti-PD-1/PD-L1 immunotherapy (**Figure 3.2**).

Another mechanism of immune suppression in the tumor microenvironment is the impairment of cytotoxic T-cells by masking the surface of activated T-cells to avoid the cancer cells being recognized by these activated T-cells. The PD-1/programmed death-ligand 1 (PD-L1) is an inhibitor of immune response T-cells effector activation<sup>133</sup>. PD-1 is predominantly expressed on the surface of the activated T-cells, B lymphocytes cells, monocytes, dendritic cells (DCs), regulatory T-cells (Tregs), NK T-cells, and macrophages<sup>133</sup>. At the same time, PD-L1 is predominantly expressed on cancer cells and antigen-presenting cells<sup>133</sup>. PD-1 and PD-L1 inhibit immune checkpoints by inhibiting tumor-infiltrating lymphocyte activation resulting in loss of cytotoxic T lymphocyte (CTL) cytotoxicity (T-cells exhaustion)<sup>133</sup>. We will discuss the possible pathway activation of PD-1/PD-L1 by DDI2 via CD44 in **Section 3.3**.

Additionally, although some studies report downregulation of MHC class I and its activators as a mechanism for escaping antitumor immunity in solid tumors, there are exclusion cases to consider<sup>127</sup>. **Figure 3.2** pathway analysis suggests DDI2 KO inhibits SPINK1 general cancer signaling pathway and Pancreatic adenocarcinoma signaling pathway. Interestingly, analysis of the TCGA data shows a significant

increase in the survival probability of pancreatic cancer patients with low DDI2 RNA tissue levels compared to high DDI2 RNA tissue levels (p-value 0.0010, **Figure 3.3**). Based on TCGA data, pancreatic cancer patients with low DDI2 RNA levels have a 35% 5-year survival rate, whereas pancreatic cancer patients with high DDI2 RNA levels have an 11% 5-year survival rate. Integrated transcriptome and pathway analysis in **Table 3.1** and **Table 3.3** suggests DDI2 involves deregulating growth factor molecules. However, the exact mechanism by which DDI2 affects pancreatic cancer is unknown.



**Figure 3.3** Pancreatic cancer patients with low DDI2 RNA levels have a significant probability of survival compared to those with high DDI2 RNA levels (p-value 0.0010). The Kaplan-Meier curve from Protein Data Bank Atlas April 2<sup>nd</sup>, 2023, shows TCGA - DDI2 FKPM RNA levels from 176 pancreatic cancer tissue patients with median FKPM DDI2 RNA levels of 3.5.

**Table 3.3 Knockout of DDI2 showed the downregulation of several genes that may inhibit Pancreatic Adenocarcinoma signaling pathways.**

<b>DEGs of DDI2 KO c010, c023, c006 Vs. HAP1 wt</b>			
<b>Genes</b>	<b>P-value (-Log<sub>10</sub>)</b>	<b>FDR step-up (-Log<sub>10</sub>)</b>	<b>Fold-change</b>
EGFR	1.17E-02	3.21E-02	<b>-2.6</b>
ERBB2	4.26E-05	4.00E-04	<b>-2.2</b>
GPLD1	9.97E-08	5.41E-06	<b>-2.8</b>
HDAC9	1.65E-02	4.20E-02	<b>2.1</b>
NOTCH1	8.10E-04	3.92E-03	<b>2.1</b>
PLD3	9.44E-03	2.72E-02	<b>-2.1</b>
VEGFC	2.11E-05	2.37E-04	<b>-2.5</b>

The serine protease inhibitor Kazal type (SPINK) is present in the pancreas. Deregulation of SPINK can lead to pancreatitis and cancer. SPINKs have been correlated with worsening cancer prognosis, including Pancreatic cancer<sup>134</sup>. SPINK general cancer pathway involves deregulating SPINK proteins, such as SPINK1, 2, 4, 5, 6, 7, and 13<sup>134</sup>. Meanwhile, based on **Table 3.1**, the SPINK general cancer pathway activation involves EGFR, IL6R, MT1E, MT1M, MT1X, and MT2A molecules. We then looked at the fold-change transcriptional expression of these molecules. **Table 3.4** shows the upregulation of metallothionein proteins and the downregulation of EGFR and Interleukin 6 (IL-6). In cancer, SPINK1 activates EGFR and increases STAT3 phosphorylation (**Figure SI 3.3**)<sup>134</sup>. At the same time, IL-6 reportedly upregulates SPINK1 via the STAT3 pathways<sup>134</sup>. Upregulation of SPINK1 promotes tumor progression via downregulation of Metallothioneins expression (**Figure SI 3.3**)<sup>134-135</sup>. SPINK1 promotes cell proliferation via PI3K/AKT and MEK/ERK signaling (**Figure SI 3.3**)<sup>134-135</sup>. In contrast, DDI2 KO downregulates EGFR and IL-6, further altering STAT3 activation and SPINK1 transcriptional expression. Alteration of SPINK1 expression further upregulates Metallothioneins

and suppresses tumor progression. This suggests that high expression of DDI2 may worsen the prognosis of SPINK1-dependent cancers, while low expression of DDI2 may inhibit SPINK1-dependent cancers.

**Table 3.4 Knockout of DDI2 showed downregulation of EGFR and Interleukin 6 (IL-6) that may inhibit SPINK general cancer signaling pathways.**

<b>DEGs of DDI2 KO c010, c023, c006 Vs. HAP1 wt</b>			
<b>Genes</b>	<b>P-value (-Log<sub>10</sub>)</b>	<b>FDR step- up (-Log<sub>10</sub>)</b>	<b>Fold-change</b>
MT1X	1.65E-05	2.00E-04	<b>2.5</b>
MT1E	1.13E-03	5.11E-03	<b>2.6</b>
MT1M	4.60E-04	2.48E-03	<b>2.0</b>
MT1F	3.63E-05	3.56E-04	<b>1.6</b>
MT2A	3.00E-06	5.69E-05	<b>2.9</b>
IL6R	1.29E-02	3.45E-02	<b>-2.6</b>
EGFR	1.17E-02	3.21E-02	<b>-2.6</b>

**Table 3.1** and **Figure 3.2** also suggest that knockout of DDI2 is significantly involved in pathways associated with xenobiotic metabolic signaling (p-values of 3.9, 3.6, and 2.6, respectively). These pathways include significantly downregulated genes ALDH1A2, ALDH1A1, NRIP1, and CYP1B1, as shown in **Figure 3.1 B**. We will limit our discussion to the ALDH1 gene in the cancer context in **Section 3.3**.

ALDH1A2 and ALDH1L1 are members of the Aldehyde dehydrogenase 1 family (ALDH1)<sup>136-138</sup>. High expression of ALDH1 (A1, A2, but not L1) worsens the prognosis of cancer<sup>136-138</sup>. ALDH1 is an oxidoreductase whose function requires nicotinamide-adenine dinucleotide phosphate–positive (NADP<sup>+</sup>) to generate co-enzyme NADPH for glycolysis and the tricarboxylic acid (TCA) cycle<sup>136</sup>. Aerobic glycolysis is one hallmark of cancer<sup>136</sup>. Alteration of ALDH1A2 further reduces the metabolites expression involved in the glycolysis pathway<sup>136</sup>. Thus, we asked

whether the downregulation of ALDH1A2 and ALDH1L1 as a consequence of DDI2 knockout decreased the transcriptional level of genes involved in glycolysis and the TCA cycle. Surprisingly, we found that the transcriptional level of genes involved in glycolysis was downregulated (GPI, PFKL, and ALDOC). Glucose-6-phosphate isomerase reduced glucose metabolism, reducing 6-phosphofructokinase (PFKL) and AldolaseC (ALDOC). We also found that DDI2 KO reduced most genes' transcriptional expression in metabolite transport and enzymatic metabolic pathways (**Table 3.5**). Although, the discussion of the downregulation of most of these genes is beyond the scope of this thesis.

**Table 3.5 Knockout of DDI2 showed genes' downregulation in xenobiotic metabolism and transporter.**

<b>DDI2 KO c010, c023, c006 Vs. HAP1 wt</b>			
<b>Genes</b>	<b>P-value (-Log<sub>10</sub>)</b>	<b>FDR step-up (-Log<sub>10</sub>)</b>	<b>Fold- change</b>
ABCG2	4.53E-06	7.59E-05	<b>-2.7</b>
AHR	1.11E-04	8.30E-04	<b>-3.1</b>
AHRR	1.03E-03	4.72E-03	<b>-2.0</b>
ALDH1A2	1.81E-07	7.94E-06	<b>-30.4</b>
ALDH1L1	1.11E-05	1.48E-04	<b>-4.6</b>
CYP1B1	3.21E-08	2.59E-06	<b>-9.5</b>
GSTM3	8.23E-07	2.29E-05	<b>3.0</b>
GSTO2	5.77E-04	2.98E-03	<b>-3.9</b>
NRIP1	2.20E-07	8.98E-06	<b>-7.0</b>
CHST15	3.41E-03	1.21E-02	<b>-2.2</b>
CHST2	1.30E-08	1.49E-06	<b>-2.5</b>
NR1I3	3.21E-04	1.89E-03	<b>-2.6</b>
PPP2R2C	1.22E-05	1.58E-04	<b>-3.4</b>
GPI	2.98E-02	6.73E-02	<b>-1.3</b>
PFKL	3.61E-05	3.55E-04	<b>-1.1</b>
ALDOC	4.66E-02	9.56E-02	<b>-2.1</b>
ASS1	8.03E-03	2.39E-02	<b>-2.7</b>
ASNS	1.20E-07	5.95E-06	<b>-2.7</b>

Based on the transcriptome and pathway analysis, we identified CD44 and ALDH1A2 as possible targets of DDI2 to accelerate tumor progression. Both of these genes have been reported as cancer stem cell markers and are involved in worsening cancer prognosis.

### **3.2.2 DDI2 knockout does not alter the transcriptional expression of proteasome subunits but reduces the transcriptional level and activity of immunoproteasome**

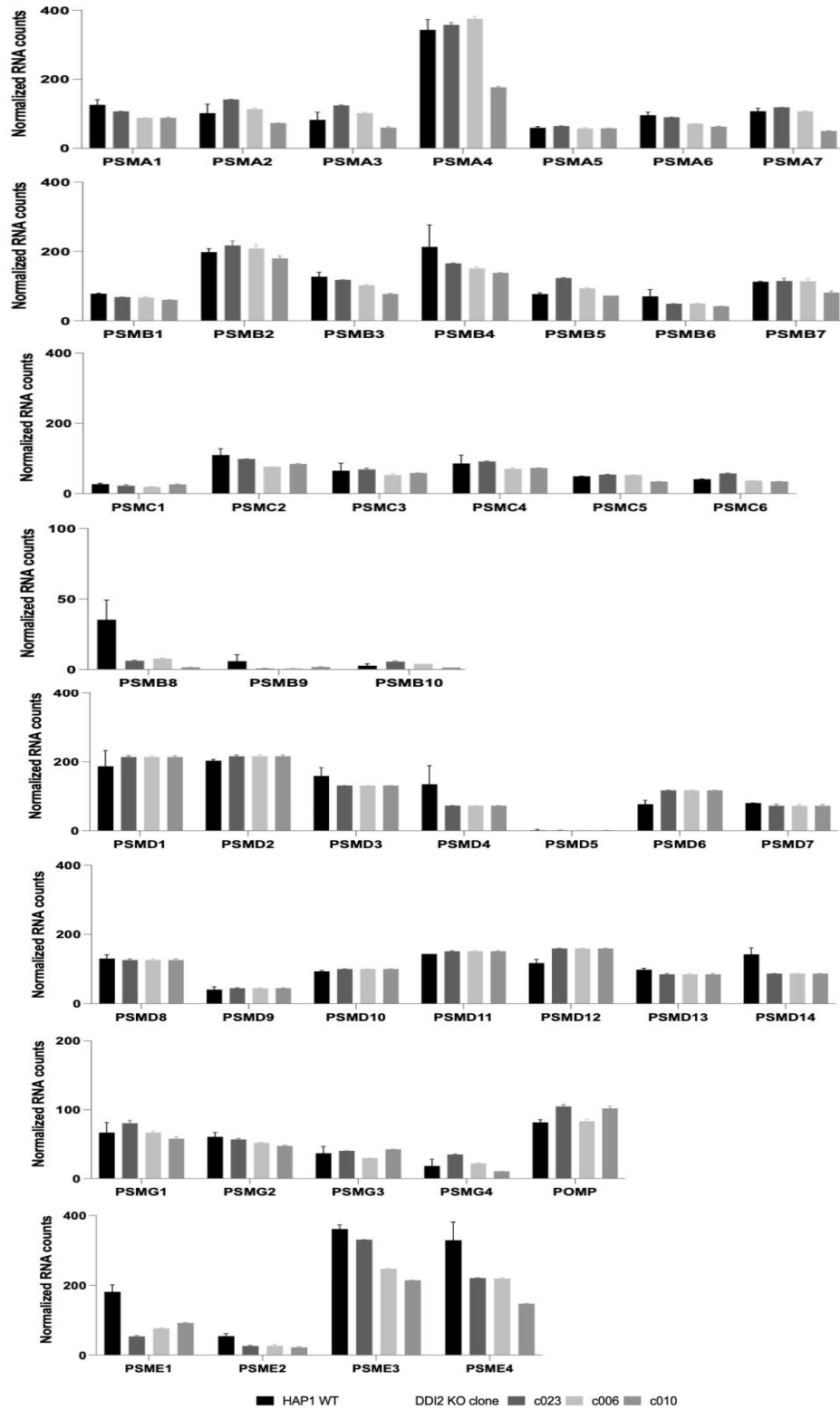
In the previous section, we found that DDI2 KO lowers the transcriptional level of immunoproteasome PSMB8 and PSMB9. Pathway analysis in **Figure 3.2** suggests PD-1/PD-L1 immunotherapy signaling and interferon signaling pathway activation. **We asked what signaling pathways might involve immunoproteasome downregulation.** Pathway analysis suggests that DDI2 KO-mediated PSMB8 and PSMB9 downregulation may affect ARE-mediated mRNA degradation pathway, Huntington's disease signaling, BAG2 signaling, Protein ubiquitination pathway, and Chaperone-mediated autophagy signaling (complete IPA data are not shown).

Because we are interested in pathways associated with the UPS, we follow up this by looking at the global transcriptional expression of proteasome subunits, proteasome-chaperones, and activators. Interestingly, differentially expressed gene analysis from transcriptome data demonstrates that DDI2 KO slightly alters the basal mRNA levels of the proteasome subunits ( $-1.7 < \text{fold-change} < 1.3$ ). Knockout of DDI2 only shows significant downregulation of PSMB8 (LMP7/ $\beta$ 5i) and PSMB9 (LMP2/ $\beta$ 1i) immunoproteasomes, but not for PSMB10 (MECL-1/LPMP10/ $\beta$ 2i). Both PSMB8 and PSMB9 were downregulated about 5 – 7-fold. This suggests the possible role of DDI2 in regulating immunoproteasome activity (see **Figure 3.1 A, Table 3.6**).

**Table 3.6 Knockout of DDI2 does not alter the transcriptional expression of proteasome subunits except the immunoproteasome PSMB8 and PSMB9.**

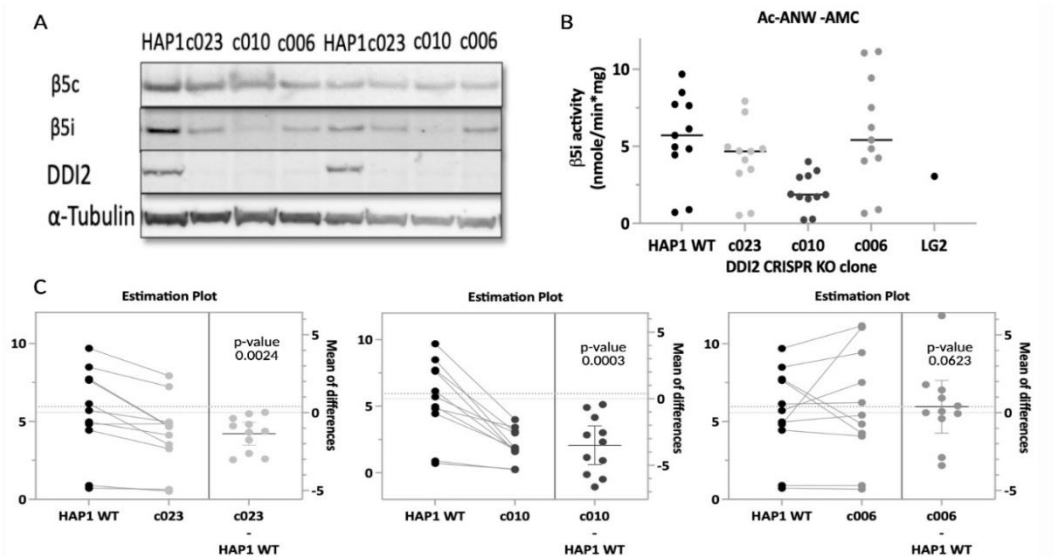
<b>DEGs of DDI2 KO c010, c023, c006 Vs. HAP1 wt</b>			
<b>Proteasome Genes</b>	<b>P-value (-Log<sub>10</sub>)</b>	<b>FDR step-up (-Log<sub>10</sub>)</b>	<b>Fold-change</b>
PSMA1	3.17	2.47	-1.3
PSMA4	2.70	2.10	-1.1
PSMA6	3.64	2.83	-1.3
PSMA7	3.20	2.49	-1.2
PSMB1	4.77	3.69	-1.2
PSMB3	3.30	2.58	-1.3
PSMB4	1.55	1.19	-1.4
PSMB5	4.49	3.48	1.3
PSMB6	1.97	1.52	-1.5
PSMB7	1.45	1.12	-1.1
PSMB8	4.99	3.86	<b>-6.8</b>
PSMB9	2.49	1.94	<b>-5.4</b>
PSMC1	1.53	1.18	-1.2
PSMC2	2.04	1.57	-1.3
PSMC5	4.16	3.24	-1.0
PSMD2	2.19	1.70	-1.1
PSMD3	2.85	2.22	-1.3
PSMD6	1.57	1.21	1.2
PSMD7	2.51	1.95	-1.2
PSMD8	1.66	1.28	-1.1
PSMD11	7.15	5.38	-1.1
PSMD12	2.56	1.99	1.2
PSMD13	4.97	3.84	-1.3
PSMD14	4.27	3.32	-1.7
PSME1	1.37	1.05	-1.2
PSME3	5.25	4.05	-1.5





**Figure 3.4** Normalized RNA count of proteasome subunits, proteasome chaperone, and proteasome activators from RNA-seq data of DDI2 KO c010 (gRNA2), DDI2 KO c023, c006 (gRNA1) vs. HAP1 wt. The Figure was generated by plotting the average normalized RNA count of 2 biological replicates of HAP1 wt against each DDI2 CRISPR KO cell (clone c006, c023, and c010).

Ideally, we should confirm whether we can see a decrease in PSMB8 and PSMB9 mRNA levels to protein levels using western blotting or ABP and see whether the activity of immunoproteasome is reduced. We first measured activity to confirm whether DDI2 functionally affects immunoproteasome activity. Here, we measured the immunoproteasome activity using the fluorogenic substrate Ac-ANW-AMC in the untreated HAP1 wt and all DDI2 CRISPR KO cells. Interestingly, our preliminary results in **Figure 3.5** suggest that DDI2 knockout significantly lowers immunoproteasome activity at the basal level in two DDI2 CRISPR KO clones (c023 and c010), but not DDI2 CRISPR KO clone c006.



**Figure 3.5 Knockout of DDI2 lowers the basal level of immunoproteasome  $\beta 5i$  activity.** *A.)* basal protein level of  $\beta 5c$ ,  $\beta 5i$ , and DDI2 in HAP1 and DDI2 KO cells. *B.)*  $\beta 5i$  activity was presented as nmole/min\*mg. Scatter plots are generated on Graphpad PRISM by plotting 11 biological replicates of each sample. Paired t-test statistical analysis was conducted by plotting the average  $\pm$  SEM of each DDI2 CRISPR KO clone against HAP1 wt. LG2 cell was used as a positive control for immunoproteasome activity measurement.

Our preliminary data suggest a role for DDI2 in lowering immunoproteasome activity. However, we cannot ascertain whether alteration of immunoproteasome activity is a direct effect of DDI2 on immunoproteasome subunits or due to downregulation of MHC class I and its activators.

### 3.3 Discussion

In **Section 3.2.1**, Integrated transcriptome and pathway analysis suggested 72 significant pathways affected by the loss-of-function of DDI2 that can be classified into antigen presentation, inflammation, xenobiotic metabolism, and cancer signaling pathways. In this study, we will limit the discussion of pathway analysis within the scope of cancer-related signaling pathways. First, we will discuss the role of DDI2 in antigen presentation and cancer signaling pathways. Then, we will discuss the role of DDI2 in xenobiotic metabolism and cancer signaling pathways.

Pathway analysis suggests DDI2 KO deregulates antigen presentation pathway via reducing transcriptional expression of MHC class I and its activator ( $\beta$ 2M, TAP1, NLRC5, HLA-A, HLA-B, HLA-C, HLA-E, PSMB8, and PSMB9). Depletion of MHC class I and its activator molecules is significantly involved in the antigen presentation pathways and tumor microenvironment pathways. The MHC class I antigen-presenting peptide modulates the activation of cytotoxic CD8<sup>+</sup> T cells effector, essential for immune responses against cancer. Likewise, High NLRC5 expression correlates with cytotoxic CD8<sup>+</sup> T-cell activation and increases cancer patient survival rates<sup>127</sup>. Conversely, the downregulation of transcriptional NLRC5 further reduces the transcription level of MHC class I genes and reduces the activation of CD8<sup>+</sup> T-cells. Disruption of MHC class I and its activator molecules is one of the immune evasion mechanisms by cancer cells, allowing cancer cells to escape the CD8<sup>+</sup> cytotoxic T-cell–dependent antitumor activity in the tumor microenvironment<sup>127</sup>. We then asked about how DDI2 KO decreases CD8<sup>+</sup> T cell activation. Our Differentially expressed genes analysis suggests that Knockout of

DDI2 reduced the expression of CD44 by 11-fold. CD44 is one of the activated CD8 T cells and memory marker<sup>130-131</sup>. CD44 is a cancer stem cell marker that plays a role in cell adhesion and migration necessary for epithelial-to-mesenchymal transition (EMT)<sup>130-132, 139</sup>. A previous study reported a positive correlation between high CD44<sup>+</sup> with PD-L1 expression from patient tumors TCGA<sup>132</sup>. Knockdown of CD44 dramatically suppressed the mRNA and protein level of PD-L1 in Triple-negative breast cancer (TNBC) and non-small cell lung cancer (NSCLC)<sup>132</sup>. Silencing CD44 also reduced colony-forming proliferation assay of these cancer<sup>132</sup>, suggesting that CD44 is tumorigenic. CD44 must be processed into the shorter CD44-ICD form by proteases before being translocated into the nucleus and activating gene transcription mediated by CREB-binding protein (CBP)/p300<sup>132, 140-141</sup>. This processed CD44 (CD44-ICD) is recruited into the PD-L1 promotor region via CREB1<sup>132, 140-141</sup>. As outlined in **Section 3.2.1**, PD-L1 is predominantly expressed in cancer and antigen-presenting cells<sup>133</sup>. PD-L1 and its receptor PD-1 inhibit immune checkpoints by inhibiting tumor-infiltrating lymphocyte activation resulting in loss of cytotoxic T lymphocyte (CTL) cytotoxicity (T-cells exhaustion)<sup>133</sup>.

High expression of CD44<sup>+</sup> results in elevated expression of PD-L1 and causes T-cell exhaustion, further worsening the tumor microenvironment and cancer prognosis. **It means that decreasing CD44 expression via DDI2 KO may reduce the PD-L1 expression and positively correlate with anti-PD-1/ PD-L1 immunotherapy.** Supporting this notion, pathway analysis showed that CD44 expression and MHC class I downregulation significantly affect the tumor microenvironment. In line with this, our pathway analysis also suggests

deregulation of antigen presentation, in turn, positively correlates with anti-PD-1/PD-L1 immunotherapy.

Moreover, our pathway analysis also suggests the role of DDI2 in xenobiotic metabolism. Knockout of DDI2 significantly downregulates specific genes such as ALDH1 (A2 and L1). High expression of ALDH1 (A1, A2, L1) worsens the prognosis of cancer<sup>136-138</sup>. ALDH1A2 was found to be a downstream target and directly affected by TAL1<sup>+</sup> T-cell acute lymphoblastic leukemia (T-ALL)<sup>136</sup>. Overexpression of ALDH1A2 augments the tumorigenicity of cancer cells in T-ALL by ensuring aerobic glycolysis for cell survival and reducing ROS production to avoid apoptosis. In contrast, depletion of ALDH1A2 affects the aerobic glycolysis metabolic pathway of hematologic cancer T-ALL<sup>136</sup>. Decreased level of energy for cancer cell viability and increased ROS level further induce cancer cell apoptosis.

Additionally, ALDH1<sup>+</sup> is a marker for cancer stem cells (CSCs) and is reportedly upregulated in ovarian and breast cancer<sup>137, 142</sup>. Kaplan–Meier survival curve analysis suggests high ALDH1 level correlates with poor prognosis and reduces the survival rate of these T-ALL and ovarian cancer patient<sup>136-137</sup>. High ALDH cells exhibit lower intracellular ROS levels<sup>137</sup>. Similarly, chemo-resistant cancer stem cells maintain lower intracellular ROS levels to escape apoptosis via upregulating antioxidant enzymes<sup>137</sup>. ALDH-high cells showed increased transcriptional expression of Nrf2, SOD2, and HO-1, markers for the antioxidant properties<sup>137</sup>. Yet, we did not see significant changes in the expression of these genes in our transcriptome data. Studies suggest that not all CSCs and tumorigenic cells depend on ALDH1<sup>+</sup><sup>136-137</sup>. Some studies reported CD44<sup>+</sup>/CD24 cancer stem cell markers

may also increase tumorigenicity and resistance to chemotherapy<sup>132, 139, 142</sup>. As mentioned above, our pathway analysis in **Table 3.1** suggests DDI2 KO may inhibit ovarian cancer by lowering CD44 levels by 11-folds.

Likewise, DDI2 KO may inhibit ALDH1<sup>+</sup>-dependent cancer by downregulating the ALDH1A2 level. Although, integrated transcriptome and pathway analysis do not correlate ALDH1 depletion with specific cancer or chronic myelogenous leukemia (CML). DDI2 may accelerate ovarian cancer tumorigenicity by increasing CD44<sup>+</sup> or ALDH1<sup>+</sup> levels, yet whether the effect of DDI2 on CD44<sup>+</sup> or ALDH1<sup>+</sup> levels is direct or indirect, again, remains for further research.

Previously, we also found that DDI2 KO lowers the transcriptional level of immunoproteasome PSMB8 and PSMB9. However, pathway analysis also shows knockout of DDI2 does not significantly affect the Protein ubiquitination pathway. We then examined the global transcriptional expression of proteasome subunits, proteasome-chaperones, and activators. We found that DDI2 KO slightly alters the basal mRNA levels of the proteasome subunits ( $-1.7 < \text{fold-change} < 1.3$ ). Knockout of DDI2 only shows significant downregulation of PSMB8 (LMP7/ $\beta$ 5i) and PSMB9 (LMP2/ $\beta$ 1i) immunoproteasomes, but not for PSMB10 (MECL-1/LPMP10/ $\beta$ 2i). This suggests the possible role of DDI2 in regulating immunoproteasome activity. To confirm this, we performed immunoproteasome activity assays. We found that immunoproteasome activity is significantly impaired in DDI2 KO CRISPR clones c023 and c010 but not in clone c006. Hence, further confirmation in other DDI2 KO hematopoietic cancer cells is still needed. Hematopoietic cells are known for expressing immunoproteasome more than

constitutive proteasome. Confirmation in other hematopoietic cancer cells by knocking out or overexpressing DDI2 may help clarify whether DDI2 indeed regulates the basal expression of the immunoproteasome. Suppose DDI2 does regulate immunoproteasome, then overexpression of DDI2 should increase immunoproteasome activity. Moreover, since the expression of PSMB8 and PSMB9 are both regulated by inflammatory cytokines interferon- $\gamma$  (IFN- $\gamma$ ) and tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), then it might be worth pursuing to experiment under the condition of overexpression/depletion of inflammatory cytokines.

To summarize, using a combination of Differentially expressed genes and pathway analysis; we identified two significantly downregulated genes in DDI2 KO cells: CD44 and ALDH1A2. These genes have been reported as markers of cancer stem cells. We also found that knocking out DDI2 has significantly downregulated the transcriptional expression of immunoproteasome PSMB8 and PSMB9, which subsequently decreases immunoproteasome activity.

### **3.4 Material and Methods**

#### **3.4.1 Cell lines and cell culture**

Cell lines and cell culture methods are described in **Chapter 3**.

#### **3.4.2 RNA Isolation**

Total RNA samples from each mammalian cell HAP1 wt, HAP1 wt c631, and DDI2 CRISPR KO (c010, c023, and c006) were extracted using Qiagen RNeasy® Mini Kit (Cat. #74104). The quality and concentration of RNA were determined using NanoDrop2000 (Thermo Scientific™) with approximately > 2µg of total RNA in each sample. The samples were kept on dry ice before being shipped on dry ice to LC Sciences – Texas.

#### **3.4.3 Total RNA Sequencing (RNA-Seq)**

##### **RNA isolation, Library Preparation, RNA-seq**

Total RNA sequencing was conducted at LC Sciences - Texas. The total RNA quality and purity were tested to ensure high-quality sequencing results. RNA samples should meet the following criteria: UV ratio 260/280 >1.8, UV ratio 230/260 >1.0, and RNA Integrity Number (RIN)  $\geq$ 7. The UV 260/280 and 230/260 ratios were measured using a UV spectrophotometer. The integrity of RNA was assessed using automated electrophoresis Agilent Technologies Bioanalyzer 2100 instrument and RNA 6000 Nano LabChip assay Kit (Agilent, Cat. #5067-1511). The obtained electropherogram was analyzed using Bioanalyzer 2100 Expert



(B.02.08.SI648) software; see **Supporting Information Figure SI 3.4** for the RIN of each sample.

