

**Design and Synthesis of Inhibitors Targeting DCN1/3 and Inhibitors Targeting MLL1  
Methyltransferase**

by

Ting-Rong Chern

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Medicinal Chemistry)  
in the University of Michigan  
2019

Doctoral Committee:

Professor Shaomeng Wang, Chair  
Professor Yali Dou  
Professor Nouri Neamati  
Associate Professor Zaneta Nikolovska-Coleska

Ting-Rong Chern

trchern@umich.edu

ORCID iD: 0000-0002-8744-6368

© Ting-Rong Chern 2019

## **Dedication**

To Mom, Dad, Yu-Hsuan, and Riley

## **Acknowledgements**

First and foremost, I would like to thank Professor Shaomeng Wang for his support and guidance throughout my graduate career. I would also like to thank my committee members, Professor Nouri Neamati, Professor Yali Dou, and Associate Professor Zaneta Nikolovska-Coleska for their advice and guidance. Last but not least, I am deeply thankful for my fellow Wang group graduate students, postdocs, and all the other brilliant scientists I had the honor of working with.

## Table of Contents

Dedication .....	ii
Acknowledgements .....	iii
List of Tables .....	ix
List of Figures.....	x
List of Appendices .....	xii
PREFACE .....	xiii
Chapter 1 Introduction.....	1
1.1 Introduction to the Neddylation Pathway.....	1
1.1.1 Neddylation Pathway.....	1
1.1.2 Cullin-RING Ligase .....	2
1.1.3 Current strategies targeting the neddylation pathway .....	3
1.2 Introduction to MLL methyltransferase .....	4
1.2.1 MLL1 methyltransferase and its relevance to disease .....	4
1.2.2 Current strategies targeting the MLL1 methyltransferase activity .....	5
1.2.3 SAM Analogs as methyltransferase inhibitors.....	6
1.2.3.1 SAM cycle and its role in biology .....	6
1.2.3.2 Early studies on SAM analogs targeting methyltransferases .....	8

1.2.3.3	The bisubstrate strategy to design SAM based analogs targeting methyltransferases .....	9
1.3	Overview of Thesis .....	11
Chapter 2 Development of Peptidomimetic Inhibitors of the DCN1-UBC12 Protein-Protein Interaction .....		12
2.1	Introduction .....	12
2.2	Results and Discussion .....	14
2.2.1	Identification of a 4-residue lead peptide .....	14
2.2.2	Development of a fluorescence polarization (FP)-based competitive binding assay 16	
2.2.3	Modification of the N-terminal acetyl group, interacting with the P2 pocket.....	17
2.2.4	Modifications at the C-terminus probing interactions with the P4 pocket.....	19
2.2.5	Modifications of Ile2 to probe interactions with the P3 pocket .....	21
2.2.6	Investigations of the interactions with the P1 core pocket.....	22
2.2.7	Optimization of the physicochemical properties of compound 33.....	23
2.2.8	Determination of the crystal structure of compound 38 in a complex with DCN1..	26
2.2.9	Cellular evaluation of compound 36 (DI-404).....	28
2.3	Chemistry .....	31
2.4	Summary .....	35
Chapter 3 Discovery of Potent and Selective DCN3 Inhibitors.....		37
3.1	Introduction .....	37

3.2	RESULTS AND DISCUSSION .....	39
3.2.1	Analysis of the binding pockets in DCN1 and DCN3 .....	39
3.2.2	Identification of initial DCN3 inhibitors by screening our DCN1 inhibitor library ....	41
3.2.3	Computational prediction of the binding of lead compounds in complexes with DCN3 43	
3.2.4	Development of a fluorescence polarization-based, competitive binding assay for DCN3	44
3.2.5	Structure-activity relationship studies of the P1 binding pocket .....	49
3.2.6	Structure-activity relationship studies of the P2 binding pocket .....	50
3.2.7	Structure-activity relationship studies of the P3 binding pocket .....	53
3.2.8	Computational docking study of (TC-6304) .....	56
3.2.9	Evaluation of the interaction of TC-6304 with DCN3 protein in cells.....	58
3.2.10	Evaluation of TC-6304 for its effect on neddylation of cullin E3 ligases in cells ..	59
3.3	Chemistry .....	61
3.4	SUMMARY .....	64
 Chapter 4 Discovery of Potent Small-Molecule Inhibitors of Mixed-Lineage Leukemia (MLL) Methyltransferase .....		
4.1	Introduction .....	66
4.2	Results and Discussion .....	68
4.2.1	Design of a Focused Chemical Library Based upon SAM .....	68
4.2.2	Identification of MLL Inhibitors by Screening the Focused Chemical Library .....	69