The total RNA sequencing was performed by LC Science using Ribo-Zero ribosomal RNA reduction chemistry. To this end, the ribosomal RNA (rRNA) was removed according to the protocol described in the Ribo-Zero Gold rRNA Removal Kit (Illumina, Cat. #MRZG12324). The remaining RNA was fragmented into short fragments using high-temperature divalent cations (NEBNext® Magnesium RNA Fragmentation Module, Cat. #E6150S) and reverse-transcribed (SuperScript™ II Reverse Transcriptase, Invitrogen, Cat. #1896649). The cDNA was further used for dUTP incorporation, Adenosine tailing, adapter ligation, and dUTP strand degradation, followed by PCR enrichment to create a cDNA library. The DNA fragments for ligation to the indexed adapter were prepared by adding an A-base to the blunt end of each strand. The adapters utilized in this process contain a T-base overhang to facilitate binding with the A-tailed fragmented DNA. Single or dual index adapters were bound to the fragments with size selection ranging from 300-600bp using AMPureXP beads to ensure top-notch sequencing results. After treating the U-labelled second-stranded DNA with heat-labile UDG enzyme (NEB, cat.m0280, USA), the bound product by PCR was amplified under the following conditions: initial denaturation at 95°C for 3 min, 8 cycles of denaturation at 98°C for 15 seconds, annealing at 60°C for 15 seconds, and extension at 72°C for 30 seconds, and final extension at 72°C for 5 min. The final cDNA library boasts an average insertion size of 300±50bp, guaranteeing optimal results for downstream applications. The transcriptome was sequenced using the Illumina paired-end RNA-

seq approach (Illumina Novaseq 6000 PE150, 2x 150bp Paired-end sequencing cycle).

### **RNA-seq Data Analysis**

The Gene expression data were pre-processed and filtered as described. The genes containing low count RNA genes <10 were removed<sup>143</sup>. The Gene Expression Profiling (GEP) data were analyzed further using the Partek Flow package to perform differential expression analysis and identify the GEP signatures<sup>143</sup>.

### **3.4.4 Immunoblotting**

Immunoblotting was performed as described in **Chapter 3** with slight modification. The gel electrophoresis was conducted using NuPAGE™ Bis-Tris 10% Midi Gel (Invitrogen™, Cat. #WG1203BOX) at 150V and GenScript Tris-MOPS-SDS running buffer (1X, containing Tris-base, MOPS, SDS, EDTA ddH<sub>2</sub>O to pH 6) Cat. #M00138. The protein was transferred on Immobilon–FL PVDF membrane 0.2μM pore-diameter (Cat. #IPFL00010) using Invitrogen™ Power Blotter 1-Step™ Transfer Buffer (1X) (Cat. #PB7300). The membrane was blocked with Li-Cor Intercept® (TBS) Blocking Buffer (Cat. #927-60001) in TBSt (1X).

**Table 3.7 List of primary and secondary antibodies**

<b>IgG</b>	<b>Type</b>	<b>Host</b>	<b>Dilution</b>	<b>Company</b>	<b>Catalog #</b>
<b>Primary</b>					
Proteasome 20S $\beta$ 5 subunit (human)	pAb	rabbit	1:1000	Enzo	BML-PW8895-0100
Proteasome 20S $\beta$ 5i subunit human(LMP7)	mAb	mouse	1:1000	Enzo	BML-PW8845-0025
$\alpha$ -tubulin (DM1A)	mAb	mouse	1:1000	Cell Signaling	3873S
DDI2	pAb	rabbit	1:2000	Bethyl®Lab	A304-629A
<b>Secondary</b>					
Anti-rabbit HRP-linked		Goat	1:1000	Cell Signaling	7074S
Anti-mouse HRP-linked		Goat	1:1000	Cell Signaling	7076S
Alexa Fluor anti-rabbit 647		Goat	1:3500	Invitrogen™	A32733
Alexa Fluor anti-rabbit 680		Goat	1:3500	Invitrogen™	A20176
IRDye® 800CW anti-Mouse		Goat	1:3500	Li-COR Biosciences	926-32210

### 3.4.5 Immunoproteasome activity assays

The immunoproteasome activity in the untreated cells was determined using fluorogenic substrate Ac-ANW-AMC (Ac-Ala-Asn-Trp-AMC, Cayman Chemicals). For this purpose, 1  $\mu$ g of cell lysates was spiked into 100  $\mu$ l of 100  $\mu$ M of Ac-ANW-AMC in 26S assay buffer (50 mM Tris HCl pH 7.5, 40 mM KCl, 2 mM EDTA, supplemented with a freshly added 1 mM DTT, 100  $\mu$ M ATP), which then preincubated at 37°C for 10 minutes. An increase in AMC-fluorescence was monitored continuously at 37°C on the Biotek Gen5™ Cytation instrument excitations wavelength of 380 nm and emission of 460 nm. The slope curves of the reaction progress were then determined. The immunoproteasome activity was presented as nmole/min\*mg. Each sample was conducted in three technical replicates.

### **3.4.6 Statistical Analysis**

Differential expression analysis (DEGs) from normalized RNA count of total RNA-sequencing was performed in collaboration with Dr. Amit K. Mitra's laboratory at the Center of Pharmacogenomics and Single-Cell Omics Initiative. Our Differential Gene Expression analysis between groups was performed using limma, with a mean fold-change  $>|1|$  and  $p < 0.05$  threshold for reporting significant differential gene expression (DEG)<sup>143</sup>.

Statistical analysis for the immunoproteasome activity was performed using the Paired t-test on Graphpad PRISM (a p-value of  $< 0.05$  is considered significant).

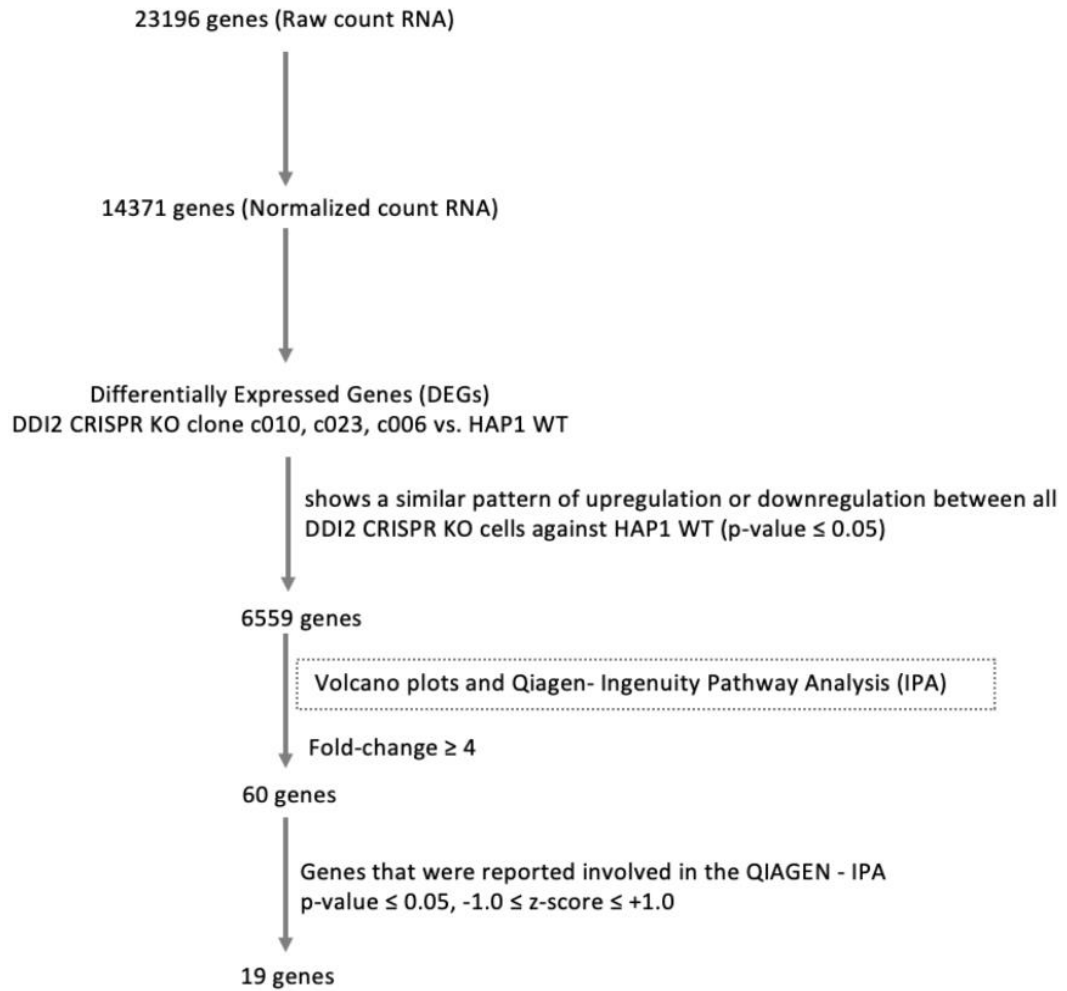
### **3.4.7 Ingenuity pathway analysis (IPA)**

Dr. Amit K. Mitra performed the Ingenuity pathway analysis at the Center of Pharmacogenomics and Single-Cell Omics Initiative. QIAGEN-Ingenuity pathway analysis (IPA) software was used to identify the most significantly affected molecular pathways and whether these predicted pathways were activated or inhibited based on the most significant DEGs<sup>143</sup>. The Benjamini–Hochberg False Discovery Rate (FDR) method was used for multiple testing corrections in DEGs from a data set consisting of all three DDI2 CRISPR KO cells (clone c006, c023, and c010) vs. HAP1 wt. A Z score ( $2.0 \leq Z \leq -2.0$ ) was considered significant for IPA analyses.

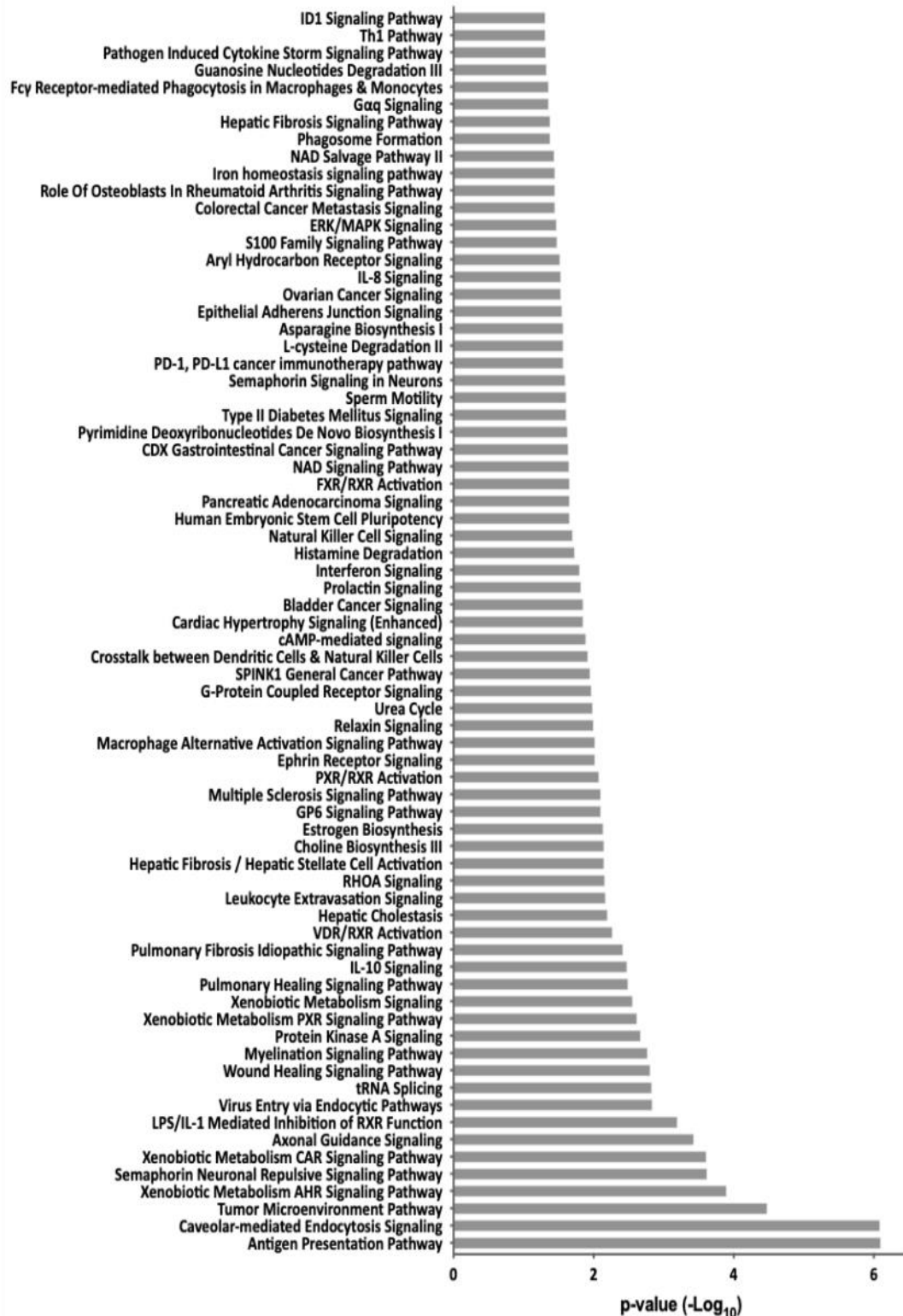
### 3.5 Supporting Information

**Table SI 3.1 Normalized RNA count values of highly variable differentially expressed genes (DEGS) in Figure 3.3**

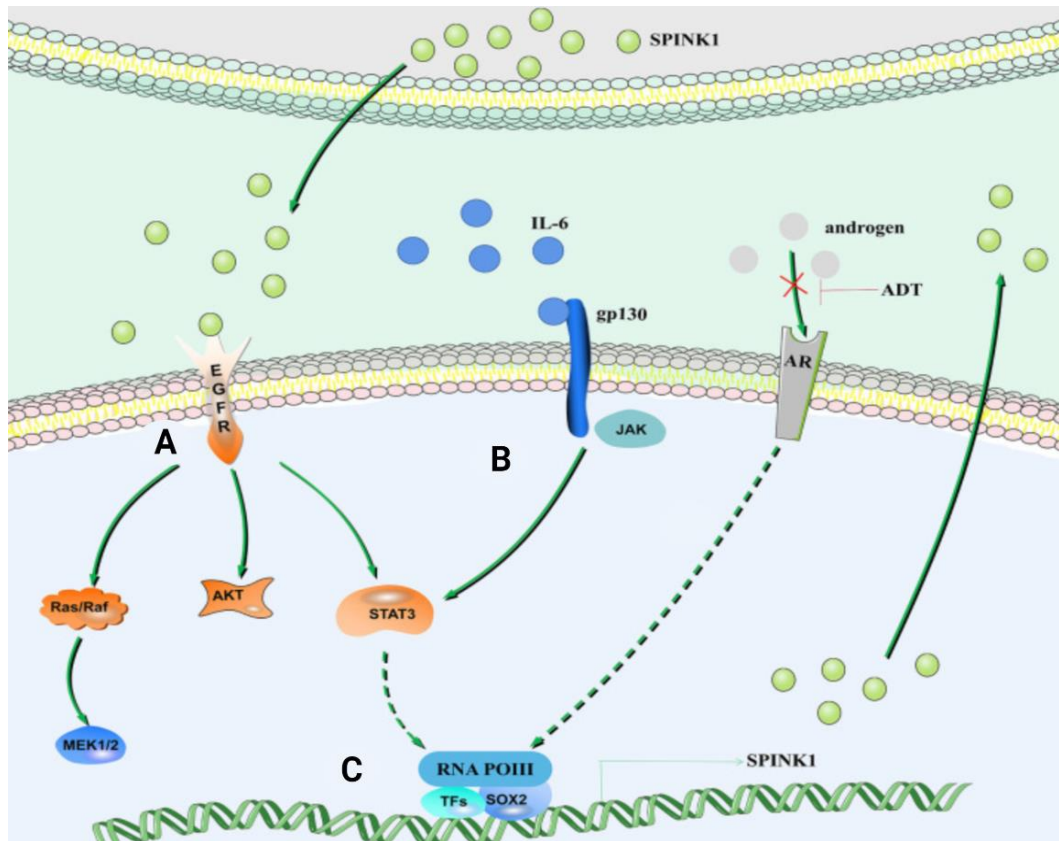
Genes Name	Normalized RNA Count				DDI2 KO vs. HAP1 wt (-Log <sub>10</sub> )			DDI2 KO c006, c023 vs. c010 (-Log <sub>10</sub> )		
	WT	c023	c006	c010	p-value	FDR set-up	Fold-change	p-value	FDR set-up	Fold-change
LIN28A	3.6	284.2	180.0	8.4	3.6	2.8	<b>43.9</b>	2.8	2.3	27.5
GLIS3	5.8	37.4	66.5	1.9	3.0	2.3	<b>6.1</b>	4.9	3.9	27.8
COL5A2	7.5	73.7	132.5	4.1	4.0	3.1	<b>9.3</b>	5.1	4.0	25.6
ANXA1	38.2	34.6	32.8	1.9	2.3	1.8	-1.2	3.6	2.9	18.3
LAMB1	108.5	149.1	145.2	8.4	3.1	2.4	-1.1	6.1	4.8	17.6
MYOF	15.9	27.7	38.9	2.1	1.9	1.5	1.4	7.6	5.7	16.2
EMP1	2.6	15.7	32.1	2.2	3.3	2.6	<b>6.5</b>	3.7	3.0	11.1
RSC1A1	12.5	10.6	25.1	1.6	1.9	1.5	-1.0	4.8	3.9	10.9
HMGN1P38	9.4	13.8	18.7	1.5	2.5	1.9	1.2	7.4	5.6	10.6
ASS1	407.9	18.2	53.9	381.1	2.1	1.6	-2.7	2.6	2.1	-10.6
ATP8A2	18.5	8.0	12.8	119.1	2.2	1.7	2.5	7.4	5.6	-11.5
EYS	1.9	1.3	3.2	25.7	2.0	1.6	<b>5.4</b>	3.3	2.6	-11.5
HPGD	14.7	19.3	17.1	230.0	4.3	3.3	<b>6.0</b>	5.7	4.5	-12.7
CMKLR2	7.8	1.2	1.1	19.7	2.8	2.2	-1.1	4.8	3.8	-18.0
SLC16A7	21.8	0.8	1.8	32.2	5.7	4.3	-1.9	6.5	5.0	-25.6
HOXB9	42.6	0.6	2.6	40.5	4.8	3.7	-2.9	5.1	4.0	-25.7



**Figure SI 3.1 Data processing from total RNA-seq to Differentially Expressed Genes (DEGs) of data set DDI2 CRISPR KO cells clone (c006, c023, c010) vs. HAP1 wt cells.**



**Figure SI 3.2 QIAGEN - Ingenuity Pathway Analysis (IPA) from differentially expressed genes (DEGs) of DDI2 CRISPR KO (clone c006, c023, and c010) vs. HAP1 wt.** Integrated transcriptome and pathway analyses suggest 72 multiple pathways significantly affected by DDI2 deletion. This graph represents the significant pathways with a p-value cut-off <0.05. HAP1 wt represents 2 biological replicates against each biological replicate of DDI2 CRISPR KO.



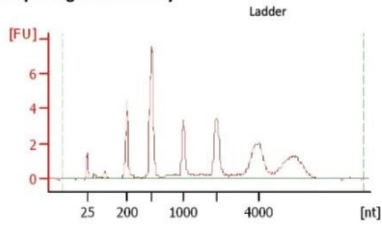
**Figure SI 3.3 SPINK general activation pathway in cancer.** SPINK1 binds and activates EGFR, which increases STAT3 phosphorylation. At the same time, IL-6 reportedly upregulates SPINK1 via the STAT3 pathways. Activation of EGFR by SPINK1 further promotes tumor progression. SPINK1 promotes cell proliferation via PI3K/AKT and MEK/ERK signaling<sup>134</sup>. This Figure is reproduced with permission. Copyright© 2022 Liao, Wang, An, Zhang, Chen, Li, Xiao, Wang, Long, Liu, and Guan.



Assay Class: Eukaryote Total RNA Nano  
 Data Path: \\0...Eukaryote Total RNA Nano\_DE34903435\_2019-11-07\_14-52-57.xad

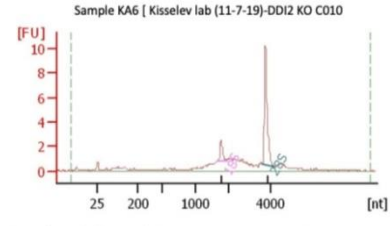
Created: 11/7/2019 2:52:57PM  
 Modified: 11/11/2019 2:58:56PM

**Electropherogram Summary**



**Overall Results for Ladder**

RNA Area: 58.2  
 RNA Concentration: 150 ng/μl  
 Result Flagging Color:    
 Result Flagging Label: All Other Samples

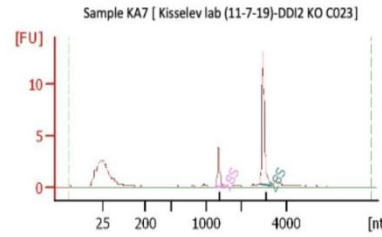


**Overall Results for sample 6 :** Sample 6

RNA Area: 43.9  
 RNA Concentration: 113 ng/μl  
 rRNA Ratio [28s / 18s]: 5.3  
 RNA Integrity Number (RIN): 7.6 (8.02.08, Anomaly Threshold(s) manually adapted)  
 Result Flagging Color:    
 Result Flagging Label: RIN: 7.60

**Fragment table for sample 6 :** Sample 6

Name	Start Size [nt]	End Size [nt]	Area % of totalArea
18S	1,677	1,978	2.0 4.5
28S	3,600	4,191	10.3 23.6

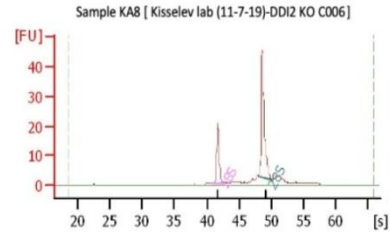


**Overall Results for sample 7 :** Sample 7

RNA Area: 27.8  
 RNA Concentration: 72 ng/μl  
 rRNA Ratio [28s / 18s]: 3.2  
 RNA Integrity Number (RIN): 9 (8.02.08, Anomaly Threshold(s) manually adapted)  
 Result Flagging Color:    
 Result Flagging Label: RIN: 9

**Fragment table for sample 7 :** Sample 7

Name	Start Size [nt]	End Size [nt]	Area % of totalArea
18S	1,250	1,537	3.7 13.2
28S	2,830	3,394	11.9 42.7

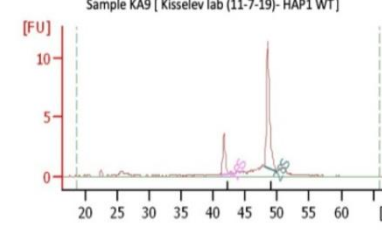


**Overall Results for sample 8 :** Sample 8

RNA Area: 118.8  
 RNA Concentration: 467 ng/μl  
 rRNA Ratio [28s / 18s]: 2.3  
 RNA Integrity Number (RIN): 9.9 (8.02.08)  
 Result Flagging Color:    
 Result Flagging Label: RIN: 9.90

**Fragment table for sample 8 :** Sample 8

Name	Start Size [nt]	End Size [nt]	Area % of total Area
18S	1,591	1,947	21.8 18.3
28S	3,574	4,173	49.3 41.5

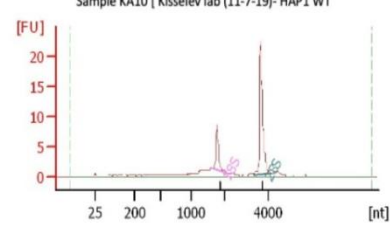


**Overall Results for sample 9 :** Sample 9

RNA Area: 31.4  
 RNA Concentration: 124 ng/μl  
 rRNA Ratio [28s / 18s]: 3.1  
 RNA Integrity Number (RIN): 8.9 (8.02.08)  
 Result Flagging Color:    
 Result Flagging Label: RIN: 8.90

**Fragment table for sample 9 :** Sample 9

Name	Start Size [nt]	End Size [nt]	Area % of total Area
18S	1,637	2,118	4.0 12.7
28S	3,533	4,119	12.3 39.2



**Overall Results for sample 10 :** Sample 10

RNA Area: 67.4  
 RNA Concentration: 174 ng/μl  
 rRNA Ratio [28s / 18s]: 3.0  
 RNA Integrity Number (RIN): 8.2 (8.02.08)  
 Result Flagging Color:    
 Result Flagging Label: RIN: 8.20

**Fragment table for sample 10 :** Sample 10

Name	Start Size [nt]	End Size [nt]	Area % of totalArea
18S	1,640	2,143	8.2 12.2
28S	3,341	4,168	24.6 36.5

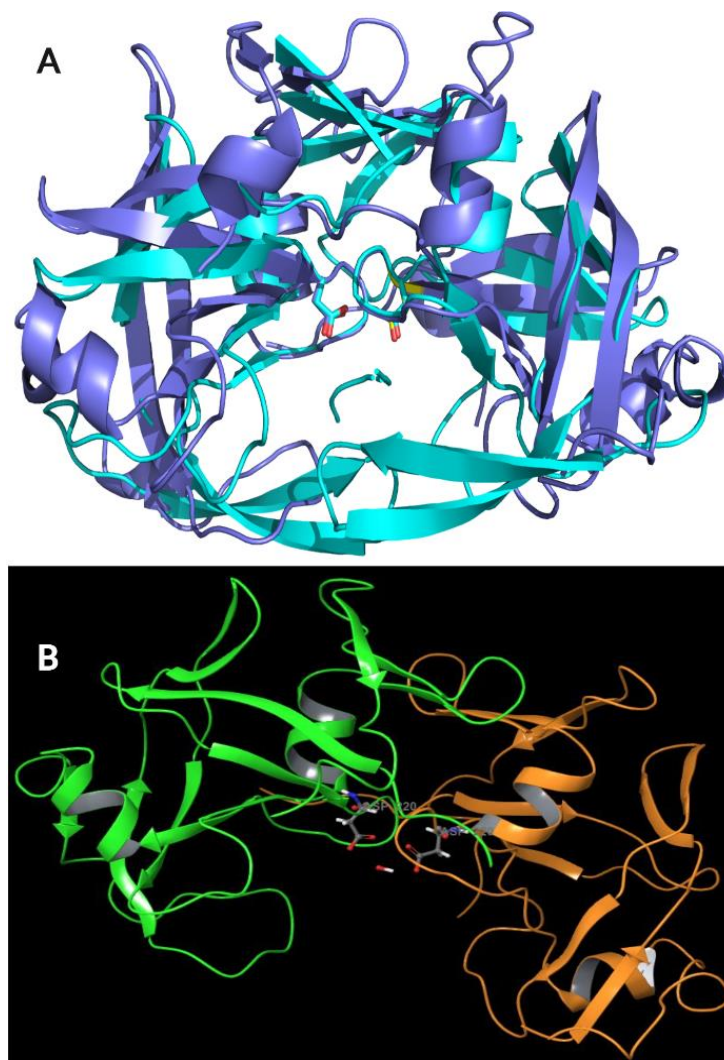
**Figure SI 3.4 The RNA Integrity Number (RIN) values and RNA quality profiles of RNA samples (HAP1 wt, DDI2 CRISPR KO clones c006, c023, and c010) for total-RNA sequencing.**

## Chapter 4 Development of Biochemical assay to target aspartic protease DDI2

### 4.1 Introduction

The human DNA Damage Inducible 1 Homolog II (hDDI2) is a novel aspartic protease whose biological function has not been widely studied and characterized<sup>15-16</sup>. The endoproteolytic activity of hDDI2 is primarily due to its retroviral protease (RVP) domain<sup>17</sup>. The RVP domain is a dimer with two functional aspartic acids forming hydrogen bonds with a water molecule (**Figure 4.1 B**). These aspartic acids are part of a conserved signature motif (D<sub>252</sub>-S<sub>253</sub>-G<sub>254</sub>-A<sub>255</sub>) identical to HIV protease (**Figure 4.1 A**)<sup>16, 88</sup>, making it an exciting target for drug development. Mutation of aspartic acid residues into asparagine at the signature motif hDDI2 (D<sub>252</sub>N) deprives its proteolytic activity of cleaving the ER-resident transcription factor Nrf1, the only known substrate of DDI2<sup>15, 17</sup>. **Apart from the proteolytic function of DDI2 in cells, no substrate-based biochemical assays or small molecule inhibitor-based biochemical assays have been successfully developed to investigate the enzymatic activity of DDI2, making efforts to study the biological function of DDI2 even more difficult.**

As mentioned in **Section 1.7.2**, only one study showed that recombinant Ddi1-like protease could cleave the HIV protease and Cathepsin D substrates at acidic pH<sup>92-93</sup>. This study also showed inhibition of Ddi1-like protease of *L. major* by Nelfinavir (NFV, 20  $\mu$ M), Pepstatin (15  $\mu$ M), and a covalent inhibitor Diazoacetyl-DL-norleucine methyl ester (DAN, 500  $\mu$ M) with a percentage of inhibition 60%, 70%, and 95% respectively<sup>92-93</sup>.



**Figure 4.1** The crystal Structure of DDI2 retroviral-like protease domain (PDB: 4RGH). **A.)** Overlays of DDI2 retroviral-like protease domain (PDB: 4RGH, purple) with HIV retroviral protease domain (PDB: 5HVP, light blue), showing the conserved RVP domain between DDI2 and HIV protease, with slight differences where HIV has beta-sheet flap closing their cavity. DDI2 cavity is considerably larger than HIV. **B.)** The putative active site of DDI2 consists of two dimers, with each chain bearing aspartic acid that can form hydrogen bonds with a water molecule (indicated by balls and sticks). Chain A and Chain B were indicated in different colors. **A)** was generated using PyMOL™ by Goodwin's laboratory, and **B)** was generated using Maestro™ Schrodinger 2019-1.

Unfortunately, we were unable to observe any DDI2 cleavage using our developed FRET substrate DBCYL-VDAWL VHS-Glu (EDANS) based on Nrf1 site cleavage under the optimized conditions (buffer and pH) mentioned in the above study (data are not shown). Supporting our results, Siva *et al.* reported no proteolytic cleavage by DDI2 using different substrates and buffers at pH 5 and 7<sup>16</sup>.

They also could not observe any binding of DDI2 with all HIV protease inhibitors and acetyl-pepstatin<sup>16</sup>. The inability of HIV protease inhibitors to bind with the hDDI2 RVP catalytic active site may be due to the size of the cavity and the opening of hDDI2 compared to HIV protease<sup>16,88</sup>. As shown in **Figure 4.1**, HIV protease has a beta-sheet flap closing its cavity in contrast to hDDI2.

To address the lack of reliable biochemical assays that allow for the visualization of DDI2 targets, we designed the first DDI2 small molecule probe-based biochemical assays. For this purpose, we took advantage of the only available covalent aspartic protease inhibitor Diazoacetyl-DL-norleucine methyl ester (DAN), which gave 95% inhibition of the Ddi1-like protease of *L. major*. In this chapter, we report our first designed DDI2 probe and the remaining challenges for future development.

## **4.2 Results and Discussion**

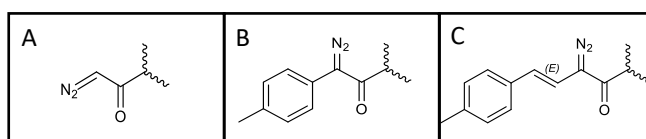
### **4.2.1 Design, *in silico* molecular docking and synthesis of DDI2-probe**

Diazoacetyl-DL-norleucine methyl ester (DAN) was first discovered in 1966 by Rajagopalan *et al.*, where the purified Pepsin, an aspartate-type protease, was rapidly inactivated<sup>144</sup>. The covalent inactivation of pepsin occurred stoichiometrically in the presence of copper ions<sup>144</sup>. Since 1961, these diazo-based non-natural amino acids have been used to inhibit aspartate-type proteases, such as the esterification of carboxylic acids or ribonucleases<sup>144-152</sup>. Thus, we based our DDI2 probe design using a diazo reactive group.

Our initial design focused on the amino acids at the P1 – P1' position that DDI2 protease cleavage sites could recognize. Since DDI2 has an active catalytic site

identical to HIV proteases, it possibly has similar preferences for cleavage sites as HIV protease (**Figure 4.1 A**). HIV proteases prefer to cleave the peptide bond of the substrates after hydrophobic residues (Phe, Tyr, Leu)<sup>153</sup>. Meanwhile, DDI2 cleaves the scissile bond of the substrate between hydrophobic amino acids (Trp and Leu) at the P1 - P1' position<sup>14</sup>. Therefore, we replaced norleucine on DAN with hydrophobic aromatic amino acids (Phe, Tyr, or Trp).

Next, the literature suggested that aliphatic diazo-containing compounds ( $R_1R_2C=N_2$ ) are highly reactive and prone to  $SN_2$  reactions<sup>146, 154</sup>. The  $\alpha$ -carbon protonation of the diazo group produces an intermediate reactive diazonium species that further releases  $N_2$  gas and forms carbenes<sup>146, 154</sup>. Adding aromatic functional groups to the  $\alpha$ -carbon of diazo-containing compounds reduces the reactivity due to the resonance effect, which can stabilize the compounds<sup>146, 154</sup>. We, therefore, modified our designed probe with a stabilized diazo warhead according to the review by Mix *et al.*<sup>146, 155</sup>, as presented in **Figure 4.2**.



**Figure 4.2** List of diazo warhead functional groups reported for aspartic proteases. A.) is a highly reactive diazo functional group. In contrast, the charge on the diazo functional groups of B.) and C.) is distributed through the olefin and phenyl, reducing the reactivity and stabilizing the diazo group.

Moreover, to visualize the binding of DDI2 using western blotting techniques and to allow the pull-down of the protein-bound with DDI2, we added a biotin tag and used lysine as a flexible linker. We conducted *in-silico* covalent docking for each modified step to predict the probe's binding affinity to DDI2 RVP protease. Since the Canonical SMILES of the diazo functional group is not available on Schrodinger Maestro™, to allow us to perform a covalent docking to the alpha

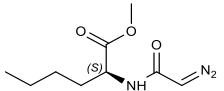
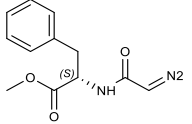
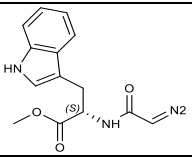
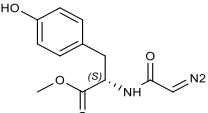
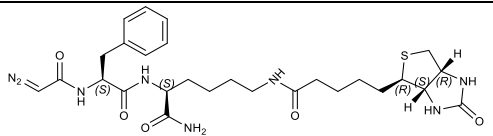
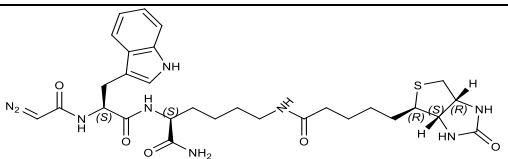
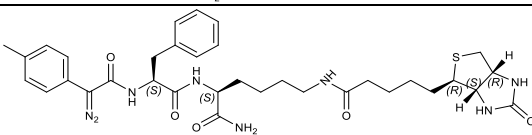
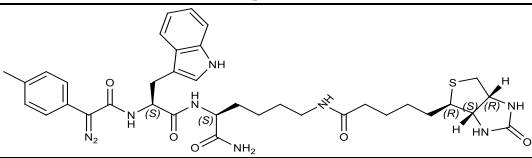
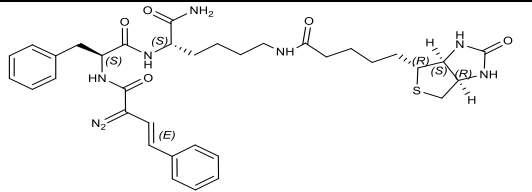
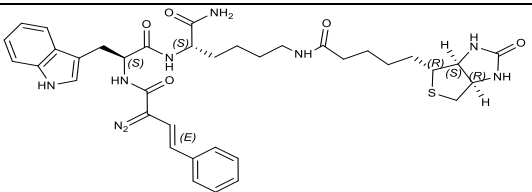
carbon of the probe, we replaced the diazo functional group on the alpha carbon with Halogen (Br). Following this, we performed a nucleophilic substitution reaction on Schrodinger Maestro<sup>TM</sup><sup>156</sup>. To simplify, the more negative the value of the covalent docking score and MM GBSA dg (NS) binding, the stronger the probe binds to the receptor<sup>156</sup>. MM GBSA dg (NS) free energy binding was calculated according to the equation as follows:

**MMGBSA dG Bind(NS) =**  
MMGBSA dG Bind of Receptor-ligand – (Receptor + Ligand).  
All values were obtained from optimized complex without calculating the strain energy or the receptor and ligand conformational changes needed to form the complex.

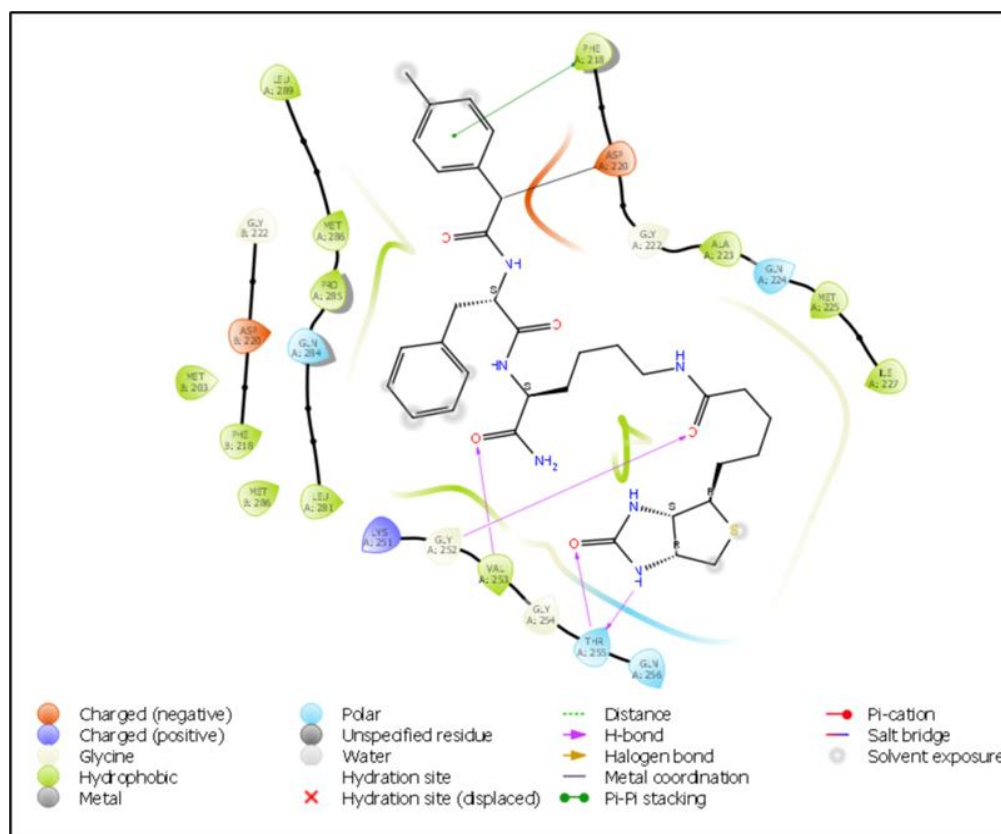
Based on the results summarized in **Table 4.1** and **Supporting Information Table SI 4.1**, Probe No. 7 and 10 gave the highest covalent docking score and MM GBSA dg (NS) binding. Both probes have reduced reactivity in their diazo group due to stabilization through resonance effects and have comparable Lipophilic/Hydrophobic bonds, H-bonds, and  $\pi - \pi$  stacking bonds. Freitas and Schapira's analysis of ligand-to-protein receptor binding interactions in the PDB suggests the three most common types of ligand-protein interaction binding in PDB are H-Bonds, hydrophobic bonds, and  $\pi - \pi$  stacking bonds<sup>157</sup>. The Ligand with Phe residue has more occurrence to engage in  $\pi - \pi$  stacking bonds (50%) than Tyr (36.8%), Trp (8.7%), and His (5.1%)<sup>157</sup>. The Phe residue also has more occurrence to form weak H-bonds CH...O (40-70 %) than Tyr and Trp<sup>157</sup>. Here, we found that probe No.7 provides a slightly better covalent docking score than probe No.10 (**Table 4.1**). Nevertheless, as Freitas and Schapira suggested, we did not observe differences in  $\pi - \pi$  stacking or weak H-bond between Phe and Trp residues. Based

on the ligand-protein interaction diagram, Probe No.7 forms H-bonds and  $\pi - \pi$  stacking to the 4RGH DDI2 receptor at a distance  $< 4\text{\AA}$ , and the covalent bond (see **Figure 4.3**). Furthermore, Probe No.7 has a slightly better MM GBSA dg (NS) binding, which may imply that probe No.7 has better binding affinity than probe 10 (**Table 4.1, Table SI 4.1**). Thus, we chose probe No.7 as our lead compound.

**Table 4.1 Covalent docking analysis for three different modified diazo-reactive functional groups**

No	Structure	Covalent Docking Score	MM GBSA dg Bind (NS) (kcal.mol)
1.		-1.983	-17.79
2.		-1.899	-24.61
3.		-3.328	-27.41
4.		-2.497	-25.69
5.		-4.805	-51.39
6.		-5.650	-69.84
7.		<b>-6.540</b>	<b>-68.04</b>
8.		-5.850	-56.81
9.		-2.897	-68.60
10.		-6.136	-62.40

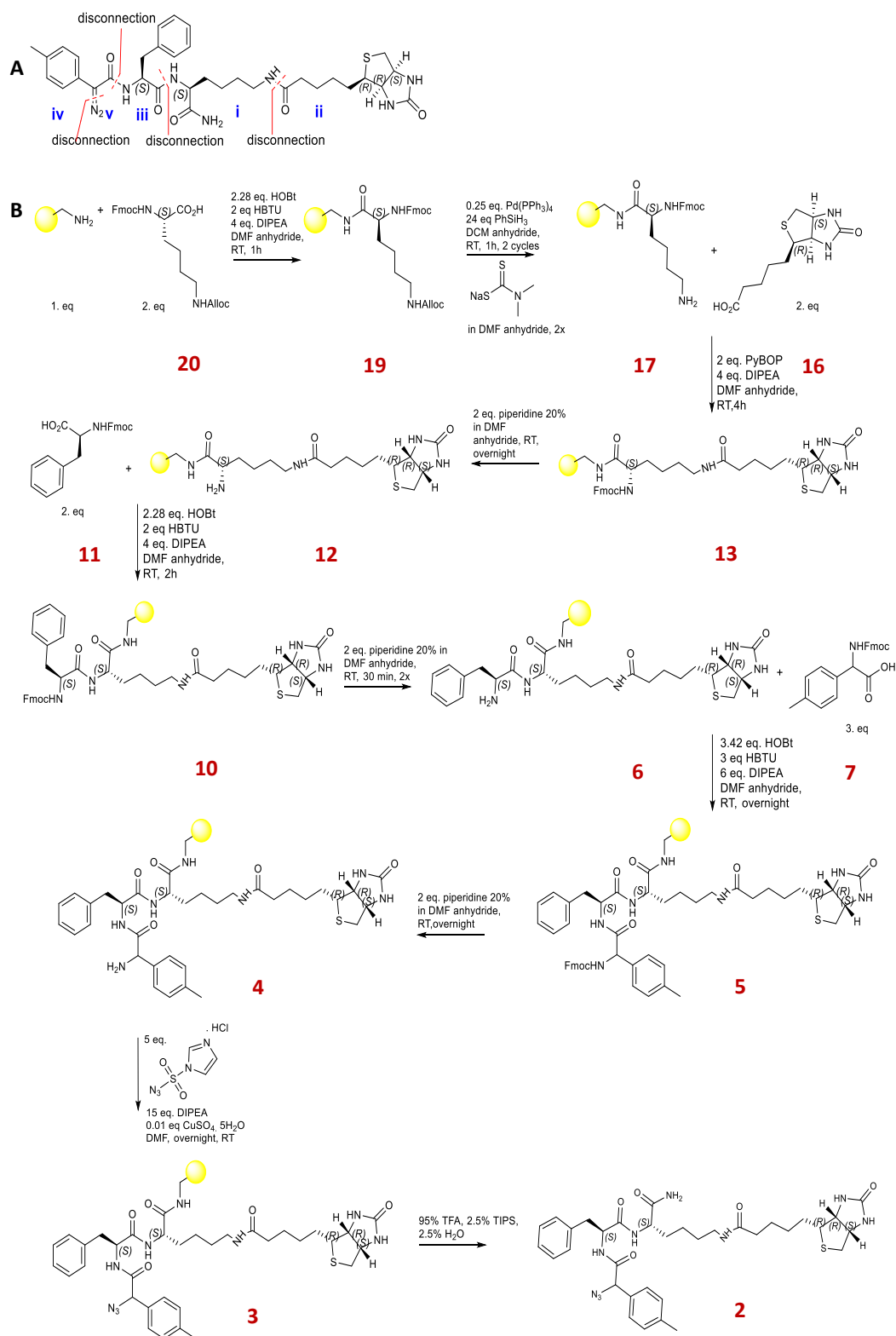




**Figure 4.3** Ligand-to-protein receptor binding interactions of Probe 7 in 4RGH receptor. The covalent binding of A: D220 to  $\alpha$ -carbon of 4-methyl-phenyl glycine.

Solid-phase peptide synthesis (SPPS) of probe No.7 was conducted according to the general setup by Amblard *et al.* (2006)<sup>158</sup>. The assembly of probe No.7 is divided into five steps, as illustrated in **Scheme 4.1 A**. Probe No.7 was synthesized as described in the synthetic route **Scheme 4.1 B**. The colorimetric Kaiser test monitored the completion of each coupling and deprotection step. To this end, the Fmoc-Lys(Alloc)-OH (**20**) was first attached to the activated rink amide resin, followed by the Alloc deprotection (2X) of **19**<sup>159</sup>. Subsequently, D-Biotin (**16**) was coupled to the unprotected NH<sub>2</sub>- $\epsilon$  lysine residue **17**. Then the Fmoc deprotection of NH<sub>2</sub>- $\alpha$ -Lysine **13** was conducted to obtain **12**. Subsequent Fmoc-N- $\alpha$ -Phenylalanine-OH (**11**) was recoupled to **12**. The Fmoc protecting group was removed from **10**, and compound **6** was obtained. Next, 4-methyl-N-Fmoc- $\alpha$ -phenyl glycine-OH (**7**) was coupled to **6** to generate **5**. The last step of the SPPS

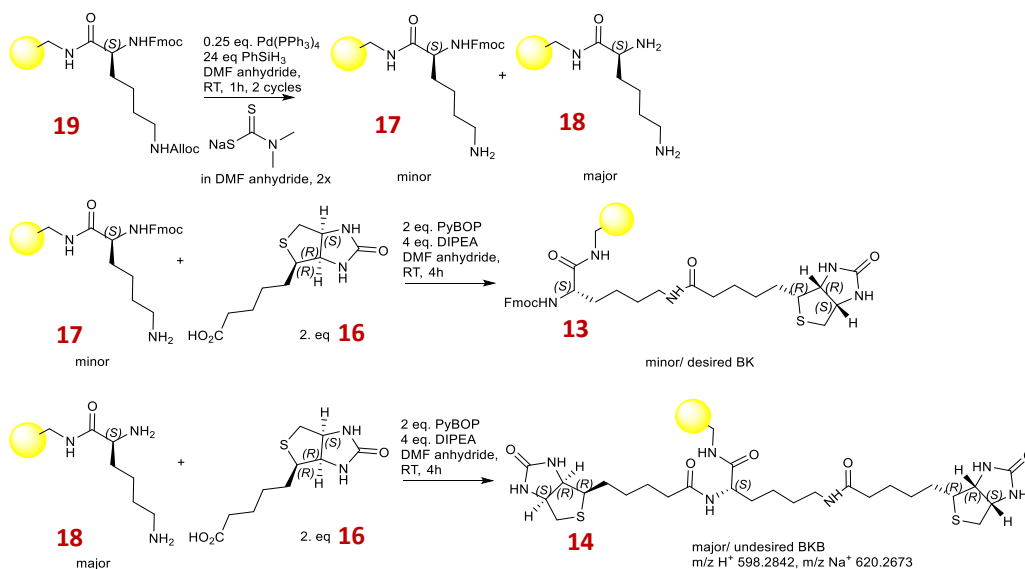
was Fmoc deprotection of **5**, followed by on-resin azide transfer using imidazole-1-sulfonyl azide as an azide donor by converting **4** to **3**<sup>160-162</sup>. For the final step, a solution-based phosphine-mediated reaction through the intermediary of acyl triazenes was designed to convert the azide into the diazo compound, as previously described<sup>146, 163-164</sup>.



**Scheme 4.1** The Disconnection approach (A) and synthetic route (B) of Probe No.7.

However, when probe No.7 was synthesized according to **Scheme 4.1 B**, compound **3** with  $m/z$   $[M+H]^+$  691.33 was obtained as a minor product. At the same time, the

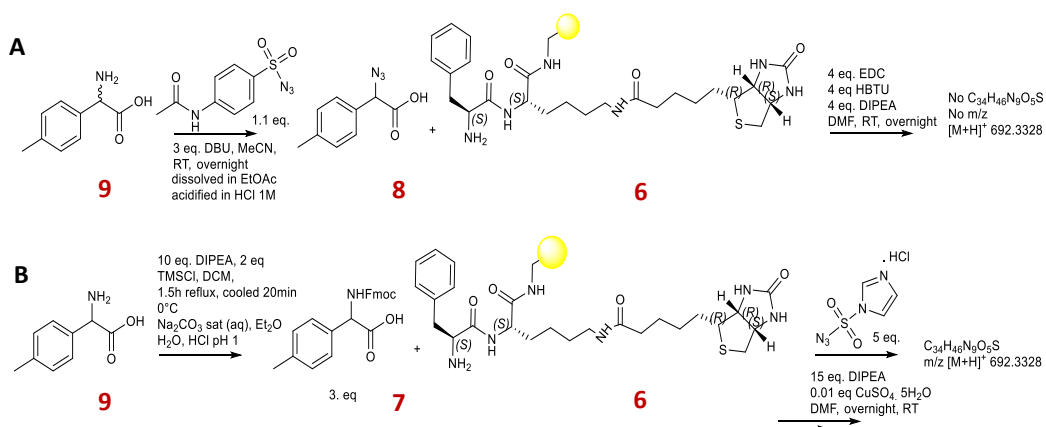
undesirable products ( $m/z$   $[M+H]^+$  598.28,  $m/z$   $[M+Na]^+$  620.26, and  $m/z$   $[M+K]^+$  636.24) were observed as main products (see **Supporting Information Figure SI 4.1**). Based on its mass-per-charge ratio, the unwanted product **14** may occur due to the polymerization of D-Biotin (**16**) on both  $NH_2$ - $\alpha$  and  $NH_2$ - $\epsilon$  of lysine **18** that may arise after Alloc deprotection from  $NH_2$ - $\epsilon$ -lysine residues (**Scheme 4.2**). The free  $NH_2$ - $\epsilon$ -lysine residue (**18**,  $pK_a \sim 10$ ) then causes an increase in the basicity within the reaction vessels, which results in partial deprotection of Fmoc- $N$ - $\alpha$ -lysine. This may result in two products being formed, the desired **13** and the side product **14**. Supporting this, studies found that  $NH_2$ - $\epsilon$ -lysine deprotection also led to the unwanted removal of the Fmoc- $N$ - $\alpha$ -lysine protecting group<sup>165-166</sup>. The Fmoc removal occurred regardless of the protecting group, whether it was 4-methyl trityl (Mtt) or allyloxy carbonyl (Alloc) type, and the deprotection method used<sup>165-166</sup>. Thus, it was suggested to carry out the deprotection of Alloc in the presence of subsequent D-Biotin to be coupled in the peptide sequence. Nevertheless, this approach is not feasible in our case since it seems that the Fmoc deprotection occurred almost as fast as the Alloc deprotection. Hence, we decided to start and attach  $N$ - $\alpha$ -Fmoc- $N$ - $\epsilon$ -biotinyl-L-lysine to the resin (**13**).



**Scheme 4.2** Possible explanation of the formation of undesirable  $m/z$   $[M+H]^+$  598.28 or  $m/z$   $[M+Na]^+$  620.26 may be due to the partial removal of Fmoc-*N*- $\alpha$ -lysine following Alloc deprotection.

Furthermore, the Incorporation of  $\alpha$ -azido-4(methyl)-phenyl glycine into Probe No. 7 was designed using two approaches as outlined in **Scheme 4.3**. First, an azide transfer was carried out to convert (**9**) into **8**. For this purpose, *p*-acetamidobenzenesulfonyl azide (*p*-ABSA) was used along with a non-nucleophilic base, 1,8-Diazabicyclo(5.4.0)undec-7-ene (DBU, pH 12.8), as previously reported by Mix *et al.*<sup>167</sup>. Compound **8** was validated by HRMS-ESI calculated for C<sub>9</sub>H<sub>9</sub>N<sub>3</sub>O<sub>2</sub>,  $m/z$   $[M-H]^-$  190.0614 (see **Supporting Information Figure SI 4.2**). Since compound **8** has been previously reported in the literature, the NMR spectra of compound **8** were not obtained<sup>167</sup>. Next, the coupling of compound **8** with compound **6** was attempted, but this was not successful (**Scheme 4.3 A**). EDC-HOBt or EDC-HCTU coupling reagents are often used in SPPS strategies. The protonated EDC is used to activate the  $\alpha$ -COOH group of **8** into an O-acylisourea active intermediate<sup>168</sup> before the coupling reaction, thus allowing the free amino group of peptide **6** to attack the activated  $\alpha$ -COOH ester. At the same time, the less reactive

coupling reagents HBTU or HOBt act as additives to carbodiimide-based coupling reagents to avoid the formation of inactive N-acylisourea<sup>168</sup>. The latter may appear after a rapid rearrangement of O-acylisourea in DMF in excess of EDC<sup>168</sup>. In addition, HBTU itself can be coupled directly with  $\alpha$ -NH<sub>2</sub> of amino acids that form the unwanted guanidine by-product and terminate the formation of amide bonds<sup>168</sup>. This may occur due to a slow  $\alpha$ -COOH pre-activation or excessive use of HBTU<sup>168</sup>. Based on this, there are several possibilities why approach A does not work, such as excessive use of EDC renders the formation of inactive N-acylisourea or excessive use of HBTU that generates guanidine by-products.



**Scheme 4.3** Synthetic route from **9** to **8** and **7**

For the second approach described in **Scheme 4.3 B**, the Fmoc protection of compound (**9**) was carried out to give **7** according to the method described by Liebeschuetz *et al.*, US patent number US20020055522<sup>169</sup>, with a slight change in the starting material from 4(methyl)-phenylphenyl glycine to 4(methyl)-phenyl glycine (**9**). Compound **7** was validated by HRMS-ESI calculated for C<sub>24</sub>H<sub>21</sub>NO<sub>4</sub> 387.1549, m/z [M+H]<sup>+</sup> 388.1514, m/z [M+Na]<sup>+</sup> 410.1364, m/z [M+K]<sup>+</sup> 426.1048 (**Figure 4.4**). The <sup>1</sup>H NMR and <sup>13</sup>C NMR spectra in DMSO-d<sub>6</sub> (**Figure 4.5**) were also obtained since compound **7** was not reported. Based on <sup>1</sup>H NMR spectra in **Figure 4.5 A** and **Supporting Information Figure SI 4.3**, the N- $\alpha$ -Fmoc-4(methyl)phenyl

glycine can be distinguished according to  $^1\text{H-NMR}$  of the Ar- $\text{CH}_3$  (s,  $J=1.0$  Hz, 3H, 2.31ppm, #7), R-NH=CO (s, 1H, 5.75ppm, #9), R-COOH (s, 1H, 12.81ppm, broad peak #30), and Ar-CH(COOH)-NHCOOR (s,  $J=1.0$ Hz, 1H, 6.3ppm, #8). Meanwhile, the  $^{13}\text{C}$  NMR chemical shift showed spectra at 172.62 for R-COOH, 47.10 and 66.41 for C-H of Fmoc, 58.29 for Ar-CH(COOH)-NHCOOR, and 21.15 for Ar- $\text{CH}_3$ .

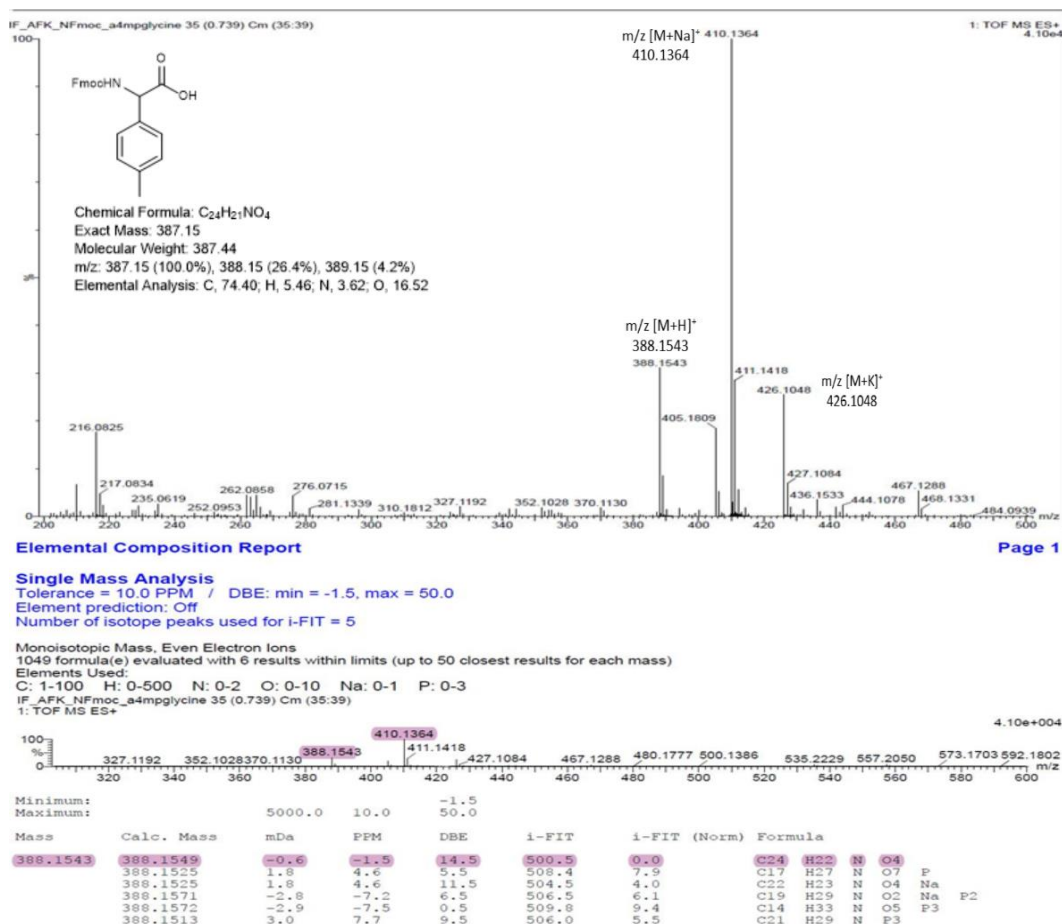
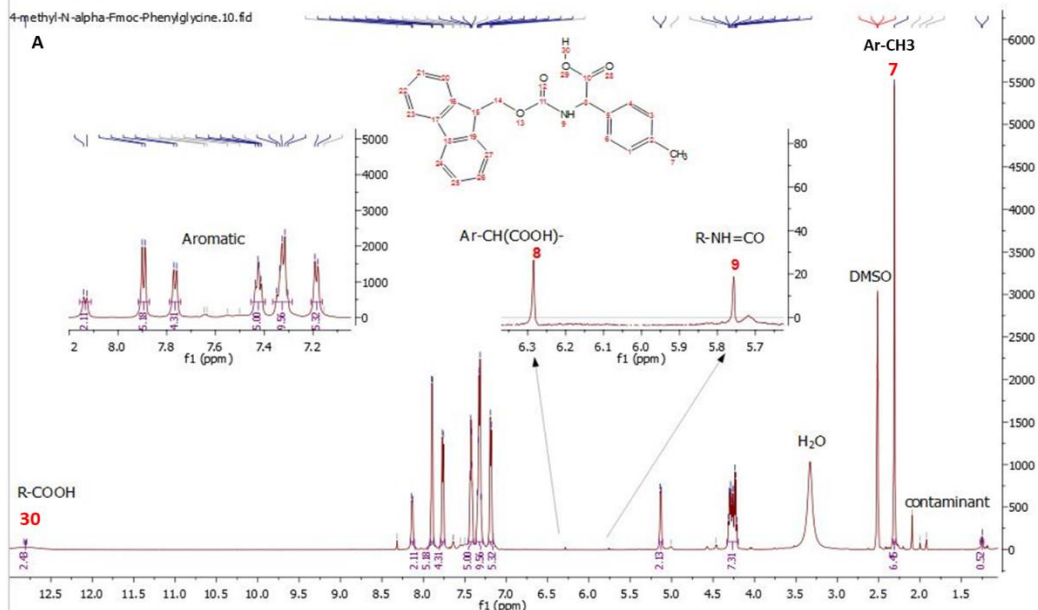
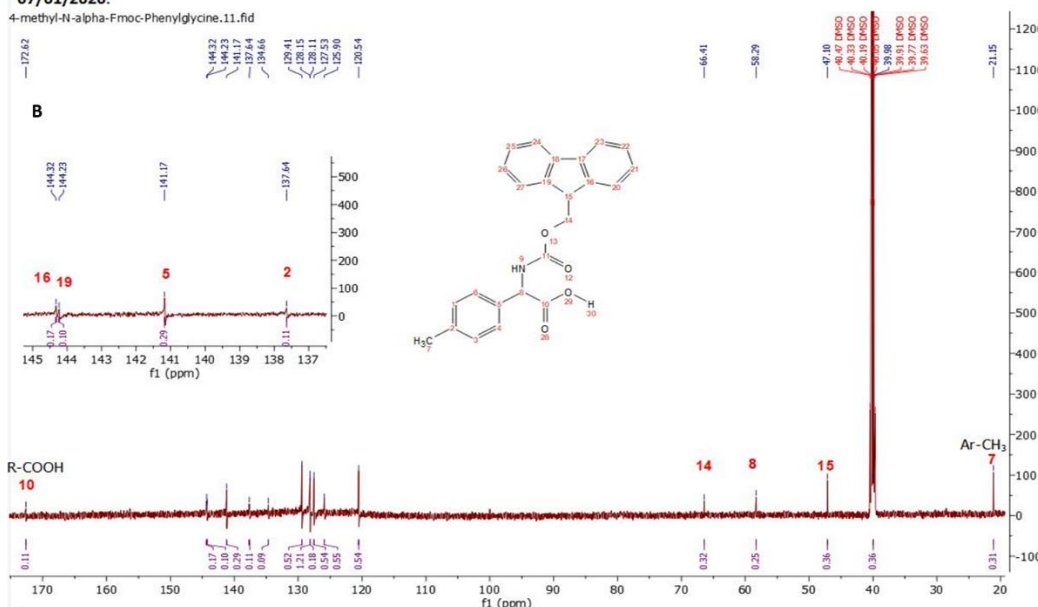


Figure 4.4 The LC-MS Spectra and HRMS-ESI of *N*- $\alpha$ -Fmoc-4(methyl)phenyl glycine compound (7) show  $m/z$   $[\text{M}+\text{H}]^+$  388.1543,  $m/z$   $[\text{M}+\text{Na}]^+$  410.1364, and  $m/z$   $[\text{M}+\text{K}]^+$  426.1048. According to HRMS-ESI, these masses corresponded to  $\text{C}_{24}\text{H}_{21}\text{NO}_4$ , calculated for 388.1549 and found as 388.1543.