4.2.3	Validation of the Screening Hits .....	70
4.2.4	Further Modifications of the Most Potent Compound 5.....	73
4.2.5	Determination of Co-Crystal Structures of MLL Inhibitors in a complex with MLL1 SET Domain.....	74
4.3	Chemistry .....	76
4.4	Summary .....	77
Chapter 5 Conclusion and Future directions .....		78
5.1	Conclusion.....	78
5.2	Future directions.....	81
5.2.1	The DCN project.....	81
5.2.1.1	Screening against other proteins in the DCN family.....	81
5.2.1.2	Targeting other protein-protein integrations among different components in the CRL complex.....	82
5.2.2	The MLL1 methyltransferase project.....	83
5.2.2.1	Structural modifications of current MLL1 inhibitors .....	83
5.2.2.1.1	Improving the physical chemical properties.....	83
5.2.2.1.2	Targeting the lysine tunnel and the substrate binding site .....	84
5.2.2.1.3	Design of MLL1 degraders (PROTAC) .....	85
5.2.2.1.4	Conformation restriction .....	85
5.2.2.2	Screening of current SAM-based library for other methyltransferases .....	86
5.2.2.3	Test current MLL1 SET domain inhibitors against a panel of HMTs. ....	86



Appendices .....	88
Bibliography .....	148

## List of Tables

Table 2.1 Structures and binding affinities of compound 3 and its analogues designed to investigate the interactions with the P2 pocket in DCN1. ....	18
Table 2.2 Structures and binding affinities of compound 7 and its analogues, designed to investigate the P4 pocket in DCN1.....	19
Table 2.3 Structures and binding affinities of compound 17 and its analogues designed to probe the interactions with the P3 pocket in DCN1. ....	22
Table 2.4 Structures and binding affinities of compound 22 and its analogues designed to investigate the P1 core pocket in DCN1.....	23
Table 3.1 Binding affinities of UBE2F peptides to DCN3 as determined in a direct BLI binding assay. ....	42
Table 3.2 Binding Affinity of Mutated UBE2F Peptides and the Tracers Determined in BLI Binding Assay .....	47
Table 3.3 Structure-activity relationship studies to probe the P1 binding pocket.....	50
Table 3.5 Structure-activity relationship studies to probe the P2 binding pocket.....	51
Table 3.6 Structure-activity relationship studies to probe the P3 binding pocket.....	54
Table 4.1 Results of hit validation using Alpha LISA assay for MLL1 methyltransferase. ....	72
Table 4.2. SAR of subset IV compounds.....	73
Table A-1 Crystallography Data Collection and Refinement Statistics .....	116

## List of Figures

Figure 1.1 The neddylation pathway and the activation of CRLs .....	2
Figure 1.2 Inhibitors targeting the neddylation pathway .....	4
Figure 1.3 WDR5 inhibitors .....	6
Figure 1.4 The SAM cycle .....	7
Figure 1.5 Summary of SAR on modifying SAH for methyltransferases.....	9
Figure 1.6 SAM analog inhibitors targeting histone methyltransferases.....	10
Figure 2.1 Proposed model of activation of CUL3 CRL and the model of inactive CUL3 CRL blocked by inhibitor .....	13
Figure 2.2 Chemical structures of designed peptides and the structural analysis of peptides-DCN1 interaction .....	15
Figure 2.3 Chemical structure of compound 4. B) Binding isotherm of compound 4 to DCN1. ....	17
Figure 2.4 Binding mode of compound 7 and 17 in DCN1 .....	20
Figure 2.5 Optimization of compound 33 and the Kinetic binding sensograms .....	25
Figure 2.6 Co-crystal structure of DCN1 complexed with compound 38. ....	26
Figure 2.7 Engagement of compound 36 with DCN1 in cells. ....	28
Figure 2.8 Effect of compound 36 on DCN1-UBC12 protein-protein interaction in H2170 cells .....	29