$^1\text{H}$  NMR (600 MHz,  $\text{DMSO}-d_6$ )  $\delta$  12.81 (s, 1H), 8.13 (d,  $J = 8.0$  Hz, 2H), 7.89 (d,  $J = 7.6$  Hz, 5H), 7.77 (d,  $J = 7.5$  Hz, 4H), 7.45–7.40 (m, 5H), 7.33 (dd,  $J = 12.6, 7.5$  Hz, 10H), 7.19 (d,  $J = 7.7$  Hz, 5H), 6.3 (s,  $J = 1.0$  Hz, 1H), 5.75 (s, 1H), 5.13 (d,  $J = 8.0$  Hz, 2H), 4.33–4.20 (m, 6.0Hz, 7H), 2.31 (s,  $J = 1.0$  Hz, 3H), (((9H-fluoren-9-ylmethoxy)carbonyl)amino)(4-methylphenyl)acetic acid,  $^1\text{H}$  NMR data analyzed on 07/01/2020.



$^{13}\text{C}$  NMR (151 MHz,  $\text{DMSO}-d_6$ )  $\delta$  172.62, 144.32, 144.23, 141.17, 137.64, 134.66, 129.41, 128.15, 128.11, 127.53, 125.90, 120.54, 66.41, 58.29, 47.10, 39.98, 21.15. (((9H-fluoren-9-ylmethoxy)carbonyl)amino)(4-methylphenyl)acetic acid,  $^{13}\text{C}$  NMR data analyzed on 07/01/2020.



**Figure 4.5** The  $^1\text{H}$  NMR and  $^{13}\text{C}$  NMR chemical shift ( $\delta$ , ppm) of *N*- $\alpha$ -Fmoc-4(methyl)phenyl glycine in  $\text{DMSO}-d_6$ . All NMR spectra were analyzed using MestReNova x64 Mestrelab™ from the experimental NMR fid. archives.

Next, compound 7 was coupled with compound 6 using HOBt and HBTU coupling reagents. Here, HBTU activates the  $\alpha$ -COOH of 7, allowing the rapid formation of amide bonds with 6 to give compound 5. At the same time, HOBt is used to avoid epimerization and accelerate the coupling reaction. The Fmoc protecting group was



removed to give compound **4** to allow on-resin azide transfer. During the SPPS reaction, the crude reaction was sampled following each deprotection and coupling step to obtain HRMS-ESI of compounds **6**, **5**, and **4**, as indicated in the supporting information **Figure SI 4.4**. HRMS-ESI of these compounds gave  $m/z$   $[M+H]^+$  of 519.2767, 888.4150, and 666.3497, consistent with the structures. Peptide **4** was converted into **3** on polystyrene-based resin utilizing imidazole-1-sulfonyl azide HCl (ISA-HCl) as an azide donor according to the optimized method by Castro *et al.*<sup>160</sup>. This method was reported suitable for Aminomethyl-polystyrene-based resin (AMPS, rink amide resin)<sup>160</sup>. The Azide transfer was carried out in three different solvents as follows CH<sub>3</sub>CN, a mixture of CH<sub>3</sub>CN: CH<sub>2</sub>Cl<sub>2</sub> 1:1, and DMF. For safety reasons, DMF was chosen as the solvent for on-resin azide transfer to avoid excessive gas formation (gas pressure buildup inside the vessel) during vigorous overnight SPPS shaking. The crude reaction azido peptide **3** was sampled for the Kaiser test and LC-MS profile, including HRMS-ESI. Based on this LC-MS profile, the highest chromatogram peak contains  $m/z$   $[M+H]^+$  692.3358 – 692.3369 at RT 6.21 – 6.25, as shown in the **Supporting Information Figure SI 4.5**. The formation of the azido peptide **2** was validated by HRMS-ESI calculated for C<sub>34</sub>H<sub>45</sub>N<sub>9</sub>O<sub>5</sub>S and observed  $m/z$   $[M+H]^+$  692.3340, which was consistent with the structure.

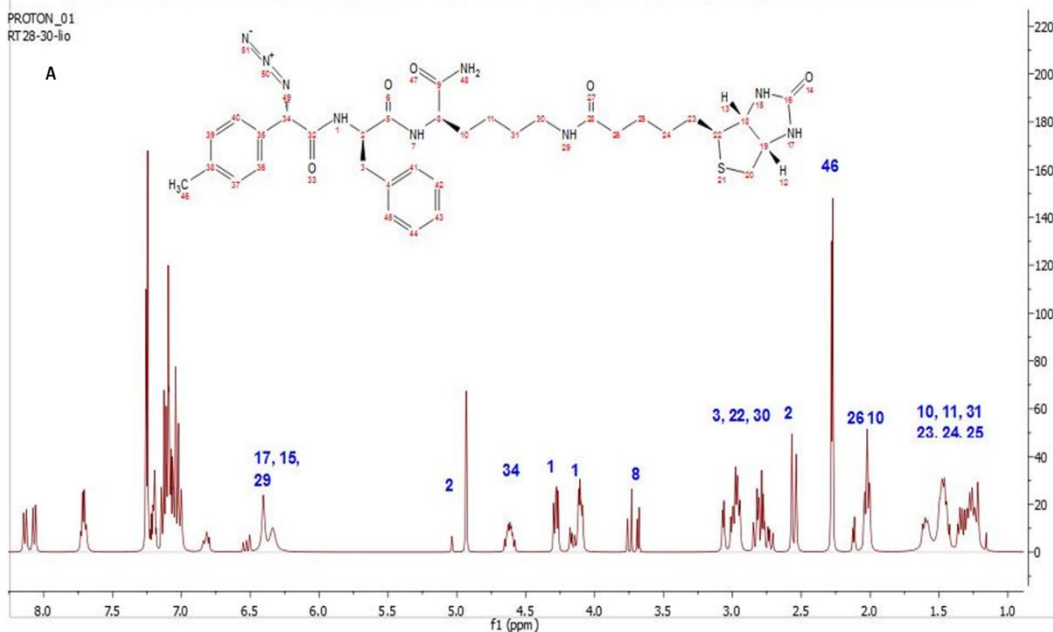
The Semi-preparative LC was conducted to purify peptide **2**, as shown in **Supporting Information Figure SI 4.6**. HRMS-ESI data showed only two peaks containing masses of  $m/z$   $[M+H]^+$  692.3354 and 692.3333 calculated for C<sub>34</sub>H<sub>45</sub>N<sub>9</sub>O<sub>5</sub>S, as shown in the **Supporting Information Figure SI 4.7**. The purified azido peptide **2** was then lyophilized and recharacterized by obtaining experimental <sup>1</sup>H NMR spectra (**Figure 4.6 A**). The experimental <sup>1</sup>H NMR was compared to the <sup>1</sup>H NMR

prediction (**Figure 4.6 A and B**) and additional references of Biotinylated lysine experimental  $^1\text{H}$  NMR as described elsewhere <sup>170</sup>.

In addition, azido peptide **2** was revalidated by HRMS-ESI calculated for  $\text{C}_{34}\text{H}_{45}\text{N}_9\text{O}_5\text{S}$  and was found as  $m/z$   $[\text{M}+\text{H}]^+$  692.3334 (**Figure 4.7 A**). This azido peptide remains stable when it is stored in DMSO (**S1**) or its lyophilized form (**S2**), as shown in the overlapping chromatogram of (**S1**) and (**S2**) in **Figure 4.7 B – C and Supporting Information Figure SI 4.8**.

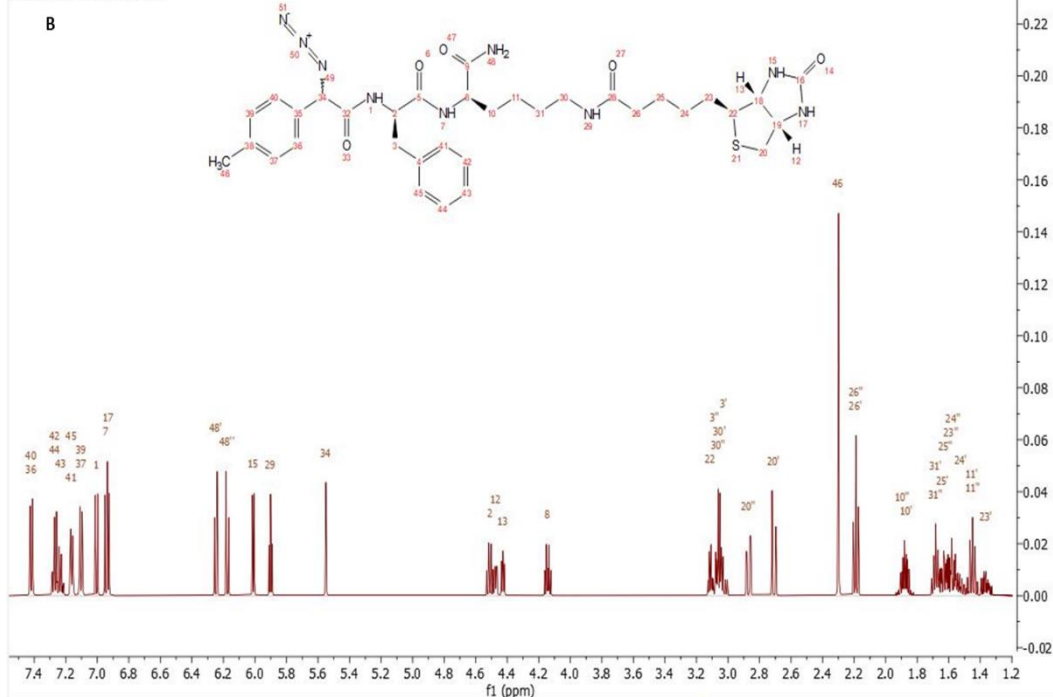
**<sup>1</sup>H NMR (399 MHz, DMSO-d<sub>6</sub>)**  $\delta$  8.46 (t, J = 8.6 Hz, 1H), 8.10 (dd, J = 26.7, 8.1 Hz, 1H), 7.71 (q, J = 5.5 Hz, 1H), 7.28 – 7.15 (m, 2H), 7.17 – 7.09 (m, 1H), 7.13 – 6.97 (m, 3H), 6.40 (s, 1H, H<sub>17</sub>), 6.34 (s, 1H, H<sub>15</sub>), 4.93 (d, J = 2.1 Hz, 1H), 4.68 – 4.55 (m, 1H), 4.28 (dd, J = 7.7, 4.9 Hz, 1H), 4.30 – 4.25 (m, 1H, H<sub>12</sub>), 4.20 – 4.06 (m, 1H, H<sub>13</sub>), 3.8 – 3.6 (m, 1H, H<sub>8</sub>), 3.08 – 3.05 (m, 2H, H<sub>22</sub>) 3.02 – 2.93 (m, 1H, H<sub>3</sub>, H<sub>30</sub>), 2.98 (dt, J = 12.9, 6.9 Hz, 1H, H<sub>20</sub>), 2.87 – 2.68 (m, 1H), 2.55 (d, J = 12.4 Hz, 1H, H<sub>20</sub>), 2.28 (s, J = 3.7 Hz, 3H, H<sub>46</sub>), 2.13 – 2.10 (m, 2H, H<sub>26</sub>), 2.13 – 1.97 (td, 2H, J = 7.4, 2.0 Hz, H<sub>10</sub>, H<sub>26</sub>), 1.65 – 1.54 (m, 2H, H<sub>10</sub>), 1.54 – 1.40 (m, 6H, H<sub>11</sub>, H<sub>23</sub>, H<sub>25</sub>), 1.38 – 1.12 (m, 4H, H<sub>24</sub>, H<sub>31</sub>). <sup>1</sup>H NMR data of compound 2 analyzed 10/27/2020.

**IUPAC:** 3-(2-(((2R)-1-((2R)-1-amino-1-oxo-6-((5-((3aR,4S,6aS)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl) pentanoyl) amino) hexan-2-yl) amino)-1-oxo-3-phenylpropan-2-yl)amino)-1-(4-methylphenyl)-2-oxoethyl triaza-1,2-dien-2-ium-1-ide

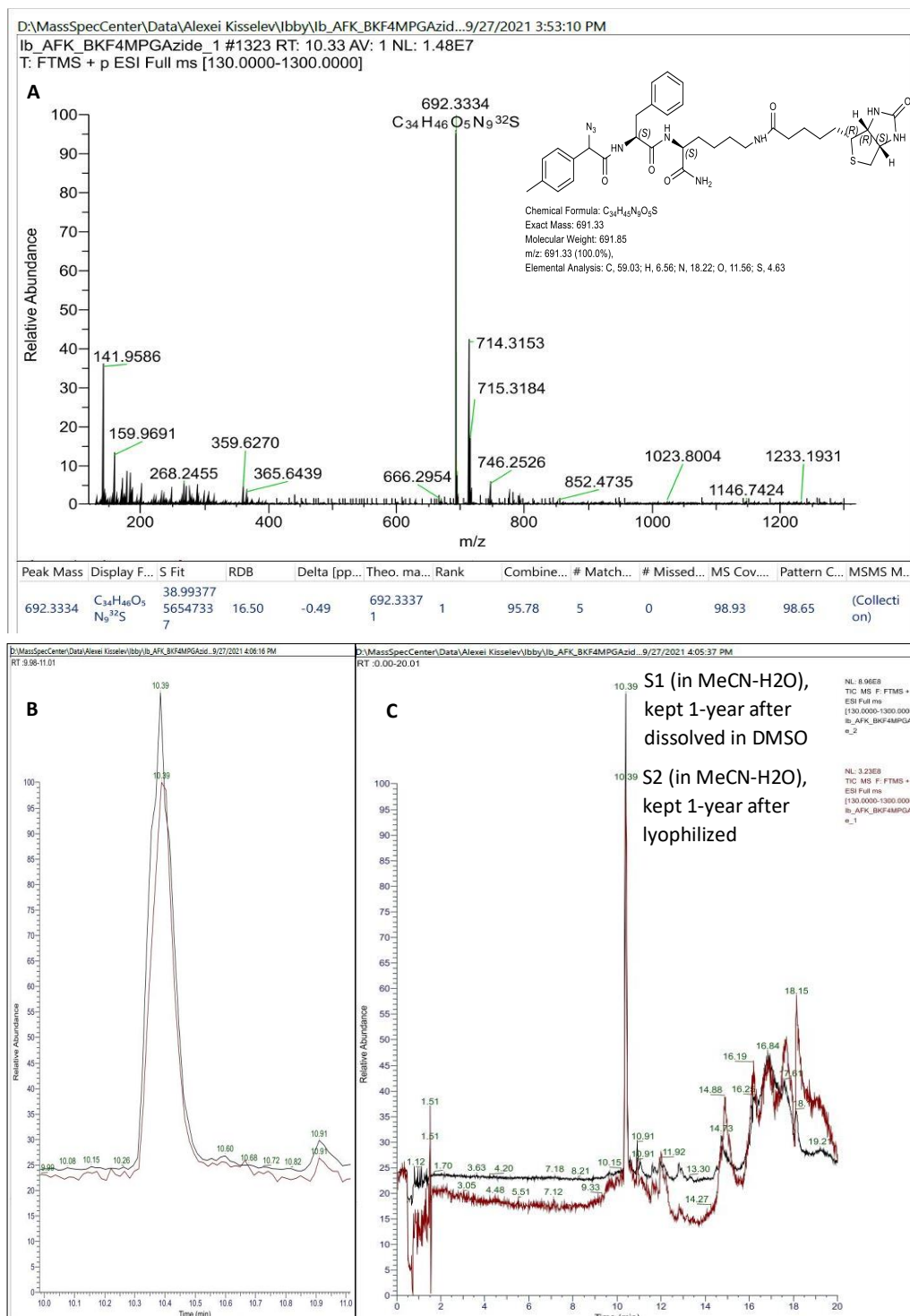


12.4, 3.3 Hz, 1H), 2.74 – 2.66 (m, 1H), 2.35 (td, J = 12.3, 3.5 Hz, 1H), 2.29 – 2.19 (m, 2H), 1.91 – 1.71 (m, 2H), 1.69 – 1.52 (m, 2H), 1.53 – 1.39 (m, 1H), 1.34 – 1.10 (m, 2H).

Predicted <sup>1</sup>H NMR Spectrum



**Figure 4.6** The <sup>1</sup>H NMR chemical shift ( $\delta$ , ppm) of azido peptide 2 in DMSO-d<sub>6</sub>. Analysis of experimental <sup>1</sup>H NMR spectra (A) from the NMR fid.file was compared to predicted (B) based on the chemical structure. (A) and (B) were analyzed using MestReNova x64 Mestrelab™.



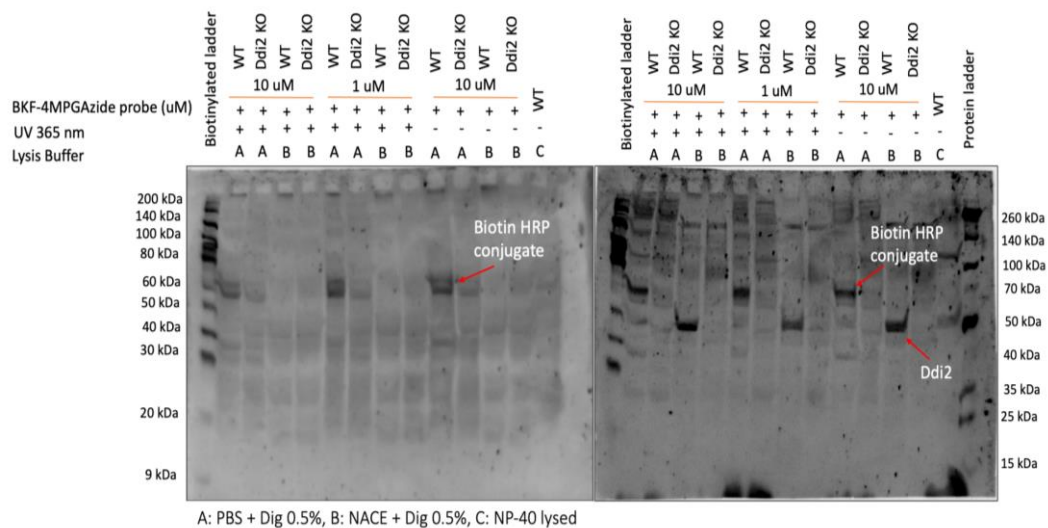
**Figure 4.7 LC-MS and HRMS-ESI of purified azido peptide 2.** A.) HRMS-ESI was calculated for  $C_{34}H_{45}N_9O_5S$  692.3337 and found  $m/z$   $[M+H]^+$  692.3340. Overlays of purified and lyophilized azido peptide 2 S1 and S2 showed one major chromatogram peak at RT 10.39 Minutes. B.) zoom scale of chromatogram peak S1 and S2 at RT 10.39 min. C.) chromatography profile of S1 and S2 suggesting both compounds remain stable after a year. S1 was dissolved in DMSO before being diluted in MeCN: H<sub>2</sub>O (3:1) for LC-MS analysis. While the lyophilized S2 was dissolved in MeCN: H<sub>2</sub>O (3:1) before LC-MS analysis.

#### 4.2.2 Biochemical assay for azido peptide 2

Up to this point, we became interested in using azido peptide **2** directly as a photolabeling probe. We reasoned that UV photo-conjugation of azides-containing compounds could form singlet nitrene reactive intermediate and liberate N<sub>2</sub> gas. Nitrene is an aza-analog of carbene that can be generated following UV photo-conjugation of the diazo-containing compound. Therefore, we used this azido peptide **2** in our biochemical assay. We performed cross-linking as described in the experimental **Section 4.3.4**. Based on **Figure 4.8**, Azido peptide **2** can bind with HAP1 cells without photo-conjugation under UV 365nm, suggesting that the binding is not due to nitrene formation.

Furthermore, the bioconjugation only occurred in HAP1 cells lysed in PBS + Digitonin 0.5% buffer but not in cells lysed in Na-acetate + Digitonin 0.5% buffer, suggesting that the bioconjugation occurred at neutral pH and not at acidic pH. Meanwhile, acidic pH of 5 is the preferred pH condition for aspartic protease Ddi1-like protein<sup>92</sup>. Although the total charge of the protein should be the determinant in the cell lysate, at neutral pH, aspartic acid mainly exists in the ionized form. As previously mentioned, DDI2 is a dimer with a distinctive motif similar to HIV protease. This signature motif contains aspartic acid, which must be present as a non-ionized (COOH) and an ionized (COO<sup>-</sup>) form to remain active. Most importantly, Azido peptide **2** only binds to HAP1 wt but not with DDI2 KO c010 cells which suggests that Azido peptide **2** can bind directly to DDI2 or to a protein that interacts with DDI2 through an unknown mechanism.

Azido peptide **2** is designed with biotin as a labeling method. Using Streptavidin-HRP conjugate to detect peptide 2-labeled proteins will also identify endogenous biotinylated proteins<sup>171</sup>. Biotin plays a role as a coenzyme in carboxylase enzymatic reactions and is essential in oxidative metabolism for lipogenesis and gluconeogenesis<sup>171-172</sup>. The Biotin and ATP-dependent carboxylases include cytoplasmic acetyl-CoA carboxylase 1, mitochondrial acetyl-CoA carboxylase 2, pyruvate carboxylase, propionyl-CoA carboxylase, and 3-methylcrotonyl-CoA carboxylase<sup>171-172</sup>. A study reported probing with biotin showed the molecular weight of the mitochondrial acetyl-CoA bands at 75 kDa (containing propionyl-CoA carboxylase and methylcrotonyl-CoA carboxylase) and 130 kDa (containing pyruvate carboxylase). Meanwhile, the cytosolic acetyl-CoA showed the band at 190 and 270-kDa, representing acetyl-CoA carboxylase<sup>173</sup>. However, we could see DDI2 labeling after probing with biotin-HRP conjugates (**Figure 4.8, left**), followed by DDI2 antibodies (**Figure 4.8, right**) at 60 kDa in PBS lysis buffer. Meanwhile, the unlabeled HAP1 cells, lysed in Na-acetate buffer (pH 5.5), showed only bands at 50 kDa with DDI2 antibody but not with Biotin-HRP-conjugate antibody. These observations may indicate that DDI2 is indeed labeled by the Azido peptide **2**, not because it binds with biotinylation-associated binding proteins carboxylase. We reasoned that if we see the endogenous biotinylation-associated binding proteins carboxylase in the blot, the smallest MW is supposed to be 75kDa. Yet, the MW on the blot following probing with Biotin-HRP-conjugate antibody is 60kDa. To prove whether this is not false positive binding, we can also perform DDI2 pull-down or affinity column for future study. Both of these approaches were reported in the identification of false-positive biotin labeling<sup>171</sup>.



**Figure 4.8 Bioconjugation of azido peptide 2 in HAP1 cells.** HAP1 wt and DDI2 KO c010 were lysed in three different lysis buffers (A) PBS containing digitonin 0.5%, (B) Na-acetate containing digitonin 0.5%, and (C) NP-40 lysis buffer. Two concentrations of azido peptide were used in the assay 1 and 10  $\mu$ M. The photoconjugation was carried out under UV 365nm. The blot was first probed with biotin HRP-conjugate, followed by DDI2. HAP1 wt cells lysed in NP-40 serve as negative control (lysis buffer containing DTT).

If DDI2 reacts with azido peptide 2, what is the mechanism of the reaction? The molecular weight of DDI2 is about 50 kDa. Probing with Biotin HRP-conjugate and DDI2 antibodies showed an increase in molecular weight to 60 kDa. Thus, it may indicate that the binding of azido peptide 2 to DDI2 occurs through protein-protein interactions but not with carboxylase. A study reported that azide (imidazole-1-sulfonyl azide hydrochloride) could be introduced to chemically label lysine of the side chain residue or the N-terminus of the enzyme<sup>174</sup>. Additionally, the azide functional group has been reported as one of the metabolic chemical reporters for labeling glycosylated proteins and cell-surface glycans<sup>175-176</sup>. Although we do not have evidence of whether there is an interaction of DDI2 with cell surface glycans or glycosylated proteins, based on transcriptome analysis in **Chapter 3, Table 3.5**, we discovered that DDI2 knockout reduced the expression of genes in most metabolic pathways, including hexosamine salvage pathways. Again, to answer the question of what proteins interact with DDI2 and azido peptide

**2**, we propose to perform a pulldown assay for our future study. To test whether azido peptide **2** can bind with the active catalytic site of DDI2, we should experiment using purified DDI2 and DDI2 mutant where the aspartic acid is mutated.

To summarize, we initially performed photoconjugation of azido peptide **2** in HAP1 cells. However, we found that azido peptide **2** can be bioconjugated with DDI2 in the absence of photoconjugation and showed an increase in the MW to 60kDa. This raises the question of whether azido peptide **2** binds directly to the DDI2 catalytic active site or through protein-protein interaction. Therefore, conducting experiments on purified DDI2 and DDI2 mutants is necessary. As we advance, if azido peptide **2** does not bind with DDI2 catalytic active site, then it may be worth trying to perform the next synthesis step to see if diazo can be used to recognize the active catalytic site of DDI2. This proposed experiment must first be carried out using purified DDI2.

### **4.3 Material and methods**

#### **4.3.1 Chemicals**

Unless stated otherwise, we obtained all organic solvents (LC-MS grade, technical, or anhydrate), acids, bases, and buffer salts through VWR life sciences.



**Table 4.2 List of chemicals used for the peptide synthesis.**

No.	CAS #	Chemical name	Chemical formula	MW (g/mol)	Source
1	145069-56-3	Rink Amide Resin 0.3 - 0.8 meq/g, 200 - 400 mesh, 1% DVB 4-(2'-4'-dimethoxyphenyl)-Fmoc-aminomethyl)phenoxy resin 0.8 mmol/g	C32H29NO7	539.58	CHEM-IMPEX INT'L INC.
2	146982-27-6	N(α)-fmoc-N(ε)-alloc-L-lysine or Fmoc-Lys(alloc)OH	C25H28N2O6	452.51	CREOSALUS, Inc.
3	58-85-5	5-((3aS,4S,6aR)-2-Oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanoic acid or D-(+)-Biotin	C <sub>10</sub> H <sub>16</sub> N <sub>2</sub> O <sub>3</sub> S	244.31	VWR liife science
4	146987-10-2	F-moc-Lys(Biotin)-OH, Nα-Fmoc-Nε-biotinyl-L-lysine	C31H38N4O6S	594.72	AmBeed
5	94750-37-1	O-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyl uronium hexa-fluoro phosphate (HBTU)	C11H16F6N5OP	379.25	CHEM-IMPEX INT'L INC.
6	80029-43-2	1-Hydroxy benzotriazole monohydrate (HOBt)	C6H5N3O.H2O	135.13	TCI America
7	7087-68-5	N,N' Diisopropylethylamine (DIPEA)	C8H19N	129.25	CHEM-IMPEX INT'L INC.
8	14221-01-3	Tetrakis (triphenylphosphine) palladium (0) Pd(PPh3) <sub>4</sub>	[(C6H5) <sub>3</sub> P] <sub>4</sub> Pd	1155.58	CHEM-IMPEX INT'L INC.
9	694-53-1	Phenylsilane	C6H8Si	108.22	TCI America
10	128-04-1	Sodium dimethyl dithiocarbamate dihydrate (DTC)	C3H6NNaS2.2H2O	143.2	TCI America
11	11024-24-1	Digitonin, water soluble	C56H92O29	1229.32	MP Biomedicals

No.	CAS #	Chemical name	Chemical formula	MW (g/mol)	Source
12	128625-52-5	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOPP)	C18H28N6OP.F6P	520.34	CHEM-IMPEX INT'L INC.
13	35661-40-6	Fmoc-L-Phenylalanine	C24H21N2O4	387.43	CHEM-IMPEX INT'L INC.
14	952234-36-5	1H-Imidazole-1-sulfonyl azide hydrochloride (ISA-HCl)	C3H4ClN5O2S	209.62	Enamine, Ltd.
15	13227-01-51	2-amino-2-(4-methylphenyl)acetic acid, DL-4-Methylphenylglycine, 4-Methylphenyl glycine	C9H11NO2	165.19	Matrix Scientific
16	75-77-4	Chlorotrimethyl silane (TMS-Cl)	C3H9ClSi	108.64	Clark's Lab
17	28920-43-6	9-Fluorenylmethoxycarbonyl chloride (Fmoc-Cl)	C15H11ClO2	258.7	Smith's Lab
18	2158-14-7	4-acetamidobenzenesulfonyl azide (p-ABSA)	C8H8N4O3S	240.24	aablocks
19	6674-22-2	1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU)	C9H16N	152.24	aablocks
20	110-89-4	Piperidine, Hexahydropyridine	C5H11N	85.15	Smith's Lab
21	25952-53-8	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, EDCI.HCl)	C8H17N3.HCl	155.245	TCI America
22	6485-79-6	Triisopropylsilane (TIPS)	C9H22Si	158.36	TCI America
23	76-05-1	Trifluoroacetic acid (TFA)	C2HF3O2	114.02	ACROS Organic

### 4.3.2 Instrumentation

#### *Maestro Schrodinger*

*In-silico* covalent molecular docking affinity and the prime MMGBSA dG Bind (NS) values were performed and calculated on Maestro® Schrodinger 11.9 2019 software. Before conducting covalent docking, the ligands were prepared using LigPrep to ionize the ligand at physiological pH of 7.4. At the same time, the protein receptor was exported from the protein data bank (PDB 4RGH) and prepared using Protein Preparation Wizard, followed by Receptor Grid Generation. The covalent docking was performed by positioning the ligand within the proximity of 30 – 35 Å centroid to the reactive residue D<sub>220</sub> of receptor chain A or chain B (A: D<sub>220</sub> or B: D<sub>220</sub>). Here, A and B indicate the receptor chain A or B. Meanwhile, the aspartic acid D<sub>220</sub> is part of the signature motif [DSGA] catalytic active site of DDI2.

The prepared ligand was then tagged to obtain CANONICAL SMILES (SMARTS) and identify the reactive side of the ligand, thereby allowing the nucleophilic substitution reaction docking to bind specifically to the  $\alpha$ -carbon atom bearing halogen Bromide in lieu of diazo functional group moiety. The docking was also set using ‘pose thorough’ mode. At the same time, the prime MMGBSA was also calculated as part of the covalent docking. Prime MMGBSA free energy includes covalent bond, coulomb, van der Waals (vdW), lipophilic or hydrophobic bond, the  $\pi - \pi$  stacking, Hydrogen bond, electrostatic, and solvation GB energy.

### *Nuclear Magnetic Resonance (NMR) spectroscopy*

Both experimental  $^1\text{H}$  and  $^{13}\text{C}$  NMR chemical shift of N- $\alpha$ -Fmoc-4(methyl)phenyl glycine in DMSO- $\text{d}_6$ , as shown in **Figure 4.5 A and B** was recorded using Bruker 600 MHz NMR spectrometer at The Auburn University NMR Center, Chemistry Building, Department of Chemistry and Biochemistry, COSAM.

Meanwhile, the experimental  $^1\text{H}$  NMR of azido peptide (**2**, BKF-4MPG-Azide) in DMSO- $\text{d}_6$ , as shown in **Figure 4.6 A**, was recorded on Varian MR400 NMR spectrometer at The Auburn University, Walker Building, HCOP.

All  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were reported as chemical shifts ( $\delta$ , ppm). The experimental  $^1\text{H}$  NMR was normalized to the internal standard TMS ( $\delta$  0.00). Meanwhile, the experimental  $^{13}\text{C}$  NMR was normalized to residual solvent DMSO- $\text{d}_6$  ( $\delta$  39.52).

All NMR spectra were analyzed using MestReNova x64 MestreLab<sup>TM</sup> from the experimental NMR fid.file, compared to predicted  $^1\text{H}$  and  $^{13}\text{C}$  NMR generated based on the chemical structures (**Figure 4.6 B**, **Figure SI 4.3 A**, and **B**). The  $^1\text{H}$  and  $^{13}\text{C}$  NMR prediction was generated using NMR 500 MHz in  $\text{CDCl}_3$ . Peak multiplicities are denoted as follows: s = singlet, d = doublet, t = triplet, q = quartet, dd = doublet of doublets, dt = doublet of triplets, and m = multiplet.

### *Liquid Chromatography - Mass Spectrometry (LC-MS)*

Unless stated otherwise, all LC-MS chromatography profiles (**Figure 4.7 B and C**, **Figure SI 4.5**, **Figure SI 4.8**); the MS spectra (**Figure 4.4**, **Figure 4.7 A**, **Figure SI 4.1**, **Figure SI 4.7**) were acquired using LC-MS Q-ToF electrospray (ESI) ionization

modes at The Auburn University Mass Spectrometry Center, Chemistry Building, Department of Chemistry and Biochemistry, COSAM. These include the High-resolution electrospray ionization mass spectra (HRMS-ESI in **Figure 4.4**, **Figure SI 4.2**, **Figure SI 4.4**, **Figure SI 4.5**, and **Figure SI 4.7**). All resulting spectra were searched against the NIST03 compound library for structural matches. All ESI ionization modes were conducted in positive mode except for HRMS ESI in **Figure SI 4.2** was operated in negative mode.

The LC profile in **Figure 4.7 B and C**, **Figure SI 4.5**, and **Figure SI 4.8** were performed using Reversed phase Waters Acquity UPLC BEH C18 Column, spherical silica (130Å, 1.7 µm, 1 mm X 50 mm, pH 1 - 12, maximum pressure 18000 psi) SKU# 186002344. The crude product separation was carried out in mobile phase A (95% H<sub>2</sub>O, 5% CH<sub>3</sub>CN, 0.1% Formic acid) and mobile phase B (95% CH<sub>3</sub>CN, 5% H<sub>2</sub>O, 0.1% Formic acid), as follows:

**Table 4.3 Analytical LC separation mobile phase**

Time (minute)	Flow Rate (mL/minute)	Solvents (%)	
		A	B
0	0.2	100	0
2	0.2	100	0
13	0.2	0	100
14.5	0.2	0	100
15.5	0.2	100	0

As shown in **Figure SI 4.6**, we purified azido peptide **2** using our instrument semi-preparative LC Bio-Rad Biologic DuoFlow QuadTec™ UV-Vis's detector (catalog #760-0135) at The Auburn University Pharmacy Research Building, HCOP. The QuadTec UV detector was set at 214 nm, 230 nm, 260 nm, and 280 nm. The LC was conducted using the reversed-phase Waters DELTA-PAK prep C18 column, spherical silica (100A, 15µM, 7.8 mm x 300 mm) SKU# WAT011798. The crude

product of azido peptide **2** was dissolved in CH<sub>3</sub>CN: H<sub>2</sub>O (3:1) and injected into the LC system, as described in **Figure SI 4.6**.

Both separation and purification of azido peptide **2** were carried out in mobile phase A (100% H<sub>2</sub>O, 0.1% TFA) and mobile phase B (80% CH<sub>3</sub>CN, 20% H<sub>2</sub>O, 0.1% TFA), as follows:

**Table 4.4 Semi-preparative LC separation mobile phase**

Time (minutes)	Flow Rate (mL/minute)	Solvents (%)	
		A	B
0	1.0	75	25
6.0	1.0	25	68
36.6	1.0	68	100
46.6	1.0	0	100
48.5	1.0	100	25
55.0	1.0	100	25

### 4.3.3 Synthetic methods

#### *General SPPS method.*

Rink amide resin (312.5 mg, 0.25 mmol, 1 equiv.) was first soaked in DMF for an hour to allow the resin beads to swell properly. Subsequently, the N- $\alpha$ -Fmoc deprotection was conducted onto the rink amide resin. Azido Peptides **2** were then synthesized using 9-fluorenylmethoxycarbonyl (Fmoc) chemistry in solid phase peptide synthesis vessels 50 mL (Medium Frit, GL 25 Thread, #CG-1860-03) and shaken vigorously using Glas-Col WS 180° Shaker at a maximum 30 rpm.

Unless stated otherwise, the resin beads were washed with organic solvents in the order of DMF – CH<sub>2</sub>Cl<sub>2</sub> – MeOH – CH<sub>2</sub>Cl<sub>2</sub> for 2x, in between every coupling or deprotection step. Then followed by the final washing step with the solvent used in the synthesis; the last wash was DMF.

The completion of each solid-phase peptide synthesis coupling and deprotection step was monitored using the colorimetric Kaiser test. The Ninhydrin reacts with free primary amines and produces intense dark blue colors of the resin beads and the solution, which can be seen during Fmoc deprotection. When the primary amine is entirely consumed during the coupling reaction, the resin beads become colorless, and the solution becomes pink - purple-ish indicating the completion of the coupling reaction.

***Coupling of N- $\alpha$ -Fmoc amino acids using HBTU and HOBt coupling reagents.***

The following procedure was conducted for the coupling of N- $\alpha$ -Fmoc-N- $\epsilon$ -biotinyl-L-lysine, N- $\alpha$ -Fmoc-Phenylalanine-OH, and N- $\alpha$ -Fmoc-4(methyl)-phenyl glycine-OH.

To this end, the N- $\alpha$ -Fmoc amino acid (0.5 mmol, 2 equiv.) was dissolved in a vial containing DIPEA (165.4  $\mu$ L,  $d = 0.782$  g/mL, 1 mmol, 4 equiv.) and 10 mL DMF. The coupling reagent HBTU (189.74 mg, 0.5 mmol, 2 equiv.) and coupling additive reagent HOBt (77.25mg, 0.57 mmol, 2.28 equiv.) were then added into the vial and allowed to dissolve in the mixture thoroughly. Subsequently, the mixture was slowly transferred into an SPPS vessel containing NH<sub>2</sub>- $\alpha$  free amino acid attached to the resin. Then, the SPPS vessel was shaken vigorously at 30 rpm. The reaction was stopped when the resin beads became colorless upon the colorimetric Kaiser test.

***Coupling of D-Biotin using PyBOP coupling reagent.***

D-Biotin (122.38 mg, 0.5 mmol, 2 equiv.) was dissolved in a previously warmed 10 mL DMF to 60°C. This solution was allowed to cool down to RT before adding DIPEA (165.4  $\mu$ L,  $d = 0.782$  g/mL, 1 mmol, 4 equiv.) and PyBOP (260.79 mg, 0.5 mmol, 2 equiv.). The mixture was manually stirred until completely dissolved. Subsequently, the mixture was slowly transferred into an SPPS vessel containing  $\text{NH}_2$ - $\epsilon$  free lysine attached to the resin and mechanically shaken vigorously at 30 rpm. The reaction was stopped when the resin beads became colorless upon the colorimetric Kaiser test. The coupling step was conducted 2x.

***Fmoc deprotection method.***

The N- $\alpha$ -terminus was then deprotected using a stock solution of 20% (v/v) piperidine in DMF (8 mL). The deprotection time is depended on the colorimetric Kaiser test, where the color of the resin beads and solution should turn to intense blue. The resin was then washed with organic solvents, as described above.

***Alloc deprotection method.***

The following procedure was conducted for the deprotection of N- $\alpha$ -Fmoc-N- $\epsilon$ -Alloc-L-lysine (**19**) to obtain **17**. For this purpose, the resin that was previously washed with  $\text{CH}_2\text{Cl}_2$  was added with a solution of  $\text{PhSiH}_3$  (740  $\mu$ L,  $d = 0.878$  g/mL, 6 mmol, 24 equiv.) in  $\text{CH}_2\text{Cl}_2$  (1 mL) and a solution of  $\text{Pd}(\text{PPh}_3)_4$  (72.86 mg, 0.0625 mmol, 0.25 equiv.) in  $\text{CH}_2\text{Cl}_2$  (3 mL). The reaction was mechanically shaken for 1h. The deprotection step was repeated 2x, and the resin was washed with  $\text{CH}_2\text{Cl}_2$  (2x) between each deprotection step. Following Alloc deprotection, the resin was



twice washed with 10 mL DMF containing 0.5% DTC for 2 minutes to remove the excess of Pd(PPh<sub>3</sub>)<sub>4</sub>.

### *Synthesis of compound 8*

Compound **8** was synthesized using a previously described method by Mix *et al.*<sup>167</sup>. To this end, 4-Methylphenyl glycine (330.3 mg, 2 mmol, 1 equiv.) was dissolved in 5 mL MeCN. Subsequently, *p*-ABSA (528.5 mg, 2.2 mmol, 1.1 equiv.) and non-nucleophilic bases DBU (897  $\mu$ L,  $d = 1.018$  g/mL, 6 mmol, 3 equiv.) were added and stirred overnight at RT. The crude reaction was then concentrated under reduced pressure, and the residue was dissolved in 5 mL EtOAc. Compound **8** was extracted from the EtOAc layer after being prewashed with 1M aq HCl. Subsequently, the organic layer was dried over MgSO<sub>4</sub> anhydrous and evaporated to dryness. The remaining residue was again dried in a low-temperature vacuum oven. No further purification was carried out. HRMS-ESI was calculated for C<sub>9</sub>H<sub>9</sub>N<sub>3</sub>O<sub>2</sub> m/z [M-H]<sup>-</sup> 190.0617 and found as 190.0614 (**Figure SI 4.2**).

### *Synthesis of compound 7*

Compound **7** was synthesized according to a previously described method by Liebeschuetz *et al.*, US patent number US20020055522<sup>169</sup>, with a slight change in the starting material. To this end, 4-Methylphenyl glycine (1.27 g, 7.7 mmol, 1 equiv.) was dissolved in 70 mL CH<sub>2</sub>Cl<sub>2</sub>. Subsequently, DIPEA (13 mL,  $d = 0.782$  g/mL, 7.7 mmol, 10 equiv.) and TMS-Cl (2 mL,  $d = 0.856$  g/mL, 15.4 mmol, 2 equiv.) were added, and refluxed at temperature > 40°C for 1.5 h. The reaction mixture was then cooled in an ice bath and added with Fmoc-Cl (1.99 g, 7.7 mmol, 1 equiv.). The reaction was stirred at 0°C for 20 minutes and stirred at RT for 1.5

h. TLC monitored the completion of the reaction. After the reaction was completed, it was concentrated under reduced pressure. The remaining residue was dissolved in a 50 mL mixture of Et<sub>2</sub>O and Na<sub>2</sub>CO<sub>3</sub> saturated (ratio 2 : 3). The newly formed yellow solid (insoluble in Et<sub>2</sub>O or saturated Na<sub>2</sub>CO<sub>3</sub>) was brought in H<sub>2</sub>O and acidified with a 50% HCl in H<sub>2</sub>O to pH 1. The mixture was then extracted with EtOAc (3x) and washed with brine (2x). The EtOAc layer was dried over MgSO<sub>4</sub> and evaporated to dryness. The remaining residue was recrystallized in a mixture of petroleum ether: EtOAc. The precipitate was rewashed with cold petroleum ether and dried in a low-temperature vacuum oven to afford a white solid (2.38 g, 80% yield). <sup>1</sup>H NMR (600 MHz, DMSO-d<sub>6</sub>, δ): δ 12.81 (s, 1H), 8.13 (d, *J* = 8.0 Hz, 2H), 7.89 (d, *J* = 7.6 Hz, 5H), 7.77 (d, *J* = 7.5 Hz, 4H), 7.45 – 7.40 (m, 5H), 7.33 (dd, *J* = 12.6, 7.5 Hz, 10H), 7.19 (d, *J* = 7.7 Hz, 5H), 6.3 (s, *J* = 1.0 Hz, 1H), 5.75 (s, 1H), 5.13 (d, *J* = 8.0 Hz, 2H), 4.33 – 4.20 (m, 6.0Hz, 7H), 2.31 (s, *J* = 1.0 Hz, 3H) (**Figure 4.5 A**). <sup>13</sup>C NMR (600 MHz, DMSO-d<sub>6</sub>) δ 172.62, 144.32, 144.23, 141.17, 137.64, 134.66, 129.41, 128.15, 128.11, 127.53, 125.90, 120.54, 66.41, 58.29, 47.10, 39.98, 21.15 (**Figure 4.5 B**). HRMS-ESI found *m/z* [M+H]<sup>+</sup> 388.1543 calculated for C<sub>24</sub>H<sub>21</sub>NO<sub>4</sub> 388.1549 (**Figure 4.4**).

### ***On-resin synthesis of azido peptide 3***

Azido peptide **3** was synthesized according to the method described by Castro *et al.*<sup>160</sup>. To this end, ISA-HCl (263.25 mg, 1.25 mmol, 5 equiv.) was dissolved in 10 mL DMF, then added with DIPEA (620 μL, *d* = 0.782, 3.75 mmol, 15 equiv. Next, this solution mixture was added to the SPPS reaction vessel containing peptide **4** tethered to the resin. The CuSO<sub>4</sub>·5H<sub>2</sub>O catalyst (100 μL of 0.25 mM, 0.0025 mmol,

0.01 equiv.) was added to accelerate the reaction. In order for azide transfer to occur, the pH of the reaction was maintained at ~10. Subsequent on-resin azide transfer was carried out overnight. The reaction was terminated when the resin beads became colorless, and the solution's color was similar to the control. Azido peptide **3** was then released from the resin, purified through semi-preparative LC, and lyophilized to afford Azido peptide **2**. Experimental <sup>1</sup>H NMR spectra (400 MHz, DMSO-d<sub>6</sub>) δ 8.46 (t, *J* = 8.6 Hz, 1H), 8.10 (dd, *J* = 26.7, 8.1 Hz, 1H), 7.71 (q, *J* = 5.5 Hz, 1H), 7.28 – 7.15 (m, 2H), 7.17 – 7.09 (m, 1H), 7.13 - 6.97 (m, 3H), **6.40** (s, 1H, **H<sub>17</sub>**), **6.34** (s, 1H, **H<sub>15</sub>**), 4.93 (d, *J* = 2.1 Hz, 1H), 4.68 – 4.55 (m, 1H), 4.28 (dd, *J* = 7.7, 4.9 Hz, 1H), 4.30 - 4.25 (m, 1H, H<sub>12</sub>), 4.20 - 4.06 (m, 1H, H<sub>13</sub>), 3.8 - 3.6 (m, 1H, H<sub>8</sub>), 3.08 - 3.05 (m, 2H, H<sub>22</sub>) 3.02 – 2.93 (m, 1H, H<sub>3</sub>, H<sub>30</sub>), 2.98 (dt, *J* = 12.9, 6.9 Hz, 1H, H<sub>20</sub>), 2.87 – 2.68 (m, 1H), 2.55 (d, *J* = 12.4 Hz, 1H, H<sub>20</sub>), **2.28** (s, *J* = 3.7 Hz, 3H, **H<sub>46</sub>**), 2.13 - 2.10 (m, 2H, H<sub>26</sub>), 2.13 - 1.97 (td, 2H, *J* = 7.4, 2.0 Hz, H<sub>10</sub>, H<sub>26</sub>), 1.65 - 1.54 (m, 2H, H<sub>10</sub>), 1.54 - 1.40 (m, 6H, H<sub>11</sub>, H<sub>23</sub>, H<sub>25</sub>), 1.38 - 1.12 (m, 4H, H<sub>24</sub>, H<sub>31</sub>) (**Figure 4.6 A**). HRMS-ESI was calculated for C<sub>34</sub>H<sub>45</sub>N<sub>9</sub>O<sub>5</sub>S 692.3337 and was found as m/z [M+H]<sup>+</sup> 692.3334 (**Figure 4.7**).

#### ***Cleavage of Peptide 2 from Rink Amide Resin.***

At the end of the SPPS, a mixture of TFA/iPr<sub>3</sub>SiH/H<sub>2</sub>O (95:2.5:2.5 v/v/v; 4 mL) was added to the resin and agitated for 2h at RT. The liquids were separated from the resin using fritted glass funnels, and the resin was washed with TFA/iPr<sub>3</sub>SiH/H<sub>2</sub>O (95:2.5:2.5 v/v/v; 2 mL) 2X. The remaining liquid was evaporated to dryness.

#### **4.3.4 Cell lysis, protein quantification, and bioconjugation using peptide 2**

Cells were washed and harvested with PBS, collected by centrifugation, and frozen at  $-80^{\circ}\text{C}$ . The frozen cell pellets were lysed in PBS pH 7.4 containing Digitonin 0.5%, Na-acetate 0.1M buffer pH 5.2 containing Digitonin 0.5%, or NP-40 lysis buffer as a negative control. The cells were incubated for 15 min on ice, centrifuged at  $20,000 \times g$  for 15 – 20 min, and the supernatants were used for experiments. The protein quantification was performed as described in **Chapter 2, Section 2.4.4**. For photoconjugation, cell pellets were lysed and quantified on the day of the experiment.

Prior to the cross-linking experiment, the stock concentration of azido peptide **2** (BKF-4MPG-azide) was prepared in DMSO to 100  $\mu\text{M}$ . At the same time, the working concentration was prepared in its lysis buffer. Then, 1  $\mu\text{g}$  of cell lysates were mixed with azido-peptide **2** at a concentration of 1  $\mu\text{M}$  and 10  $\mu\text{M}$  in its lysis buffer. The reaction mixture was incubated for  $\pm 30$  minutes in 384 well-dark plates under long wave UV 365nm. The UV lamp (Spectroline model ENF-280C) was positioned over the reaction mixture to allow direct photoactivation. During photoactivation, the reaction mixture was covered with aluminum foil. Then, the samples were collected for immunoblotting.

#### ***Western blotting***

The gel electrophoresis was performed using GenScript SurePAGE™ Bis-Tris 10% mini gel (Cat #M00666) at 170V and GenScript MES SDS running buffer (1X, containing 0.5M MES, ddH<sub>2</sub>O to pH 6) Cat#M00677. The protein was transferred on Immobilon–FL PVDF membrane 0.2 $\mu\text{M}$  pore-diameter (Cat #IPFL00010) using

wet transfer buffer GenScript (1X, containing Tris-base, Bicine, and 10% methanol) Cat # M00139. The membrane was blocked with Li-Cor Intercept® (TBS) Blocking Buffer (Cat #927-60001) in TBSt (1X).

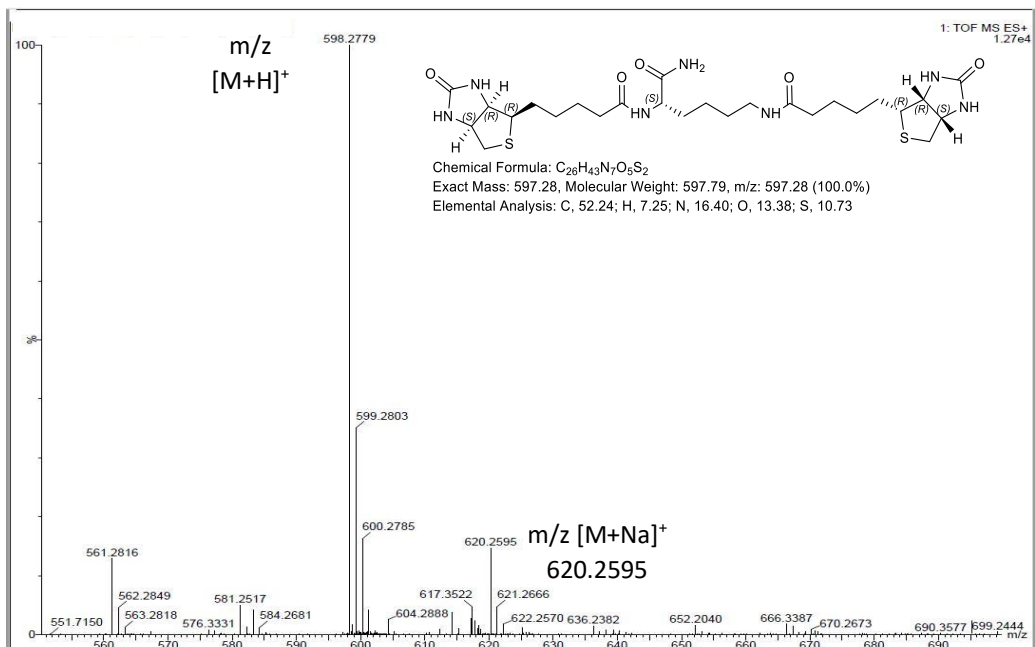
**Table 4.5 List of antibodies**

<b>IgG</b>	<b>Type</b>	<b>Host</b>	<b>Dilution</b>	<b>Company</b>	<b>Catalog #</b>
<b>Primary</b>					
Anti-biotin(D5A7) HRP-conjugate	mAb	rabbit	1:1000	Cell Signaling	5571S
DDI2	pAb	rabbit	1:2000	Bethyl®Lab	A304- 629A
<b>Secondary</b>					
Anti-rabbit HRP-linked		Goat	1:1000	Cell Signaling	7074S
<b>Ladder</b>					
Spectra™ Multicolor Broad Range ladder	protein			Thermofisher Scientific	26623
Biotinylated ladder	protein			Cell Signaling	818515

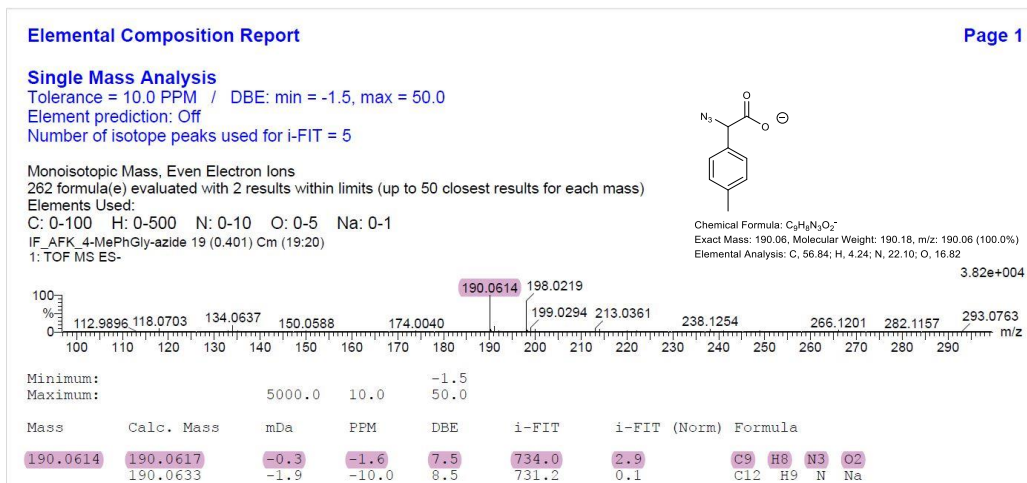
#### 4.4 Supporting Information

**Table SI 4.1 The prime MMGBSA dG Bind (NS) energy components of designed probes (not calculated over the strain energy of conformational changes).**

Probe #	Covalent Docking score	MM GBSA dg bind (NS) kcal/mol										Covalent binding Distance
		total	Cov.	Co $\mu$ Lomb	vdW	Hydrophobic	H-bond	Solv GB	$\pi - \pi$			
1	-1.983	-17.79	0.48	-2.19	-18.25	-6.51	0.00	8.68	0.00	1.49A		
2	-1.899	-24.61	0.46	-2.26	-24.15	-8.96	0.00	10.30	-0.00	1.44A		
3	-3.328	-27.41	0.48	-9.40	-21.48	-9.88	-0.47	13.61	-0.28	1.46A		
4	-2.497	-25.69	0.46	-13.14	-17.15	-7.78	-0.81	12.82	-0.09	1.44A		
5	-4.805	-51.39	0.00	-22.24	-44.28	-14.37	-1.63	32.47	-1.34	1.44A		
6	-5.650	-69.84	0.00	-42.09	-49.00	-18.41	-2.70	43.69	-1.34	1.45A		
7	<b>-6.540</b>	<b>-68.04</b>	0.00	-31.67	-52.71	-19.30	-2.01	38.42	-0.78	1.47A		
8	-5.848	-56.81	0.00	-34.70	-48.55	-17.94	-1.82	46.22	-0.01	1.45A		
9	-2.897	-68.60	0.00	-38.18	-53.16	-17.97	-3.44	44.16	-0.00	1.47A		
10	<b>-6.136</b>	<b>-62.40</b>	0.00	-36.13	-55.42	-19.74	-1.88	51.89	-1.11	1.46A		



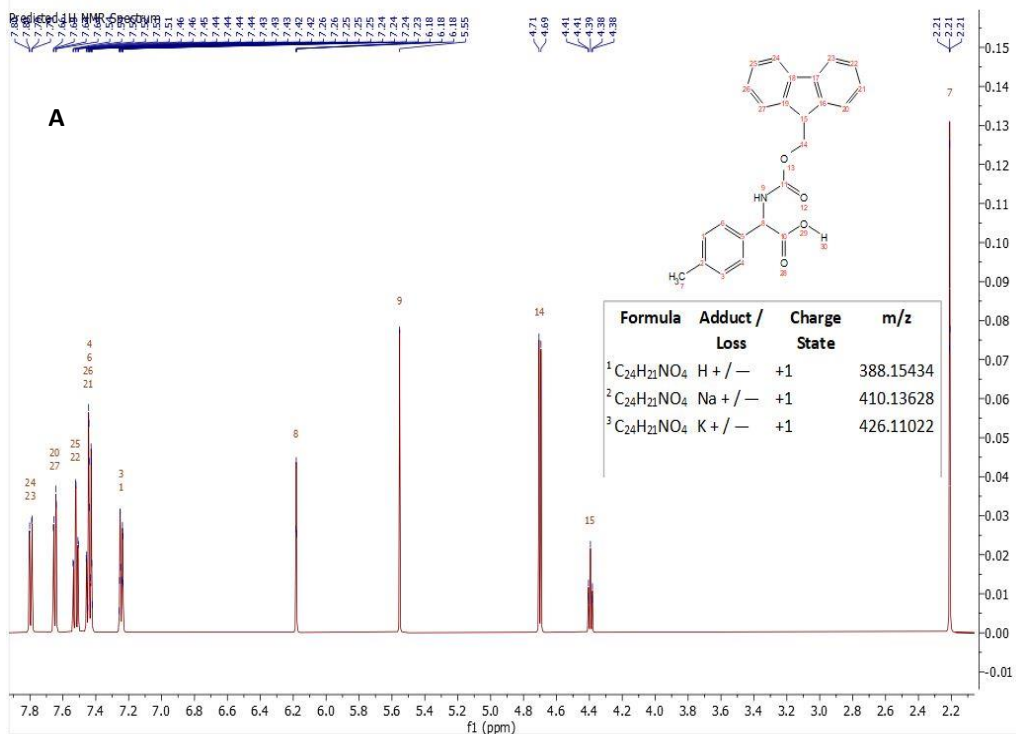
**Figure SI 4.1 HPLC-MS Spectra of undesirable products calculated for  $C_{26}H_{43}N_7O_5S_2$ ,  $m/z$   $[M+H]^+$  598.28 (found 598.2779) and  $C_{26}H_{43}N_7O_5S_2$  Na  $m/z$   $[M+Na]^+$  620.26 (found 620.2595).**



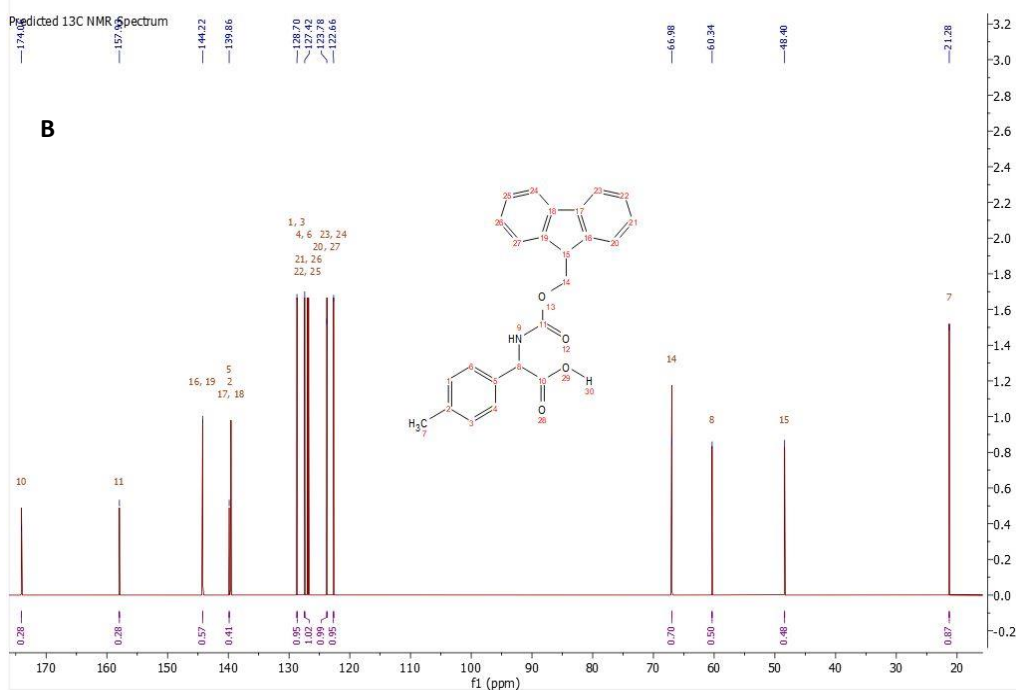
**Figure SI 4.2 HRMS-ESI of 4(methyl) phenyl glycine azide calculated for  $C_9H_9N_3O_2$   $m/z$   $[M-H]^-$  190.0614.**

$^1\text{H}$  NMR (500 MHz, Chloroform-d)  $\delta$  7.80 (dd,  $J = 7.4, 1.5$  Hz, 2H), 7.65 (dd,  $J = 7.4, 1.5$  Hz, 2H), 7.52 (td,  $J = 7.5, 1.6$  Hz, 2H), 7.48 – 7.40 (m, 4H), 7.24 (dq,  $J = 7.5, 1.1$  Hz, 2H), 6.18 (t,  $J = 1.0$  Hz, 1H), 5.55 (s, 1H), 4.70 (d,  $J = 5.9$  Hz, 2H), 4.39 (t,  $J = 6.0$  Hz, 1H), 2.21 (t,  $J = 1.0$  Hz, 3H).

**Predicted  $^1\text{H}$  NMR spectra**

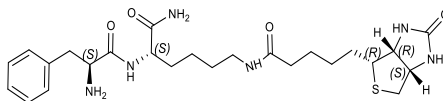


$^{13}\text{C}$  NMR (125 MHz, Common NMR Solvents)  $\delta$  174.06, 157.93, 144.22, 139.86, 128.70, 127.42, 123.78, 122.66, 66.98, 60.34, 48.40, 21.28. **Predicted  $^{13}\text{C}$  NMR spectra**

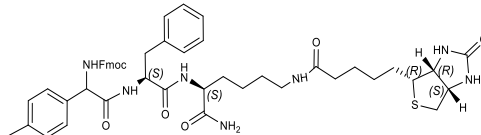
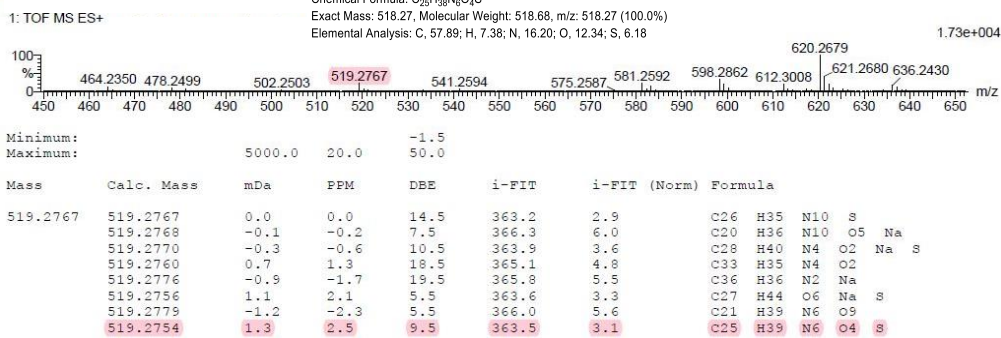


**Figure SI 4.3 Predicted  $^1\text{H}$  NMR and  $^{13}\text{C}$  NMR chemical shift (ppm) of *N*- $\alpha$ -Fmoc-4(methyl)phenyl glycine in  $\text{CDCl}_3$ . These predicted NMR spectra were generated on MestReNova x64 Mestrelab™ based on the chemical structure compared to experimental NMR in Figure 4.5.**

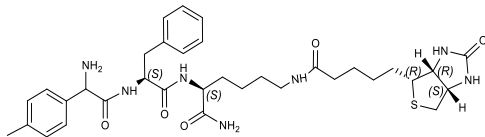
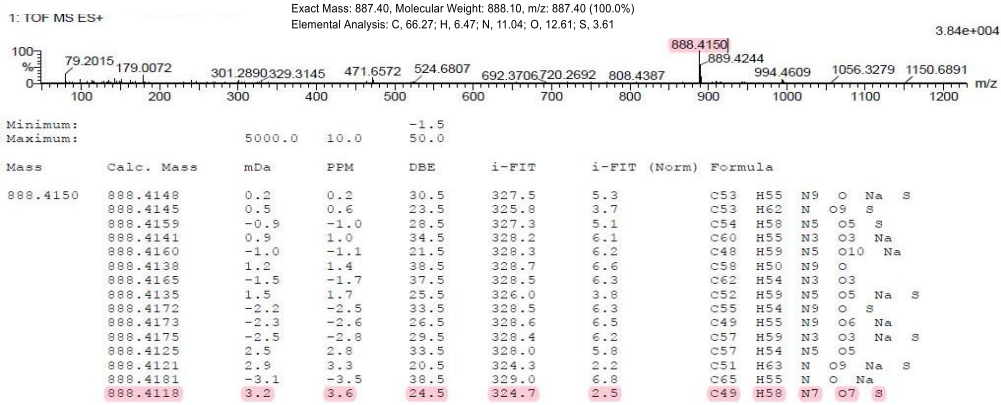




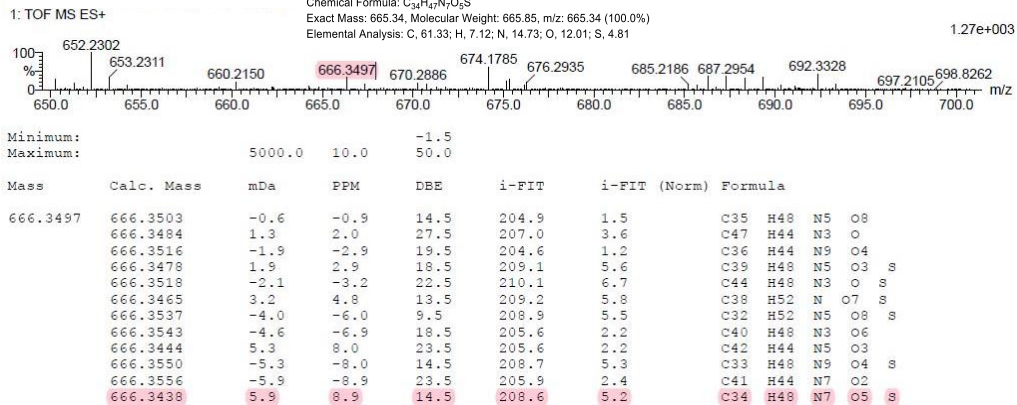
Chemical Formula:  $C_{25}H_{38}N_6O_4S$   
 Exact Mass: 518.27, Molecular Weight: 518.68, m/z: 518.27 (100.0%)  
 Elemental Analysis: C, 57.89; H, 7.38; N, 16.20; O, 12.34; S, 6.18



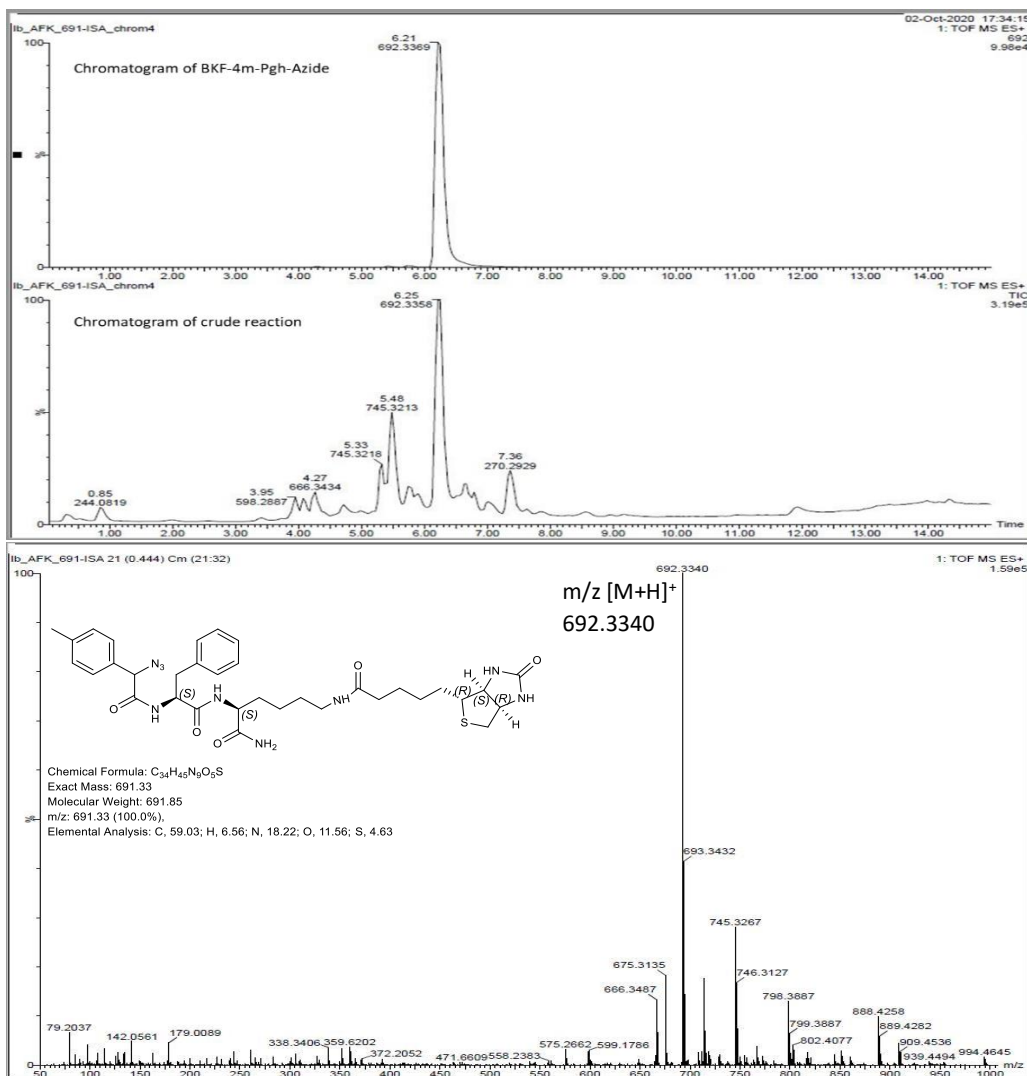
Chemical Formula:  $C_{49}H_{57}N_7O_7S$   
 Exact Mass: 887.40, Molecular Weight: 888.10, m/z: 887.40 (100.0%)  
 Elemental Analysis: C, 66.27; H, 6.47; N, 11.04; O, 12.61; S, 3.61



Chemical Formula:  $C_{34}H_{47}N_7O_5S$   
 Exact Mass: 665.34, Molecular Weight: 665.85, m/z: 665.34 (100.0%)  
 Elemental Analysis: C, 61.33; H, 7.12; N, 14.73; O, 12.01; S, 4.81



**Figure SI 4.4 HRMS-ESI of compounds 6, 5, and 4 sampled from crude SPPS reaction. Compound 6 was calculated for  $C_{25}H_{38}N_6O_4S$  and found as  $[M+H]^+$  519.2767. Compound 5 (calculated for  $C_{49}H_{57}N_7O_7S$ ) was detected as  $[M+H]^+$  888.4150. Compound 4 (calculated for  $C_{34}H_{47}N_7O_5S$ ) was observed as  $[M+H]^+$  666.3497.**



### Elemental Composition Report

Page 1

#### Single Mass Analysis

Tolerance = 10.0 PPM / DBE: min = -1.5, max = 50.0

Element prediction: Off

Number of isotope peaks used for i-FIT = 5

Monoisotopic Mass, Even Electron Ions

3608 formula(e) evaluated with 36 results within limits (up to 50 closest results for each mass)

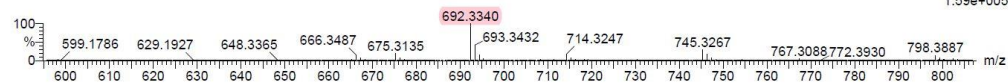
Elements Used:

C: 1-100 H: 0-100 N: 0-10 O: 0-10 Na: 0-1 S: 0-1

Ib\_AFK\_691-ISA 21 (0.444) Cm (21:32)

1: TOF MS ES+

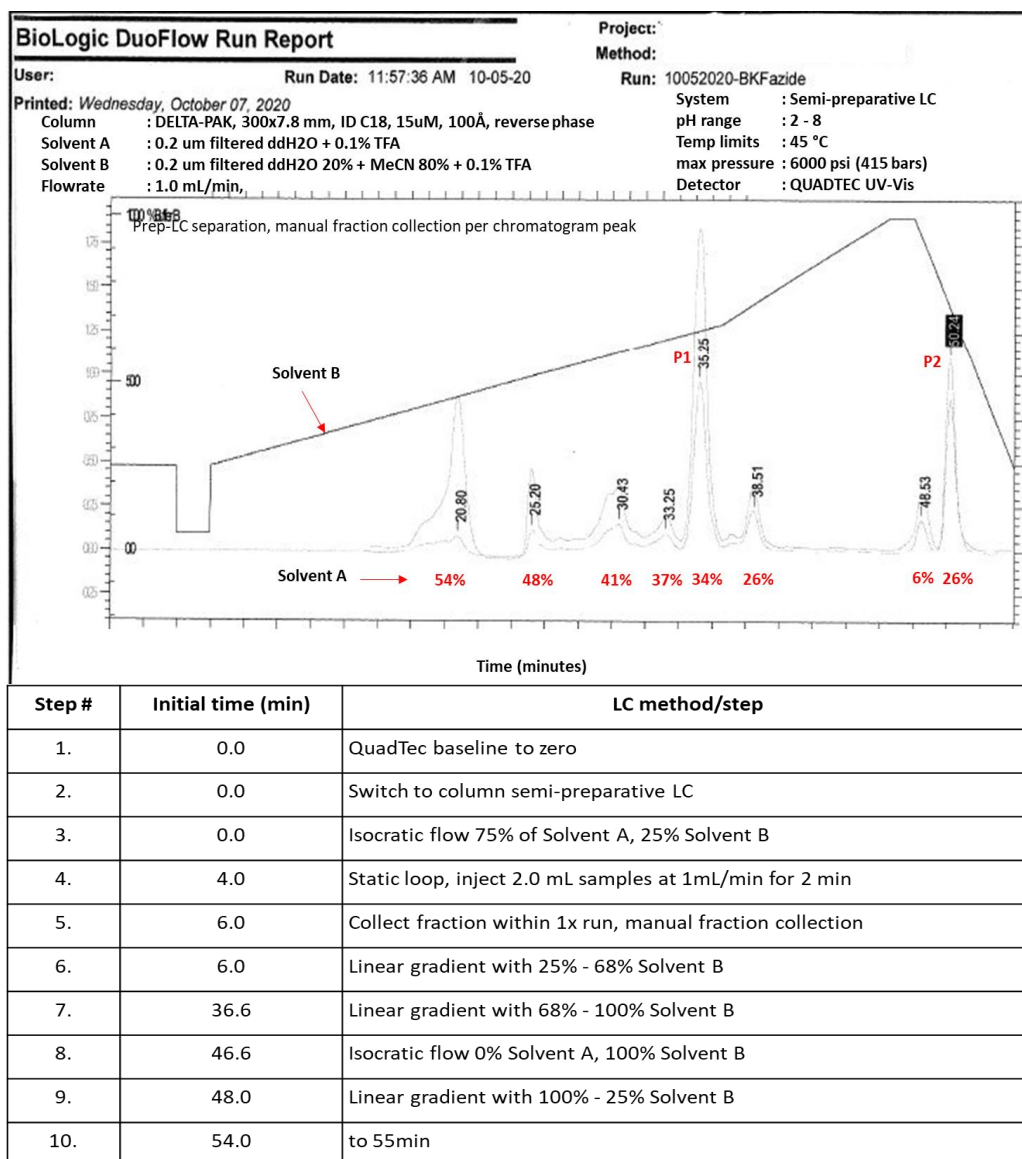
1.59e+005



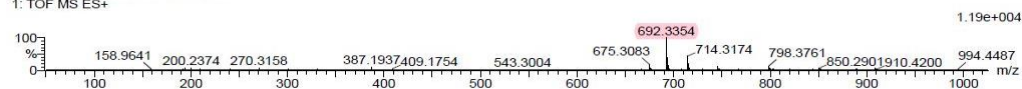
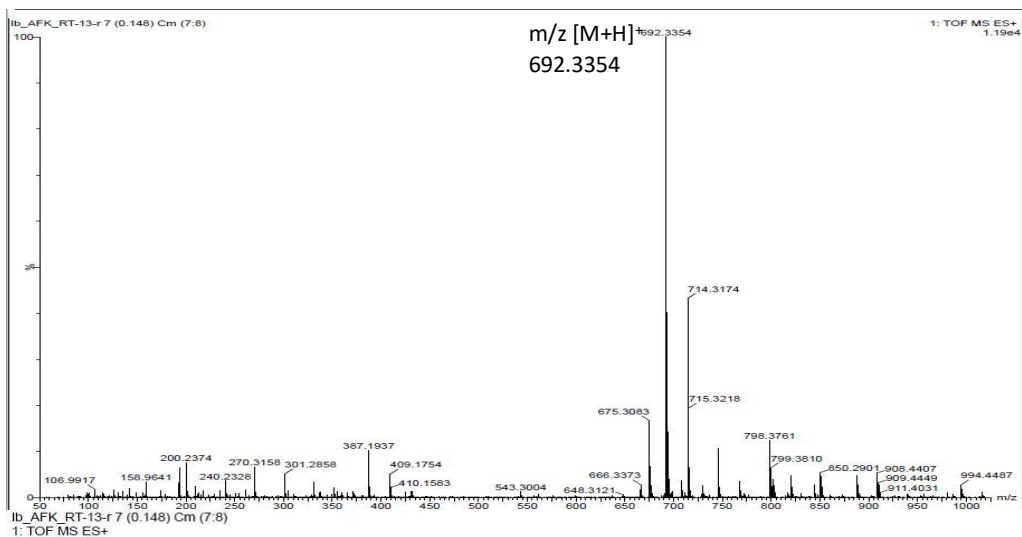
Minimum: -1.5  
Maximum: 50.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
692.3340	692.3343	-0.3	-0.4	16.5	728.5	2.4	C34 H46 N9 O5 S
	692.3336	0.4	0.6	20.5	735.4	9.3	C41 H46 N3 O7
	692.3344	-0.4	-0.6	9.5	738.8	12.7	C28 H47 N9 O10 Na
	692.3345	-0.5	-0.7	12.5	727.0	0.8	C36 H51 N3 O7 Na S
	692.3349	-0.9	-1.3	25.5	736.4	10.2	C42 H42 N7 O3
	692.3329	1.1	1.6	11.5	726.9	0.8	C33 H50 N5 O9 S

**Figure SI 4.5** LC-MS chromatography profile of azido peptide 2 before purification showed chromatogram peak at RT between  $\pm 6.2$  min for exact mass 692.3358. LC-MS spectra and HRMS-ESI were calculated for  $C_{34}H_{45}N_9O_5S$  and found as m/z [M+H]<sup>+</sup> 692.3340.

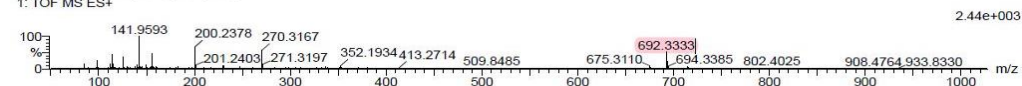
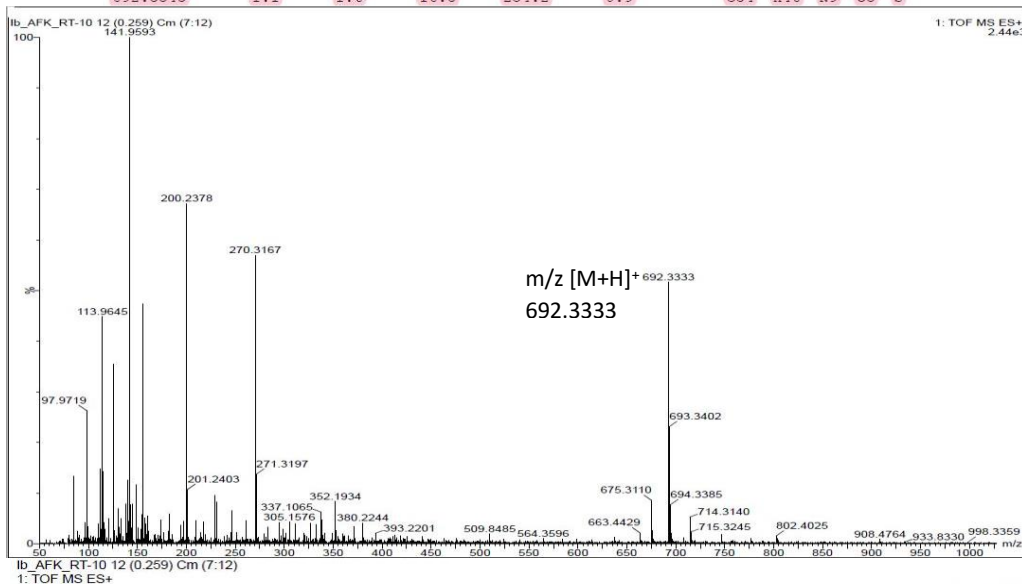


**Figure SI 4.6 Chromatography profile of azido peptide 2 and its purification using semi-preparative LC.** The peptide fraction of each chromatogram peak was manually collected and examined for calculated mass  $C_{34}H_{45}N_9O_5S$ . Masses of 691.33 were contained in chromatogram peaks 1 (P1) and 2 (P2).



Minimum: -1.5  
Maximum: 5000.0 10.0 50.0

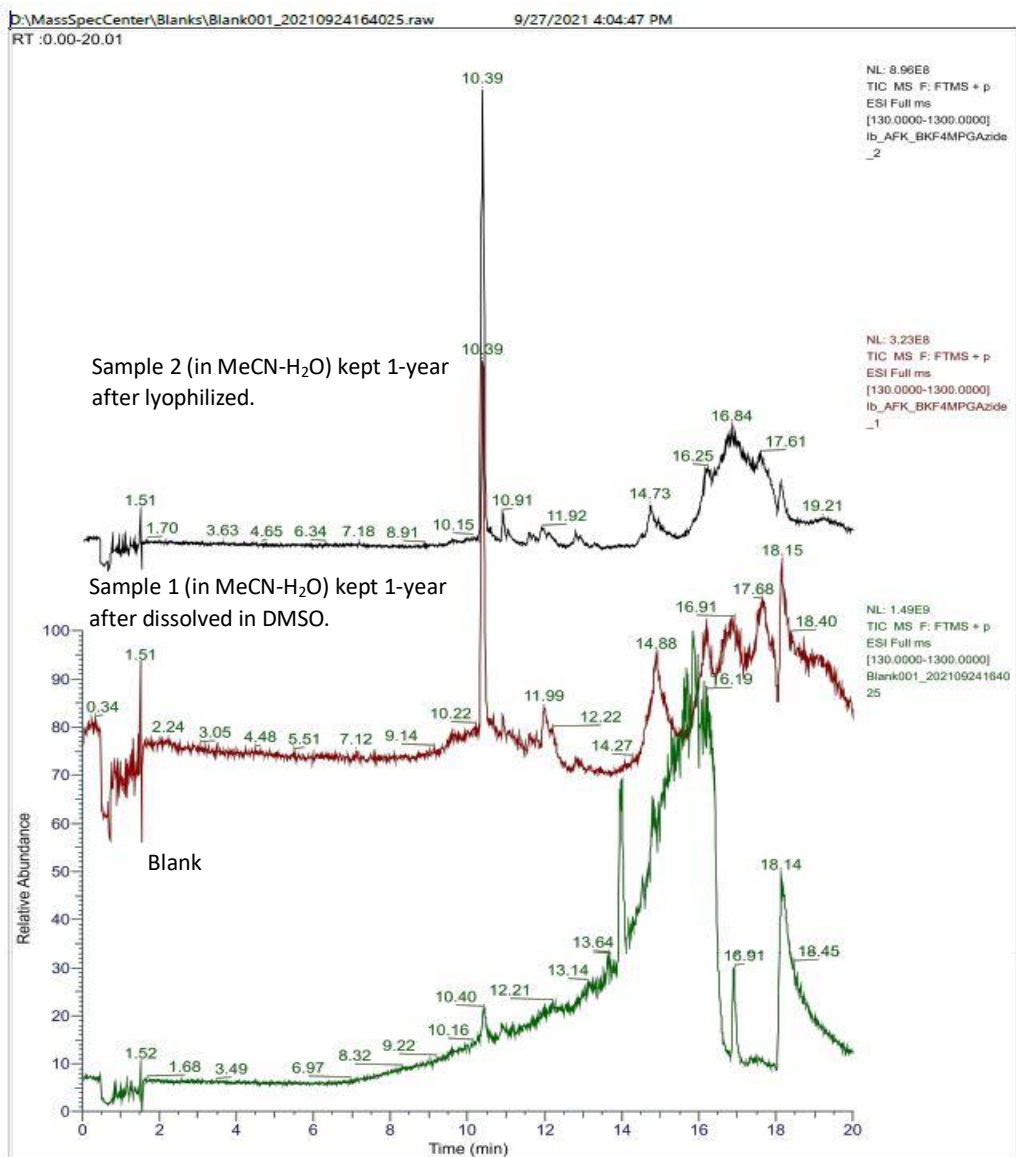
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
692.3354	692.3352	0.2	0.3	21.5	260.0	6.7	C44 H47 N O5 Na
	692.3351	0.3	0.4	28.5	261.0	7.7	C50 H46 N S
	692.3359	-0.5	-0.7	17.5	255.7	2.4	C37 H47 N7 O3 Na S
	692.3349	0.5	0.7	25.5	260.2	6.9	C42 H42 N7 O3
	692.3365	-1.1	-1.6	26.5	260.9	7.6	C45 H43 N5 O Na
	692.3343	1.1	1.6	16.5	254.2	0.9	C34 H46 N9 O5 S



Minimum: -1.5  
Maximum: 5000.0 10.0 50.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
692.3333	692.3327	0.6	0.9	25.5	225.3	6.2	C48 H47 N Na S
	692.3325	0.8	1.2	22.5	224.6	5.5	C40 H43 N7 O3 Na
	692.3343	-1.0	-1.4	16.5	220.2	1.2	C34 H46 N9 O5 S
	692.3319	1.4	2.0	13.5	222.0	3.0	C32 H47 N9 O5 Na S
	692.3349	-1.6	-2.3	25.5	223.9	4.9	C42 H42 N7 O5
	692.3317	1.6	2.3	33.5	226.4	7.4	C53 H42 N

**Figure SI 4.7** LC-MS Spectra profile of chromatogram peak 1 (P1) and peak 2 (P2) in **Figure SI 4.6**. HRMS-ESI of the calculated mass  $C_{34}H_{45}N_9O_5S$ . Chromatogram peak 1 shows  $m/z$   $[M+H]^+$  692.3354. Meanwhile, chromatogram peak 2 shows  $m/z$   $[M+H]^+$  692.3333.



**Figure SI 4.8** Overlays of LC-MS chromatogram profile of purified azido peptide 2 and blank showed a major peak at retention time 10.39 min. Both samples were initially kept in lyophilized form. Sample 1 (S1) was stored in DMSO and kept for a year before being diluted in MeCN: H<sub>2</sub>O (3:1). Sample 2 (S2) was stored in its lyophilized form for a year prior to being diluted in MeCN: H<sub>2</sub>O (3:1).

## Chapter 5 Concluding Remarks and Future Directions

DDI2 is a novel aspartic protease whose biological function has not been widely studied and characterized<sup>16</sup>. Interestingly, this protein consists of a retroviral protease domain with a distinctive motif similar to a well-characterized HIV protease<sup>16</sup>, making it an exciting target for drug development. So far, the only known substrate of DDI2 is the Nrf1 transcription factor involved in proteasome degradation<sup>17</sup>. Moreover, DDI2 is also a ubiquitin-binding protein that only cleaves polyubiquitylated substrates similar to the proteasome<sup>17, 89</sup>.

Upon proteasome inhibition, DDI2 precisely cleaves a single peptide bond in polyubiquitylated Nrf1<sup>14</sup>. Nrf1 processing further induces proteasome gene activation and is reportedly involved in proteasome activity recovery<sup>5, 11-14</sup>. To our knowledge, the function of hDDI2 in the ubiquitin-proteasome system results in only slight differences (~25%) in proteasome recovery between hDDI2 KO and parental cells<sup>15, 94</sup>, raising a question on the significance of hDDI2 in the recovery of proteasome activity. If DDI2 is essential for Nrf1 processing and Nrf1 activation is a primary determinant for the induction of proteasome genes when the proteasome is inhibited, disabling hDDI2 should result in a >50% reduction of proteasome recovery activity. We address this question in **Chapter 2**. We initially hypothesized that inhibiting the DDI2 protein would impair Nrf1 processing, thereby blocking the induction of proteasome genes and the recovery of proteasome activity. However, our results indicate that the proteasome activity recovery occurred independently of the DDI2 – Nrf1 pathway. We showed that the late onset of proteasome activity recovery predominantly occurred through increased

activation of the 26S proteasome. However, the early start of proteasome activity recovery is transcriptionally independent of the induction of proteasome genes. **The practical implication of our finding is that the proteasome activity recovery depends on the cell's physiology and can not be generalized to all cancer cells.** HAP1 cells do not accumulate polyubiquitylated proteins under basal conditions<sup>95</sup>—treatment of HAP1 cells with Btz (100nM) only slightly accumulated polyubiquitylated proteins after 16h<sup>95</sup>. Likewise, a higher concentration of proteasome inhibitors is required to accumulate polyubiquitylated substrates in other cancer cells<sup>95</sup>. Meanwhile, multiple myeloma (MM) endures basal proteotoxic stress due to high monoclonal immunoglobulin secretion and misfolded protein accumulation<sup>123, 177</sup>. These increase the proteasome load to degrade abnormal proteins<sup>123, 177</sup>. Treatment of MM cells with proteasome inhibitors increases the amount of polyubiquitylated protein in cells<sup>17, 123</sup>. The burdens of proteasome seem higher in MM cells than in HAP1 cells. Inactivating DDI2 leads accumulation of high molecular weight polyubiquitylated substrates in HAP1 and MM cells due to impaired function of DDI2 as ubiquitin protein shuttles<sup>17, 95</sup>. However, knocking out DDI2 has a more significant effect in exacerbating the increased load of non-degraded polyubiquitylated in MM cells because those cells are always under basal proteotoxic stress<sup>17, 38, 95</sup>.

Moreover, HAP1 cells still exhibit efficient mRNA translation and protein synthesis of proteasome subunits during the recovery of proteasome activity after a one-hour treatment with a proteasome inhibitor, as we showed in our study (**Figure 2.4 – Figure 2.7**). In our research, the processing of Nrf1 by DDI2 and induction of mRNA proteasome subunits do not contribute to the recovery of proteasome

activity because it does not increase the rate of protein synthesis (**Figure 2.2, Figure 2.4 – Figure 2.7**). Thus, when DDI2 is impaired in HAP1 cells resulting in the accumulation of unprocessed Nrf1, it does not reduce the recovery of proteasome activity. Meanwhile, although MM's protein synthesis rate is high, it is not followed by proper folding<sup>178</sup>. Hence, it tends to result in more misfolded proteins and damaged ribosomal products<sup>178</sup>. Lower efficiency in protein folding and increased defective ribosomal products coupled with a rapid increase in the degradation of nascent polypeptides reduces the capacity of functional proteasome<sup>123, 178</sup>. Both increased proteasomal load and decreased active proteasome augmented proteasomal stress in MM<sup>123, 177-178</sup>. The proteasome gene's induction due to the activation of Nrf1 by DDI2 may help MM cells increase their proteasomal capacity<sup>38</sup>. Conversely, knocking out DDI2 will significantly reduce the number of functional proteasomes in MM cells and impair their ability to restore proteasome activity<sup>38</sup>. Hence, the processing of Nrf1 by DDI2 may be necessary for cancer with lower efficiency protein synthesis or folding.

Based on the findings in **Chapter 2**, moving forward, we still have homework to prove that rapid assembly of stably nascent proteasome subunits takes place in the early start of the proteasome recovery in HAP1 cells by conducting pulse labeling and chase experiments. During native gel, we observed a subset of untreated controls did not migrate into the gel compared to the treated samples, suggesting some changes in the shape and total charges of the proteasome. This observation was more pronounced in the untreated DDI2 KO cells (data are not shown). In addition, we also observed an unknown band (labeled as \*\*) between the latent 20S and 26S proteasome (**Figure 2.8**). This unknown complex is present in untreated



controls and during the proteasome activity recovery. However, the extent to which this band contributed to the rescue of proteasome activity remains for future study. Addressing these two questions will help develop a better approach for studying 26S/the 30S and 20S proteasome ratios during proteasome assembly or recovery of proteasome activity.

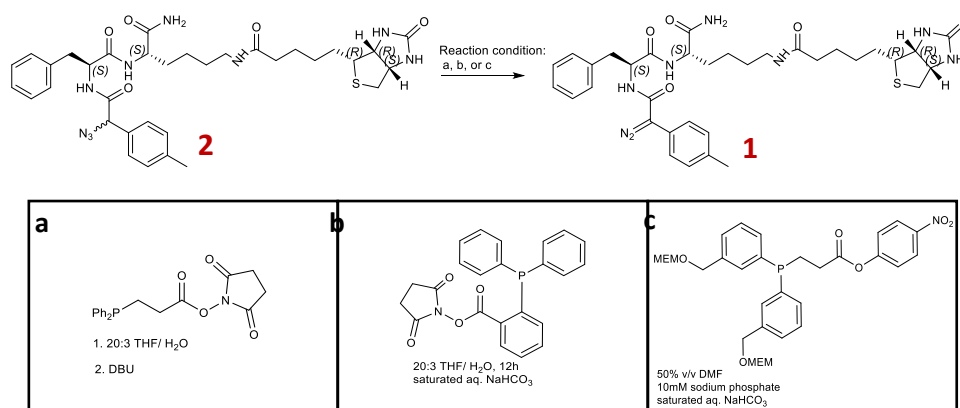
At the basal level, deletion of DDI2 causes severe developmental failure and mid-late gestation embryonic lethality<sup>20</sup>, demonstrating the importance of DDI2 in normal function and development. However, the extent to which this deletion of DDI2 leads to the disruption of normal physiological homeostasis is still unknown. In **Chapter 3**, we attempted to answer the challenge of understanding the biological function of DDI2 in cells using a combination of the CRISPR KO gene-editing screening and a high-throughput whole-transcriptome analysis approach. Assessing the changes in RNA molecules level and integrating them with pathway analysis helps us determine the likelihood of affected pathways in DDI2 knockout. We found that knockout of DDI2 significantly downregulates antigen presentation pathway essential for T-cell effector activation, particularly in the tumor microenvironment. Knockout of DDI2 also significantly downregulates most of the genes involved in metabolic pathways. Using a combination of Differentially expressed genes and pathway analysis; we found two significantly downregulated genes, CD44 and ALDH1A2. These two genes have been reported as markers of cancer stem cells<sup>130-132, 136-139</sup>. Pathway analysis suggests that the downregulation of CD44 significantly affects the tumor microenvironment. Supporting this, previous studies also reported CD44 positively regulates Programmed death-ligand 1 (PD-L1) expression in the tumor microenvironment<sup>132</sup>. Hence, it will be

interesting to study the mechanism of how DDI2 affects CD44 expression. ALDH1A2 is an essential coenzyme in glycolysis and the tricarboxylic acid cycle (TCA)<sup>136</sup>. Based on our transcriptome analysis, knocking out of DDI2 downregulates not only ALDH1 but also most of the genes' transcriptional expression in metabolite transport and enzymatic metabolic pathways. Downregulating these genes may have contributed to the severe developmental failure and mid-late gestation embryonic lethality, as previously reported in DDI2 knockout mice<sup>20</sup>. It may also contribute to critical pathways involving proteostasis. Therefore, further investigation is needed to address the knockout effects of DDI2 on metabolic pathways, particularly the oxidative metabolism pathway.

In addition, transcriptome analysis suggested that DDI2 knockout decreased transcriptional expression of immunoproteasome subunits and immunoproteasome activity. However, these findings are not yet conclusive; further investigation in DDI2 knockout cells derived from different hematopoietic cells is still needed to confirm that these effects are not only specific to HAP1 cells.

At last, during the study of DDI2 in **Chapter 2** and **Chapter 3**, we realized one limitation of studying the biological function of DDI2 is the lack of biochemical assays to probe or inhibit the proteolytic activity of DDI2. As a multi-complex and multi-functional protein, DDI2 is still regarded as an undruggable protein. Hence, in **Chapter 4**, we attempted to design a probe that can bind covalently with the aspartic acid of the DDI2 catalytic active site. However, during the experiment, we found that azido peptide 2 can bind with DDI2 without photo-conjugation. In the future, we can perform pulldown assays or affinity chromatography to answer how

this binding occurs. To address whether azido peptide 2 can bind with the proteolytic active site DDI2, we must conduct photo-conjugation experiments in purified DDI2 and DDI2 DN. If azido peptide 2 cannot bind to the active catalytic site of DDI2, we can move forward to synthesize the diazo probe. Based on the literature, we propose to convert peptide azide 2 into a phosphinoester-mediated diazo probe, as outlined below<sup>163-164</sup>.



**Figure 5.1 Conversion of Azide into Diazo compounds via Azide deimidogenation mediated by several phosphinoesters (a)<sup>163</sup>, (b)<sup>163</sup>, and (c)<sup>164</sup>.** When  $\alpha$ -azido containing compounds react with phosphinoesters, the resulting phosphazide intermediate will exclusively yield diazo compounds under controlled  $pK_a < 7.1$  at RT.

Overall, although our finding suggests that DDI2 does not play a role in the recovery of proteasome activity, results from integrated transcriptome and pathway analysis suggest that DDI2 is involved in worsening cancer prognosis through the modulating transcriptional expression of multiple genes. Yet, whether DDI2's direct or indirect role in affecting these genes remains to be studied. To facilitate this, we must first have suitable biochemical assays for DDI2, establishing the importance of continuing to develop probes for DDI2.

## Bibliography

1. Calamini, B.; Morimoto, R. I., Protein homeostasis as a therapeutic target for diseases of protein conformation. *Curr. Top. Med. Chem.* **2012**, *12* (22), 2623-2640.
2. Tahmaz, I.; Shahmoradi, G. S.; Topf, U., Prefoldin Function in Cellular Protein Homeostasis and Human Diseases. *Frontiers in Cell and Developmental Biology* **2022**, *9*.
3. Buszczak, M.; Signer, R. A. J.; Morrison, S. J., Cellular differences in protein synthesis regulate tissue homeostasis. *Cell* **2014**, *159* (2), 242-251.
4. Shiber, A.; Döring, K.; Friedrich, U.; Klann, K.; Merker, D.; Zedan, M.; Tippmann, F.; Kramer, G.; Bukau, B., Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature* **2018**, *561* (7722), 268-272.
5. Deshaies, R. J., Proteotoxic crisis, the ubiquitin-proteasome system, and cancer therapy. *BMC Biol.* **2014**, *12* (1).
6. Hanahan, D.; Weinberg, R. A., The Hallmarks of Cancer. *Cell* **2000**, *100* (1), 57-70.
7. Oromendia, A. B.; Dodgson, S. E.; Amon, A., Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* **2012**, *26* (24), 2696-2708.
8. Padovani, C.; Jevtić, P.; Rapé, M., Quality control of protein complex composition. *Mol. Cell* **2022**, *82* (8), 1439-1450.
9. Ciechanover, A.; Kwon, Y. T., Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies. *Exp. Mol. Med.* **2015**, *47* (3), e147.
10. Thibaudeau, T. A.; Smith, D. M., A Practical Review of Proteasome Pharmacology. *Pharmacol. Rev.* **2019**, *71* (2), 170-197.
11. Papadopoulos, K. P.; Siegel, D. S.; Vesole, D. H.; Lee, P.; Rosen, S. T.; Zojwalla, N.; Holahan, J. R.; Lee, S.; Wang, Z.; Badros, A., Phase I Study of 30-Minute Infusion of Carfilzomib As Single Agent or in Combination With Low-Dose Dexamethasone in Patients With Relapsed and/or Refractory Multiple Myeloma. *J. Clin. Oncol.* **2015**, *33* (7), 732-739.
12. Weyburne, E. S.; Wilkins, O. M.; Sha, Z.; Williams, D. A.; Pletnev, A. A.; de Bruin, G.; Overkleeft, H. S.; Goldberg, A. L.; Cole, M. D.; Kisselev, A. F., Inhibition of the Proteasome  $\beta$ 2 Site Sensitizes Triple-Negative Breast Cancer Cells to  $\beta$ 5 Inhibitors and Suppresses Nrf1 Activation. *Cell Chem. Biol.* **2017**, *24*, 218.
13. Shabaneh, T. B.; Downey, S. L.; Goddard, A. L.; Screen, M.; Lucas, M. M.; Eastman, A.; Kisselev, A. F., Molecular Basis of Differential Sensitivity of Myeloma Cells to Clinically Relevant Bolus Treatment with Bortezomib. *PLoS One* **2013**, *8* (2), e56132.
14. Radhakrishnan, S. K.; Lee, C. S.; Young, P.; Beskow, A.; Chan, J. Y.; Deshaies, R. J., Transcription Factor Nrf1 Mediates the Proteasome Recovery Pathway after Proteasome Inhibition in Mammalian Cells. *Mol. Cell* **2010**, *38* (1), 17-28.
15. Koizumi, S.; Irie, T.; Hirayama, S.; Sakurai, Y.; Yashiroda, H.; Naguro, I.; Ichijo, H.; Hamazaki, J.; Murata, S., The aspartyl protease DD12 activates Nrf1 to compensate for proteasome dysfunction. *eLife* **2016**, *5*, e18357.
16. Sivá, M.; Svoboda, M.; Veverka, V.; Trempe, J.-F.; Hofmann, K.; Kožíšek, M. n.; Hexnerová, R.; Sedlák, F.; Belza, J.; Brynda, J.; Šácha, P.; Hubálek, M.; Starková, J.; Flaisigová, I.; Konvalinka, J.; Šašková, K. G., Human DNA-Damage-Inducible 2 Protein Is Structurally and Functionally Distinct from Its Yeast Ortholog. *Sci. Rep.* **2016**, *6* (1), 30443.
17. Dirac-Svejstrup, A. B.; Walker, J.; Faull, P.; Encheva, V.; Akimov, V.; Puglia, M.; Perkins, D.; Kümper, S.; Hunjan, S. S.; Blagoev, B.; Snijders, A. P.; Powell, D. J.;

- Svejstrup, J. Q., DDI2 Is a Ubiquitin-Directed Endoprotease Responsible for Cleavage of Transcription Factor NRF1. *Mol. Cell* **2020**, *79* (2), 332-341.e7.
18. Ciechanover, A.; Kwon, Y. T., Protein Quality Control by Molecular Chaperones in Neurodegeneration. *Front. Neurosci.* **2017**, *11*, 185-185.
  19. Tanaka, K., The proteasome: overview of structure and functions. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **2009**, *85* (1), 12-36.
  20. Siva, M.; Haberecht-Müller, S.; Prochazkova, M.; Prochazka, J.; Sedlak, F.; Chawengsaksophak, K.; Kasperek, P.; Sedlacek, R.; Konvalinka, J.; Krüger, E.; Saskova, K. G., DDI2 protease activity controls embryonic development and inflammation via TCF11/NRF1. *bioRxiv* **2020**, 2020.12.16.423023.
  21. Kottemann, M. C.; Conti, B. A.; Lach, F. P.; Smogorzewska, A., Removal of RTF2 from Stalled Replisomes Promotes Maintenance of Genome Integrity. *Mol. Cell* **2018**, *69* (1), 24-35.e5.
  22. Svoboda, M.; Konvalinka, J.; Trempe, J. F.; Grantz Saskova, K., The yeast proteases Ddi1 and Wss1 are both involved in the DNA replication stress response. *DNA Repair (Amst)* **2019**, *80*, 45-51.
  23. Serbyn, N.; Noireterre, A.; Bagdiul, I.; Plank, M.; Michel, A.; Loewith, R.; Kornmann, B.; Stutz, F., The Aspartic Protease Ddi1 Contributes to DNA-Protein Crosslink Repair in Yeast. *Mol. Cell* **2020**, *77*.
  24. Kisselev, A. F.; Akopian, T. N.; Woo, K. M.; Goldberg, A. L., The Sizes of Peptides Generated from Protein by Mammalian 26 and 20 S Proteasomes: IMPLICATIONS FOR UNDERSTANDING THE DEGRADATIVE MECHANISM AND ANTIGEN PRESENTATION \*. *J. Biol. Chem.* **1999**, *274* (6), 3363-3371.
  25. Glickman, M. H.; Ciechanover, A., The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiol. Rev.* **2002**, *82* (2), 373-428.
  26. Ciechanover, A., The ubiquitin-proteasome proteolytic pathway. *Cell* **1994**, *79* (1), 13-21.
  27. Hershko, A.; Ciechanover, A.; Heller, H.; Haas, A. L.; Rose, I. A., Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77* (4), 1783-6.
  28. Goldberg, A. L., Development of proteasome inhibitors as research tools and cancer drugs. *J. Cell Biol.* **2012**, *199* (4), 583-588.
  29. Koizumi, S.; Hamazaki, J.; Murata, S., Transcriptional regulation of the 26S proteasome by Nrf1. *Proceedings of the Japan Academy. Series B, Physical and biological sciences* **2018**, *94* (8), 325-336.
  30. Groll, M.; Bajorek, M.; Köhler, A.; Moroder, L.; Rubin, D. M.; Huber, R.; Glickman, M. H.; Finley, D., A gated channel into the proteasome core particle. *Nat. Struct. Biol.* **2000**, *7* (11), 1062-1067.
  31. Livneh, I.; Cohen-Kaplan, V.; Cohen-Rosenzweig, C.; Avni, N.; Ciechanover, A., The life cycle of the 26S proteasome: from birth, through regulation and function, and onto its death. *Cell Res.* **2016**, *26* (8), 869-885.
  32. Rousseau, A.; Bertolotti, A., Regulation of proteasome assembly and activity in health and disease. *Nature Reviews Molecular Cell Biology* **2018**, *19* (11), 697-712.
  33. Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A., Cancer statistics, 2022. *CA Cancer J. Clin.* **2022**, *72* (1), 7-33.
  34. Rajkumar, S. V., Multiple myeloma: 2022 update on diagnosis, risk stratification, and management. *Am. J. Hematol.* **2022**.
  35. Laubach, J.; Richardson, P.; Anderson, K., Multiple Myeloma. *Annu. Rev. Med.* **2011**, *62* (1), 249-264.
  36. Anderson, K. C.; Carrasco, R. D., Pathogenesis of Myeloma. *Annual Review of Pathology: Mechanisms of Disease* **2011**, *6* (1), 249-274.

37. Kisselev, A. F.; van der Linden, W. A.; Overkleeft, H. S., Proteasome inhibitors: an expanding army attacking a unique target. *Chem. Biol.* **2012**, *19*, 99.
38. Chen, T.; Ho, M.; Briere, J.; Moscvin, M.; Czarnecki, P. G.; Anderson, K. C.; Blackwell, T. K.; Bianchi, G., Multiple myeloma cells depend on the DDI2/NRF1-mediated proteasome stress response for survival. *Blood Advances* **2022**, *6* (2), 429-440.
39. Op, M.; Ribeiro, S. T.; Chavarria, C.; De Gassart, A.; Zaffalon, L.; Martinon, F., The aspartyl protease DDI2 drives adaptation to proteasome inhibition in multiple myeloma. *Cell Death Dis.* **2022**, *13* (5), 475.
40. Pérez-Galán, P.; Roué, G.; Villamor, N.; Montserrat, E.; Campo, E.; Colomer, D., The proteasome inhibitor bortezomib induces apoptosis in mantle-cell lymphoma through generation of ROS and Noxa activation independent of p53 status. *Blood* **2006**, *107* (1), 257-264.
41. Hambley, B.; Caimi, P. F.; William, B. M., Bortezomib for the treatment of mantle cell lymphoma: an update. *Therapeutic Advances in Hematology* **2016**, *7* (4), 196-208.
42. Reece, D.; Imrie, K.; Stevens, A.; Smith, C. A., Bortezomib in multiple myeloma and lymphoma: a systematic review and clinical practice guideline. *Curr. Oncol.* **2006**, *13* (5), 160-72.
43. Molineaux, S. M., Molecular Pathways: Targeting Proteasomal Protein Degradation in Cancer. *Clin. Cancer Res.* **2012**, *18* (1), 15-20.
44. Groll, M.; Berkers, C. R.; Ploegh, H. L.; Ovaa, H., Crystal Structure of the Boronic Acid-Based Proteasome Inhibitor Bortezomib in Complex with the Yeast 20S Proteasome. *Structure* **2006**, *14* (3), 451-456.
45. Kisselev, A. F.; Callard, A.; Goldberg, A. L., Importance of the Different Proteolytic Sites of the Proteasome and the Efficacy of Inhibitors Varies with the Protein Substrate. *J. Biol. Chem.* **2006**, *281* (13), 8582-8590.
46. Schwartz, R.; Davidson, T., Pharmacology, pharmacokinetics, and practical applications of bortezomib. *Oncology (Williston Park)* **2004**, *18* (14 Suppl 11), 14-21.
47. Hamilton, A. L.; Eder, J. P.; Pavlick, A. C.; Clark, J. W.; Liebes, L.; Garcia-Carbonero, R.; Chachoua, A.; Ryan, D. P.; Soma, V.; Farrell, K.; Kinchla, N.; Boyden, J.; Yee, H.; Zeleniuch-Jacquotte, A.; Wright, J.; Elliott, P.; Adams, J.; Muggia, F. M., Proteasome Inhibition With Bortezomib (PS-341): A Phase I Study With Pharmacodynamic End Points Using a Day 1 and Day 4 Schedule in a 14-Day Cycle. *J. Clin. Oncol.* **2005**, *23* (25), 6107-6116.
48. Kim, K. B.; Crews, C. M., From epoxomicin to carfilzomib: chemistry, biology, and medical outcomes. *Nat. Prod. Rep.* **2013**, *30* (5), 600-4.
49. Groll, M.; Kim, K. B.; Kairies, N.; Huber, R.; Crews, C. M., Crystal Structure of Epoxomicin:20S Proteasome Reveals a Molecular Basis for Selectivity of  $\alpha'$ , $\beta'$ -Epoxyketone Proteasome Inhibitors. *J. Am. Chem. Soc.* **2000**, *122*, 1237.
50. O'Connor, O. A.; Stewart, A. K.; Vallone, M.; Molineaux, C. J.; Kunkel, L. A.; Gerecitano, J. F.; Orlovski, R. Z., A Phase 1 Dose Escalation Study of the Safety and Pharmacokinetics of the Novel Proteasome Inhibitor Carfilzomib (PR-171) in Patients with Hematologic Malignancies. *Clin. Cancer Res.* **2009**, *15* (22), 7085-7091.
51. Siegel, D. S.; Martin, T.; Wang, M.; Vij, R.; Jakubowiak, A. J.; Lonial, S.; Trudel, S.; Kukreti, V.; Bahlis, N.; Alsina, M.; Chanan-Khan, A.; Buadi, F.; Reu, F. J.; Somlo, G.; Zonder, J.; Song, K.; Stewart, A. K.; Stadtmauer, E.; Kunkel, L.; Wear, S.; Wong, A. F.; Orlovski, R. Z.; Jagannath, S., A phase 2 study of single-agent carfilzomib (PX-171-003-A1) in patients with relapsed and refractory multiple myeloma. *Blood* **2012**, *120* (14), 2817-2825.

52. Siegel, D. S., From clinical trials to clinical practice: single-agent carfilzomib adverse events and their management in patients with relapsed and/or refractory multiple myeloma. *Therapeutic Advances in Hematology* **2013**, *4* (6), 354-365.
53. Yang, J.; Wang, Z.; Fang, Y.; Jiang, J.; Zhao, F.; Wong, H.; Bennett, M. K.; Molineaux, C. J.; Kirk, C. J., Pharmacokinetics, Pharmacodynamics, Metabolism, Distribution, and Excretion of Carfilzomib in Rats. *Drug Metab. Disposition* **2011**, *39* (10), 1873-1882.
54. Raedler, L. A., Ninlaro (Ixazomib): First Oral Proteasome Inhibitor Approved for the Treatment of Patients with Relapsed or Refractory Multiple Myeloma. *Am Health Drug Benefits* **2016**, *9* (Spec Feature), 102-5.
55. Kumar, S. K.; LaPlant, B.; Roy, V.; Reeder, C. B.; Lacy, M. Q.; Gertz, M. A.; Laumann, K.; Thompson, M. A.; Witzig, T. E.; Buadi, F. K.; Rivera, C. E.; Mikhael, J. R.; Bergsagel, P. L.; Kapoor, P.; Hwa, L.; Fonseca, R.; Stewart, A. K.; Chanan-Khan, A.; Rajkumar, S. V.; Dispenzieri, A., Phase 2 trial of ixazomib in patients with relapsed multiple myeloma not refractory to bortezomib. *Blood Cancer Journal* **2015**, *5* (8), e338-e338.
56. Richardson, P. G.; Baz, R.; Wang, M.; Jakubowiak, A. J.; Laubach, J. P.; Harvey, R. D.; Talpaz, M.; Berg, D.; Liu, G.; Yu, J.; Gupta, N.; Di Bacco, A.; Hui, A.-M.; Lonial, S., Phase 1 study of twice-weekly ixazomib, an oral proteasome inhibitor, in relapsed/refractory multiple myeloma patients. *Blood* **2014**, *124* (7), 1038-1046.
57. Kupperman, E.; Lee, E. C.; Cao, Y.; Bannerman, B.; Fitzgerald, M.; Berger, A.; Yu, J.; Yang, Y.; Hales, P.; Bruzzese, F.; Liu, J.; Blank, J.; Garcia, K.; Tsu, C.; Dick, L.; Fleming, P.; Yu, L.; Manfredi, M.; Rolfe, M.; Bolen, J., Evaluation of the Proteasome Inhibitor MLN9708 in Preclinical Models of Human Cancer. *Cancer Res.* **2010**, *70* (5), 1970-1980.
58. Xie, Y.; Varshavsky, A., RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: A negative feedback circuit. *Proceedings of the National Academy of Sciences* **2001**, *98* (6), 3056-3061.
59. Mannhaupt, G.; Schnall, R.; Karpov, V.; Vetter, I.; Feldmann, H., Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. *FEBS Lett.* **1999**, *450* (1-2), 27-34.
60. Hamazaki, J.; Murata, S., ER-Resident Transcription Factor Nrf1 Regulates Proteasome Expression and Beyond. *Int. J. Mol. Sci.* **2020**, *21* (10), 3683.
61. Li, X.; Matilainen, O.; Jin, C.; Glover-Cutter, K. M.; Holmberg, C. I.; Blackwell, T. K., Specific SKN-1/Nrf Stress Responses to Perturbations in Translation Elongation and Proteasome Activity. *PLoS Genet.* **2011**, *7* (6), e1002119.
62. Xu, H.; Fu, J.; Ha, S.-W.; Ju, D.; Zheng, J.; Li, L.; Xie, Y., The CCAAT box-binding transcription factor NF-Y regulates basal expression of human proteasome genes. *Biochim. Biophys. Acta* **2012**, *1823* (4), 818-825.
63. Vilchez, D.; Boyer, L.; Morantte, L.; Lutz, M.; Merkwirth, C.; Joyce, D.; Spencer, B.; Page, L.; Masliah, E.; Berggren, W. T.; Gage, F. H.; Dillin, A., Increased proteasome activity in human embryonic stem cells is regulated by PSMD11. *Nature* **2012**, *489* (7415), 304-308.
64. Vilchez, D.; Morantte, L.; Liu, Z.; Douglas, P. M.; Merkwirth, C.; Rodrigues, A. P. C.; Manning, G.; Dillin, A., RPN-6 determines *C. elegans* longevity under proteotoxic stress conditions. *Nature* **2012**, *489* (7415), 263-268.
65. Radhakrishnan, S. K.; den Besten, W.; Deshaies, R. J., p97-dependent retrotranslocation and proteolytic processing govern formation of active Nrf1 upon proteasome inhibition. *eLife* **2014**, *3*, e01856.
66. Sekine, H.; Motohashi, H., Roles of CNC Transcription Factors NRF1 and NRF2 in Cancer. *Cancers (Basel)* **2021**, *13* (3), 541.
67. Chowdhury, A. M. M. A.; Katoh, H.; Hatanaka, A.; Iwanari, H.; Nakamura, N.; Hamakubo, T.; Natsume, T.; Waku, T.; Kobayashi, A., Multiple regulatory

- mechanisms of the biological function of NRF3 (NFE2L3) control cancer cell proliferation. *Sci. Rep.* **2017**, *7* (1), 12494.
68. Kannan, M. B.; Dodard-Friedman, I.; Blank, V., Stringent Control of NFE2L3 (Nuclear Factor, Erythroid 2-Like 3; NRF3) Protein Degradation by FBW7 (F-box/WD Repeat-containing Protein 7) and Glycogen Synthase Kinase 3 (GSK3) \*. *J. Biol. Chem.* **2015**, *290* (43), 26292-26302.
  69. Han, J. W.; Valdez, J. L.; Ho, D. V.; Lee, C. S.; Kim, H. M.; Wang, X.; Huang, L.; Chan, J. Y., Nuclear factor-erythroid-2 related transcription factor-1 (Nrf1) is regulated by O-GlcNAc transferase. *Free Radical Biol. Med.* **2017**, *110*, 196-205.
  70. Kobayashi, A.; Waku, T., New addiction to the NRF2-related factor NRF3 in cancer cells: ubiquitin-independent proteolysis through the 20S proteasome. *Cancer Sci.* **2020**, *111* (1), 6-14.
  71. Vangala, J. R.; Sotzny, F.; Krüger, E.; Deshaies, R. J.; Radhakrishnan, S. K., Nrf1 can be processed and activated in a proteasome-independent manner. *Curr. Biol.* **2016**, *26* (18), R834-r835.
  72. Steffen, J.; Seeger, M.; Koch, A.; Krüger, E., Proteasomal Degradation Is Transcriptionally Controlled by TCF11 via an ERAD-Dependent Feedback Loop. *Mol. Cell* **2010**, *40* (1), 147-158.
  73. Waku, T.; Nakamura, N.; Koji, M.; Watanabe, H.; Katoh, H.; Tatsumi, C.; Tamura, N.; Hatanaka, A.; Hirose, S.; Katayama, H.; Tani, M.; Kubo, Y.; Hamazaki, J.; Hamakubo, T.; Watanabe, A.; Murata, S.; Kobayashi, A., NRF3-POMP-20S proteasome assembly axis promotes cancer development via ubiquitin-independent proteolysis of p53 and Rb. *Mol. Cell. Biol.* **2020**, MCB.00597-19.
  74. Kobayashi, A., Roles of NRF3 in the hallmarks of cancer: proteasomal inactivation of tumor suppressors. *Cancers (Basel)* **2020**, *12* (9), 2681.
  75. Waku, T.; Katayama, H.; Hiraoka, M.; Hatanaka, A.; Nakamura, N.; Tanaka, Y.; Tamura, N.; Watanabe, A.; Kobayashi, A., NRF1 and NRF3 complementarily maintain a basal proteasome activity in cancer cells through CPEB3-mediated translational repression. *bioRxiv* **2020**, 2020.01.10.902718.
  76. Aono, S.; Hatanaka, A.; Hatanaka, A.; Gao, Y.; Hippo, Y.; Taketo, M. M.; Waku, T.; Kobayashi, A.,  $\beta$ -Catenin/TCF4 Complex-Mediated Induction of the NRF3 (NFE2L3) Gene in Cancer Cells. *Int. J. Mol. Sci.* **2019**, *20* (13), 3344.
  77. Bury, M.; Le Calvé, B.; Lessard, F.; Dal Maso, T.; Saliba, J.; Michiels, C.; Ferbeyre, G.; Blank, V., NFE2L3 Controls Colon Cancer Cell Growth through Regulation of DUX4, a CDK1 Inhibitor. *Cell Reports* **2019**, *29* (6), 1469-1481.e9.
  78. Wang, H.; Zhan, M.; Yang, R.; Shi, Y.; Liu, Q.; Wang, J., Elevated expression of NFE2L3 predicts the poor prognosis of pancreatic cancer patients. *Cell cycle (Georgetown, Tex.)* **2018**, *17* (17), 2164-2174.
  79. Chan, J. Y.; Kwong, M.; Lu, R.; Chang, J.; Wang, B.; Yen, T. S. B.; Kan, Y. W., Targeted disruption of the ubiquitous CNC-bZIP transcription factor, Nrf-1, results in anemia and embryonic lethality in mice. *The EMBO Journal* **1998**, *17* (6), 1779-1787.
  80. Leung, L.; Kwong, M.; Hou, S.; Lee, C.; Chan, J. Y., Deficiency of the Nrf1 and Nrf2 Transcription Factors Results in Early Embryonic Lethality and Severe Oxidative Stress \*. *J. Biol. Chem.* **2003**, *278* (48), 48021-48029.
  81. Wang, W.; Chan, J. Y., Nrf1 is targeted to the endoplasmic reticulum membrane by an N-terminal transmembrane domain. Inhibition of nuclear translocation and transacting function. *J. Biol. Chem.* **2006**, *281* (28), 19676-87.
  82. Tomlin, F. M.; Gerling-Driessen, U. I. M.; Liu, Y.-C.; Flynn, R. A.; Vangala, J. R.; Lentz, C. S.; Clauder-Muenster, S.; Jakob, P.; Mueller, W. F.; Ordoñez-Rueda, D.; Paulsen, M.; Matsui, N.; Foley, D.; Rafalko, A.; Suzuki, T.; Bogoyo, M.; Steinmetz, L. M.; Radhakrishnan, S. K.; Bertozzi, C. R., Inhibition of NGLY1 Inactivates the



- Transcription Factor Nrf1 and Potentiates Proteasome Inhibitor Cytotoxicity. *ACS Central Science* **2017**, *3* (11), 1143-1155.
83. Tian, W.; Rojo de la Vega, M.; Schmidlin, C. J.; Ooi, A.; Zhang, D. D., Kelch-like ECH-associated protein 1 (KEAP1) differentially regulates nuclear factor erythroid-2-related factors 1 and 2 (NRF1 and NRF2). *The Journal of biological chemistry* **2018**, *293* (6), 2029-2040.
  84. Sekine, H.; Okazaki, K.; Kato, K.; Alam, M. M.; Shima, H.; Katsuoka, F.; Tsujita, T.; Suzuki, N.; Kobayashi, A.; Igarashi, K.; Yamamoto, M.; Motohashi, H., O-GlcNAcylation Signal Mediates Proteasome Inhibitor Resistance in Cancer Cells by Stabilizing NRF1. *Mol. Cell. Biol.* **2018**, *38* (17), e00252-18.
  85. Vangala, J. R.; Radhakrishnan, S. K., Nrf1-mediated transcriptional regulation of the proteasome requires a functional TIP60 complex. *J. Biol. Chem.* **2019**, *294* (6), 2036-2045.
  86. Sha, Z.; Goldberg, A. L., Proteasome-mediated processing of Nrf1 is essential for coordinate induction of all proteasome subunits and p97. *Current biology : CB* **2014**, *24* (14), 1573-1583.
  87. Sha, Z.; Goldberg, A. L., Reply to Vangala et al.: Complete inhibition of the proteasome reduces new proteasome production by causing Nrf1 aggregation. *Curr. Biol.* **2016**, *26* (18), R836-R837.
  88. Sirkis, R.; Gerst, J. E.; Fass, D., Ddi1, a Eukaryotic Protein With the Retroviral Protease Fold. *J. Mol. Biol.* **2006**, *364* (3), 376-387.
  89. Yip, M. C. J.; Bodnar, N. O.; Rapoport, T. A., Ddi1 is a ubiquitin-dependent protease. *Proceedings of the National Academy of Sciences* **2020**, *117* (14), 7776.
  90. Nowicka, U.; Zhang, D.; Walker, O.; Krutauz, D.; Castañeda, C. A.; Chaturvedi, A.; Chen, T. Y.; Reis, N.; Glickman, M. H.; Fushman, D., DNA-damage-inducible 1 protein (Ddi1) contains an uncharacteristic ubiquitin-like domain that binds ubiquitin. *Structure (London, England : 1993)* **2015**, *23* (3), 542-557.
  91. White, R. E.; Dickinson, J. R.; Semple, C. A. M.; Powell, D. J.; Berry, C., The retroviral proteinase active site and the N-terminus of Ddi1 are required for repression of protein secretion. *FEBS Lett.* **2011**, *585* (1), 139-142.
  92. Perteguer, M. J.; Gómez-Puertas, P.; Cañavate, C.; Dagger, F.; Gárate, T.; Valdivieso, E., Ddi1-like protein from *Leishmania major* is an active aspartyl proteinase. *Cell Stress Chaperones* **2013**, *18* (2), 171-81.
  93. White, R. E.; Powell, D. J.; Berry, C., HIV proteinase inhibitors target the Ddi1-like protein of *Leishmania* parasites. *FASEB J.* **2011**, *25* (5), 1729-36.
  94. Northrop, A.; Vangala, J. R.; Feygin, A.; Radhakrishnan, S. K., Disabling the Protease DDI2 Attenuates the Transcriptional Activity of NRF1 and Potentiates Proteasome Inhibitor Cytotoxicity. *Int. J. Mol. Sci.* **2020**, *21* (1), 327.
  95. Collins, G. A.; Sha, Z.; Kuo, C.-L.; Erbil, B.; Goldberg, A. L., Mammalian Ddi2 is a shuttling factor containing a retroviral protease domain that influences binding of ubiquitylated proteins and proteasomal degradation. *J. Biol. Chem.* **2022**, *298* (5).
  96. Bertolaet, B. L.; Clarke, D. J.; Wolff, M.; Watson, M. H.; Henze, M.; Divita, G.; Reed, S. I., UBA domains mediate protein-protein interactions between two DNA damage-inducible proteins. *J. Mol. Biol.* **2001**, *313* (5), 955-63.
  97. Díaz-Martínez, L. A.; Kang, Y.; Walters, K. J.; Clarke, D. J., Yeast UBL-UBA proteins have partially redundant functions in cell cycle control. *Cell Div.* **2006**, *1*, 28.
  98. Mitsiades, N.; Mitsiades, C. S.; Poulaki, V.; Chauhan, D.; Fanourakis, G.; Gu, X.; Bailey, C.; Joseph, M.; Libermann, T. A.; Treon, S. P.; Munshi, N. C.; Richardson, P. G.; Hideshima, T.; Anderson, K. C., Molecular sequelae of proteasome inhibition in human multiple myeloma cells. *Proceedings of the National Academy of Sciences* **2002**, *99* (22), 14374-14379.

99. Meiners, S.; Heyken, D.; Weller, A.; Ludwig, A.; Stangl, K.; Kloetzel, P.-M.; Krüger, E., Inhibition of Proteasome Activity Induces Concerted Expression of Proteasome Genes and de Novo Formation of Mammalian Proteasomes\*. *J. Biol. Chem.* **2003**, *278* (24), 21517-21525.
100. Beigl, T. B.; Kjosås, I.; Seljeseth, E.; Glomnes, N.; Aksnes, H., Efficient and crucial quality control of HAP1 cell ploidy status. *Biol Open* **2020**, *9* (11).
101. Essletzbichler, P.; Konopka, T.; Santoro, F.; Chen, D.; Gapp, B. V.; Kralovics, R.; Brummelkamp, T. R.; Nijman, S. M.; Bürckstümmer, T., Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* **2014**, *24* (12), 2059-65.
102. Kotecki, M.; Reddy, P. S.; Cochran, B. H., Isolation and characterization of a near-haploid human cell line. *Exp. Cell Res.* **1999**, *252* (2), 273-80.
103. Yin, Z.; Chen, L., Simple Meets Single: The Application of CRISPR/Cas9 in Haploid Embryonic Stem Cells. *Stem Cells Int* **2017**, *2017*, 2601746.
104. Chen, R.; Keating, M. J.; Gandhi, V.; Plunkett, W., Transcription inhibition by flavopiridol: mechanism of chronic lymphocytic leukemia cell death. *Blood* **2005**, *106* (7), 2513-9.
105. Bensaude, O., Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2011**, *2* (3), 103-108.
106. Ratnadiwakara, M.; Änkö, M. L., mRNA Stability Assay Using transcription inhibition by Actinomycin D in Mouse Pluripotent Stem Cells. *Bio Protoc* **2018**, *8* (21), e3072.
107. Zhang, B.; Gojo, I.; Fenton, R. G., Myeloid cell factor-1 is a critical survival factor for multiple myeloma. *Blood* **2002**, *99* (6), 1885-93.
108. Byrd, J. C.; Shinn, C.; Waselenko, J. K.; Fuchs, E. J.; Lehman, T. A.; Nguyen, P. L.; Flinn, I. W.; Diehl, L. F.; Sausville, E.; Grever, M. R., Flavopiridol Induces Apoptosis in Chronic Lymphocytic Leukemia Cells Via Activation of Caspase-3 Without Evidence of bcl-2 Modulation or Dependence on Functional p53. *Blood* **1998**, *92* (10), 3804-3816.
109. Takada, Y.; Aggarwal, B. B., Flavopiridol inhibits NF-kappaB activation induced by various carcinogens and inflammatory agents through inhibition of IkappaBalpha kinase and p65 phosphorylation: abrogation of cyclin D1, cyclooxygenase-2, and matrix metalloproteinase-9. *J. Biol. Chem.* **2004**, *279* (6), 4750-9.
110. Kleeff, J.; Kornmann, M.; Sawhney, H.; Korc, M., Actinomycin D induces apoptosis and inhibits growth of pancreatic cancer cells. *Int. J. Cancer* **2000**, *86* (3), 399-407.
111. Morita, M.; Alain, T.; Topisirovic, I.; Sonenberg, N., Polysome profiling analysis. *Bio-protocol* **2013**, *3* (14), e833-e833.
112. Sobhany, M.; Stanley, R. E., Polysome Profiling without Gradient Makers or Fractionation Systems. *JoVE* **2021**, (172), e62680.
113. He, S. L.; Green, R., Chapter Ten - Polysome Analysis of Mammalian Cells. In *Methods Enzymol.*, Lorsch, J., Ed. Academic Press: 2013; Vol. 530, pp 183-192.
114. King, H. A.; Gerber, A. P., Translatome profiling: methods for genome-scale analysis of mRNA translation. *Briefings in Functional Genomics* **2016**, *15* (1), 22-31.
115. Hoffman, L.; Pratt, G.; Rechsteiner, M., Multiple forms of the 20 S multicatalytic and the 26 S ubiquitin/ATP-dependent proteases from rabbit reticulocyte lysate. *J. Biol. Chem.* **1992**, *267* (31), 22362-8.
116. Kisselev, A. F.; Kaganovich, D.; Goldberg, A. L., Binding of hydrophobic peptides to several non-catalytic sites promotes peptide hydrolysis by all active sites of 20 S proteasomes. Evidence for peptide-induced channel opening in the alpha-rings. *J. Biol. Chem.* **2002**, *277* (25), 22260-70.
117. Li, C.; Wang, X.; Li, X.; Qiu, K.; Jiao, F.; Liu, Y.; Kong, Q.; Liu, Y.; Wu, Y., Proteasome Inhibition Activates Autophagy-Lysosome Pathway Associated With

- TFEB Dephosphorylation and Nuclear Translocation. *Frontiers in Cell and Developmental Biology* **2019**, *7*.
118. Roelofs, J.; Suppahia, A.; Waite, K. A.; Park, S., Native Gel Approaches in Studying Proteasome Assembly and Chaperones. *Methods Mol. Biol.* **2018**, *1844*, 237-260.
  119. Morozov, A.; Astakhova, T.; Erokhov, P.; Karpov, V., The ATP/Mg(2+) Balance Affects the Degradation of Short Fluorogenic Substrates by the 20S Proteasome. *Methods Protoc* **2022**, *5* (1).
  120. Verdoes, M.; Florea, B. I.; Menendez-Benito, V.; Maynard, C. J.; Witte, M. D.; van der Linden, W. A.; van den Nieuwendijk, A. M. C. H.; Hofmann, T.; Berkers, C. R.; van Leeuwen, F. W. B.; Groothuis, T. A.; Leeuwenburgh, M. A.; Ovaa, H.; Neeffjes, J. J.; Filippov, D. V.; van der Marel, G. A.; Dantuma, N. P.; Overkleeft, H. S., A Fluorescent Broad-Spectrum Proteasome Inhibitor for Labeling Proteasomes In Vitro and In Vivo. *Chem. Biol.* **2006**, *13* (11), 1217-1226.
  121. Welk, V.; Coux, O.; Kleene, V.; Abeza, C.; Trümbach, D.; Eickelberg, O.; Meiners, S., Inhibition of Proteasome Activity Induces Formation of Alternative Proteasome Complexes. *The Journal of biological chemistry* **2016**, *291* (25), 13147-13159.
  122. McShane, E.; Sin, C.; Zauber, H.; Wells, J. N.; Donnelly, N.; Wang, X.; Hou, J.; Chen, W.; Storchova, Z.; Marsh, J. A.; Valleriani, A.; Selbach, M., Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell* **2016**, *167* (3), 803-815.e21.
  123. Bianchi, G.; Oliva, L.; Cascio, P.; Pengo, N.; Fontana, F.; Cerruti, F.; Orsi, A.; Pasqualetto, E.; Mezghrani, A.; Calbi, V.; Palladini, G.; Giuliani, N.; Anderson, K. C.; Sitia, R.; Cenci, S., The proteasome load versus capacity balance determines apoptotic sensitivity of multiple myeloma cells to proteasome inhibition. *Blood* **2009**, *113* (13), 3040-3049.
  124. Britton, M.; Lucas, M. M.; Downey, S. L.; Screen, M.; Pletnev, A. A.; Verdoes, M.; Tokhunts, R. A.; Amir, O.; Goddard, A. L.; Pelphrey, P. M.; Wright, D. L.; Overkleeft, H. S.; Kisselev, A. F., Selective Inhibitor of Proteasome's Caspase-like Sites Sensitizes Cells to Specific Inhibition of Chymotrypsin-like Sites. *Chem. Biol.* **2009**, *16* (12), 1278-1289.
  125. Kumatori, A.; Tanaka, K.; Inamura, N.; Sone, S.; Ogura, T.; Matsumoto, T.; Tachikawa, T.; Shin, S.; Ichihara, A., Abnormally high expression of proteasomes in human leukemic cells. *Proceedings of the National Academy of Sciences* **1990**, *87* (18), 7071-7075.
  126. Dhatchinamoorthy, K.; Colbert, J. D.; Rock, K. L., Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. *Front. Immunol.* **2021**, *12*, 636568.
  127. Yoshihama, S.; Roszik, J.; Downs, I.; Meissner, T. B.; Vijayan, S.; Chapuy, B.; Sidiq, T.; Shipp, M. A.; Lizee, G. A.; Kobayashi, K. S., NLRC5/MHC class I transactivator is a target for immune evasion in cancer. *Proceedings of the National Academy of Sciences* **2016**, *113* (21), 5999-6004.
  128. Cornel, A. M.; Mimpen, I. L.; Nierkens, S., MHC Class I Downregulation in Cancer: Underlying Mechanisms and Potential Targets for Cancer Immunotherapy. *Cancers (Basel)* **2020**, *12* (7).
  129. Früh, K.; Yang, Y., Antigen presentation by MHC class I and its regulation by interferon  $\gamma$ . *Curr. Opin. Immunol.* **1999**, *11* (1), 76-81.
  130. Lee-Sayer, S. S. M.; Maeshima, N.; Dougan, M. N.; Dahiya, A.; Arif, A. A.; Dosanjh, M.; Maxwell, C. A.; Johnson, P., Hyaluronan-binding by CD44 reduces the memory potential of activated murine CD8 T cells. *Eur. J. Immunol.* **2018**, *48* (5), 803-814.
  131. Baaten, B.; Tinoco, R.; Chen, A.; Bradley, L., Regulation of Antigen-Experienced T Cells: Lessons from the Quintessential Memory Marker CD44. *Front. Immunol.* **2012**, *3*.
  132. Kong, T.; Ahn, R.; Yang, K.; Zhu, X.; Fu, Z.; Morin, G.; Bramley, R.; Cliffe, N. C.; Xue, Y.; Kuasne, H.; Li, Q.; Jung, S.; Gonzalez, A. V.; Camilleri-Broet, S.; Guiot,

- M.-C.; Park, M.; Ursini-Siegel, J.; Huang, S., CD44 Promotes PD-L1 Expression and Its Tumor-Intrinsic Function in Breast and Lung Cancers. *Cancer Res.* **2020**, *80* (3), 444-457.
133. Jiang, Y.; Chen, M.; Nie, H.; Yuan, Y., PD-1 and PD-L1 in cancer immunotherapy: clinical implications and future considerations. *Hum. Vaccin. Immunother.* **2019**, *15* (5), 1111-1122.
  134. Liao, C.; Wang, Q.; An, J.; Zhang, M.; Chen, J.; Li, X.; Xiao, L.; Wang, J.; Long, Q.; Liu, J.; Guan, X., SPINKs in Tumors: Potential Therapeutic Targets. *Front. Oncol.* **2022**, *12*, 833741.
  135. Tiwari, R.; Pandey, S. K.; Goel, S.; Bhatia, V.; Shukla, S.; Jing, X.; Dhanasekaran, S. M.; Ateeq, B., SPINK1 promotes colorectal cancer progression by downregulating Metallothioneins expression. *Oncogenesis* **2015**, *4* (8), e162.
  136. Zhang, C.; Amanda, S.; Wang, C.; King Tan, T.; Zulfaqar Ali, M.; Zhong Leong, W.; Moy Ng, L.; Kitajima, S.; Li, Z.; Eng Juh Yeoh, A.; Hao Tan, S.; Sanda, T., Oncorequisite role of an aldehyde dehydrogenase in the pathogenesis of T-cell acute lymphoblastic leukemia. *Haematologica* **2021**, *106* (6), 1545-1558.
  137. Mizuno, T.; Suzuki, N.; Makino, H.; Furui, T.; Morii, E.; Aoki, H.; Kunisada, T.; Yano, M.; Kuji, S.; Hirashima, Y.; Arakawa, A.; Nishio, S.; Ushijima, K.; Ito, K.; Itani, Y.; Morishige, K., Cancer stem-like cells of ovarian clear cell carcinoma are enriched in the ALDH-high population associated with an accelerated scavenging system in reactive oxygen species. *Gynecol. Oncol.* **2015**, *137* (2), 299-305.
  138. Yang, C. K.; Wang, X. K.; Liao, X. W.; Han, C. Y.; Yu, T. D.; Qin, W.; Zhu, G. Z.; Su, H.; Yu, L.; Liu, X. G.; Lu, S. C.; Chen, Z. W.; Liu, Z.; Huang, K. T.; Liu, Z. T.; Liang, Y.; Huang, J. L.; Xiao, K. Y.; Peng, M. H.; Winkle, C. A.; O'Brien, S. J.; Peng, T., Aldehyde dehydrogenase 1 (ALDH1) isoform expression and potential clinical implications in hepatocellular carcinoma. *PLoS One* **2017**, *12* (8), e0182208.
  139. Al-Hajj, M.; Wicha, M. S.; Benito-Hernandez, A.; Morrison, S. J.; Clarke, M. F., Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (7), 3983-8.
  140. Okamoto, I.; Kawano, Y.; Murakami, D.; Sasayama, T.; Araki, N.; Miki, T.; Wong, A. J.; Saya, H., Proteolytic release of CD44 intracellular domain and its role in the CD44 signaling pathway. *J. Cell Biol.* **2001**, *155* (5), 755-62.
  141. De Falco, V.; Tamburrino, A.; Ventre, S.; Castellone, M. D.; Malek, M.; Manié, S. N.; Santoro, M., CD44 proteolysis increases CREB phosphorylation and sustains proliferation of thyroid cancer cells. *Cancer Res.* **2012**, *72* (6), 1449-58.
  142. Tanei, T.; Morimoto, K.; Shimazu, K.; Kim, S. J.; Tanji, Y.; Taguchi, T.; Tamaki, Y.; Noguchi, S., Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential Paclitaxel and epirubicin-based chemotherapy for breast cancers. *Clin. Cancer Res.* **2009**, *15* (12), 4234-41.
  143. Kumar, H.; Mazumder, S.; Sharma, N.; Chakravarti, S.; Long, M. D.; Meurice, N.; Petit, J.; Liu, S.; Chesi, M.; Sanyal, S.; Stewart, A. K.; Kumar, S.; Bergsagel, L.; Rajkumar, S. V.; Baughn, L. B.; Van Ness, B. G.; Mitra, A. K., Single-Cell Proteomics and Tumor RNAseq Identify Novel Pathways Associated With Clofazimine Sensitivity in PI- and IMiD- Resistant Myeloma, and Putative Stem-Like Cells. *Front. Oncol.* **2022**, *12*.
  144. Rajagopalan, T. G.; Stein, W. H.; Moore, S., The Inactivation of Pepsin by Diazoacetyl norleucine Methyl Ester. *J. Biol. Chem.* **1966**, *241* (18), 4295-4297.
  145. Takahashi, K.; Mizobe, F.; Change, W. J., Inactivation of acid proteases from *Rhizopus chinensis*, *Aspergillus saitoi* and *Mucor pusillus*, and calf rennin by diazoacetyl norleucine methyl ester. *J. Biochem.* **1972**, *71* (1), 161-4.
  146. Mix, K. A.; Aronoff, M. R.; Raines, R. T., Diazo Compounds: Versatile Tools for Chemical Biology. *ACS Chem. Biol.* **2016**, *11* (12), 3233-3244.

147. Ong, E. B.; Perlmann, G. E., Specific inactivation of pepsin by benzyloxycarbonyl-L-phenylalanyldiazomethane. *Nature* **1967**, *215* (5109), 1492-4.
148. Delpierre, G. R.; Fruton, J. S., Specific Inactivation Of Pepsin by a Diazo Ketone. *Proceedings of the National Academy of Sciences* **1966**, *56* (6), 1817.
149. Lundblad, R. L.; Stein, W. H., On the reaction of diazoacetyl compounds with pepsin. *J. Biol. Chem.* **1969**, *244* (1), 154-60.
150. Delpierre, G. R.; Fruton, J. S., Inactivation of pepsin by diphenyldiazomethane. *Proc. Natl. Acad. Sci. U. S. A.* **1965**, *54* (4), 1161-1167.
151. Hamilton, G. A.; Spona, J.; Crowell, L. D., The inactivation of pepsin by an equimolar amount of 1-diazo-4-phenylbutanone-2. *Biochem. Biophys. Res. Commun.* **1967**, *26* (2), 193-198.
152. Stepanov, V. M.; Vaganova, T. I., Identification of the carboxyl group of pepsin reacting with diazoacetamide derivatives. *Biochem. Biophys. Res. Commun.* **1968**, *31* (5), 825-830.
153. Pettit, S. C.; Simsic, J.; Loeb, D. D.; Everitt, L.; Hutchison, C. A.; Swanstrom, R., Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. *J. Biol. Chem.* **1991**, *266* (22), 14539-14547.
154. Doscher, M. S.; Wilcox, P. E., Chemical derivatives of alpha-chymotrypsinogen IV. A comparison of the reactions of alpha-chymotrypsinogen and of simple carboxylic acids with diazoacetamide. *J. Biol. Chem.* **1961**, *236*, 1328-37.
155. Mix, K. A.; Raines, R. T., Optimized Diazo Scaffold for Protein Esterification. *Org. Lett.* **2015**, *17* (10), 2358-2361.
156. Maestro, v.-. *Schrödinger, LLC* **2019**.
157. de Freitas, R. F.; Schapira, M., A systematic analysis of atomic protein–ligand interactions in the PDB. *Medchemcomm* **2017**, *8* (10), 1970-1981.
158. Amblard, M.; Fehrentz, J. A.; Martinez, J.; Subra, G., Methods and protocols of modern solid phase Peptide synthesis. *Mol. Biotechnol.* **2006**, *33* (3), 239-54.
159. Thieriet, N.; Alsina, J.; Giral, E.; Guibé, F.; Albericio, F., Use of Alloc-amino acids in solid-phase peptide synthesis. Tandem deprotection-coupling reactions using neutral conditions. *Tetrahedron Lett.* **1997**, *38* (41), 7275-7278.
160. Castro, V.; Blanco-Canosa, J. B.; Rodriguez, H.; Albericio, F., Imidazole-1-sulfonyl Azide-Based Diazo-Transfer Reaction for the Preparation of Azido Solid Supports for Solid-Phase Synthesis. *ACS Combinatorial Science* **2013**, *15* (7), 331-334.
161. Marine, J. E.; Liang, X.; Song, S.; Rudick, J. G., Azide-rich peptides via an on-resin diazotransfer reaction. *Peptide Science* **2015**, *104* (4), 419-426.
162. Goddard-Borger, E. D.; Stick, R. V., An Efficient, Inexpensive, and Shelf-Stable Diazotransfer Reagent: Imidazole-1-sulfonyl Azide Hydrochloride. *Org. Lett.* **2007**, *9* (19), 3797-3800.
163. Myers, E. L.; Raines, R. T., A Phosphine-Mediated Conversion of Azides into Diazo Compounds. *Angew. Chem. Int. Ed.* **2009**, *48* (13), 2359-2363.
164. Chou, H.-H.; Raines, R. T., Conversion of Azides into Diazo Compounds in Water. *J. Am. Chem. Soc.* **2013**, *135* (40), 14936-14939.
165. Farrera-Sinfreu, J.; Royo, M.; Albericio, F., Undesired removal of the Fmoc group by the free ε-amino function of a lysine residue. *Tetrahedron Lett.* **2002**, *43* (43), 7813-7815.
166. Protection for the Amino Group. In *Greene's Protective Groups in Organic Synthesis*, 4th ed.; Wuts, P. G. M.; Greene, T. W., Eds. John Wiley & Sons, Inc.: Hoboken, NJ., 2006; pp 696-926.
167. Mix, K. A.; Lomax, J. E.; Raines, R. T., Cytosolic Delivery of Proteins by Bioreversible Esterification. *J. Am. Chem. Soc.* **2017**, *139* (41), 14396-14398.
168. El-Faham, A.; Albericio, F., Peptide Coupling Reagents, More than a Letter Soup. *Chem. Rev.* **2011**, *111* (11), 6557-6602.

169. Liebeschuetz, J. W.; Wylie, W. A.; Waszkowycz, B.; Murray, C. W.; Rimmer, A. D.; Welsh, P. M.; Jones, S. D.; Roscoe, J. M. E.; Young, S. C.; Morgan, P. J., Preparation of meta-benzamidine derivatives of amino acids or dipeptides as serine protease inhibitors. *Patent US20020055522* **2002**, 35.
170. Tong, G.; Lawlor, J. M.; Tregear, G. W.; Haralambidis, J., The synthesis of oligonucleotide-polyamide conjugate molecules suitable as PCR primers. *The Journal of Organic Chemistry* **1993**, *58* (8), 2223-2231.
171. Tytgat, H. L.; Schoofs, G.; Driesen, M.; Proost, P.; Van Damme, E. J.; Vanderleyden, J.; Lebeer, S., Endogenous biotin-binding proteins: an overlooked factor causing false positives in streptavidin-based protein detection. *Microb Biotechnol* **2015**, *8* (1), 164-8.
172. Zempleni, J.; Wijeratne, S. S.; Hassan, Y. I., Biotin. *Biofactors* **2009**, *35* (1), 36-46.
173. Kirkeby, S.; Moe, D.; Bøg-Hansen, T. C.; van Noorden, C. J., Biotin carboxylases in mitochondria and the cytosol from skeletal and cardiac muscle as detected by avidin binding. *Histochemistry* **1993**, *100* (6), 415-21.
174. van Dongen, S. F. M.; Teeuwen, R. L. M.; Nallani, M.; van Berkel, S. S.; Cornelissen, J. J. L. M.; Nolte, R. J. M.; van Hest, J. C. M., Single-Step Azide Introduction in Proteins via an Aqueous Diazo Transfer. *Bioconjugate Chem.* **2009**, *20* (1), 20-23.
175. Zaro, B. W.; Batt, A. R.; Chuh, K. N.; Navarro, M. X.; Pratt, M. R., The Small Molecule 2-Azido-2-deoxy-glucose Is a Metabolic Chemical Reporter of O-GlcNAc Modifications in Mammalian Cells, Revealing an Unexpected Promiscuity of O-GlcNAc Transferase. *ACS Chem. Biol.* **2017**, *12* (3), 787-794.
176. Vocadlo, D. J.; Hang, H. C.; Kim, E.-J.; Hanover, J. A.; Bertozzi, C. R., A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proceedings of the National Academy of Sciences* **2003**, *100* (16), 9116-9121.
177. Cenci, S.; Mezghrani, A.; Cascio, P.; Bianchi, G.; Cerruti, F.; Fra, A.; Lelouard, H.; Masciarelli, S.; Mattioli, L.; Oliva, L.; Orsi, A.; Pasqualetto, E.; Pierre, P.; Ruffato, E.; Tagliavacca, L.; Sitia, R., Progressively impaired proteasomal capacity during terminal plasma cell differentiation. *EMBO J.* **2006**, *25* (5), 1104-13.
178. Meister, S.; Schubert, U.; Neubert, K.; Herrmann, K.; Burger, R.; Gramatzki, M.; Hahn, S.; Schreiber, S.; Wilhelm, S.; Herrmann, M.; Jäck, H.-M.; Voll, R. E., Extensive Immunoglobulin Production Sensitizes Myeloma Cells for Proteasome Inhibition. *Cancer Res.* **2007**, *67* (4), 1783-1792.