Figure 2.9 Effect of cullin neddylation of compound 36 in H2170 cells. ....	30
Figure 3.1 Crystal structure analysis of DCN1 and DCN3.....	40
Figure 3.2 Chemical structures and binding affinities of previously synthesized DCN1 inhibitors. Compounds 1-3 display significant binding affinities to DCN3. ....	42
Figure 3.3 Modeling of DCN inhibitors on DCN1 and DCN3 proteins. ....	44
Figure 3.4 Modeled truncated UBE2F with T3R mutation with DCN3.....	48
Figure 3.5 Binding isotherm of DCN3-Flu1 to DCN3 protein.....	49
Figure 3.6 Computatioanl Docking of TC-6304 with DCN1/3.....	57
Figure 3.7 The cellular thermal shift assay.....	58
Figure 3.8 . Evaluation of the effect of a selective DCN3 inhibitor TG-6304 on neddylation of different cullins in cell. ....	60
Figure 4.1 Mechanism of lysine methylation and the Designed focused library. ....	69
Figure 4.2 Result of Single Dose Screening against MLL1 Methyltransferase.....	70
Figure 4.3 Crystal structures of inhibitor and MLL1 methyltransferase.....	75
Figure 5.1 Analysis of the protein-protein interactions in the CUL1 CRL complex .....	82
Figure 5.2 Further modifications of MLL1 inhibitors .....	83
Figure 5.3 An SMYD3 inhibitor, GSK2807 .....	86

## List of Appendices

Appendix A: Experimental Section of Chapter 2 .....	89
Appendix B: Experimental Section of Chapter 3 .....	118
Appendix C: Experimental Section of Chapter 4 .....	138

## Preface

The goal of my work presented in this dissertation was to fill the unmet need for compounds that can be used to interrogate the biology of the Cullin-RING ubiquitin ligases (CRLs) and the MLL1 methyltransferase. We conducted structure-based design, organic synthesis, and biochemical/biophysical assays to discover potent inhibitors targeting DCN1, DCN3 and, MLL1 methyltransferase.

Chapter 2 describes the design, synthesis and evaluation of peptidomimetics targeting the DCN1-UBC12 protein-protein interaction. Starting from a 12-residue UBC12 peptide, we successfully obtained a series of peptidomimetic compounds that bind to DCN1 protein with  $K_D$  values of  $<10$  nM. Determination of a co-crystal structure of a potent peptidomimetic inhibitor complexed with DCN1 provided the structural basis for their high-affinity interaction. Cellular investigation of one potent DCN1 inhibitor, compound **DI-404**, reveals that it effectively and selectively inhibits the neddylation of cullin 3 over other cullin members. Further optimization of **DI-404** may yield a new class of therapeutics for the treatment of human diseases in which cullin 3 CRL plays a key role. This work has been published in the *Journal of Medicinal Chemistry* **2018**, 61, 1934–1950.

In Chapter 3, we report the discovery of first-in-class, small-molecule inhibitors targeting the DCN3-UBE2F interaction, hereafter called DCN3 inhibitors. Our efforts have yielded potent and highly selective small-molecule DCN3 inhibitors, exemplified by **TC-**

**6304** which binds to DCN3 with a  $K_i$  value of 35 nM and fails to bind to DCN1 at concentrations as high as 30  $\mu$ M. Cellular thermal shift assays showed that **TC-6304** engages DCN3 in cells in a dose-dependent manner but does not affect the neddylation of any of the cullins that were examined, indicating the different roles of DCN3 when compared to DCN1. This study provides first-in-class, potent and selective small-molecule inhibitors of DCN3. This work has been submitted at the time of thesis writing for publication in the *Journal of Medicinal Chemistry*.

Chapter 4 describes the design, synthesis and evaluation of a chemical library focused on S-adenosylmethionine-based compounds which led to the discovery of first-in-class, potent small-molecule inhibitors directly targeting the MLL1 SET domain. These are exemplified by compound **TC-5115** with an  $IC_{50}$  value of 15 nM against MLL1 methyltransferase, which is 48 times more potent than the pan-HMT inhibitor, SAH. Determination of co-crystal structures for a number of these MLL1 inhibitors reveals that they adopt a unique binding mode that interacts with SET-I domain of MLL1 methyltransferase. This work is in preparation at the time of thesis writing for publication in the *Journal of Medicinal Chemistry*.

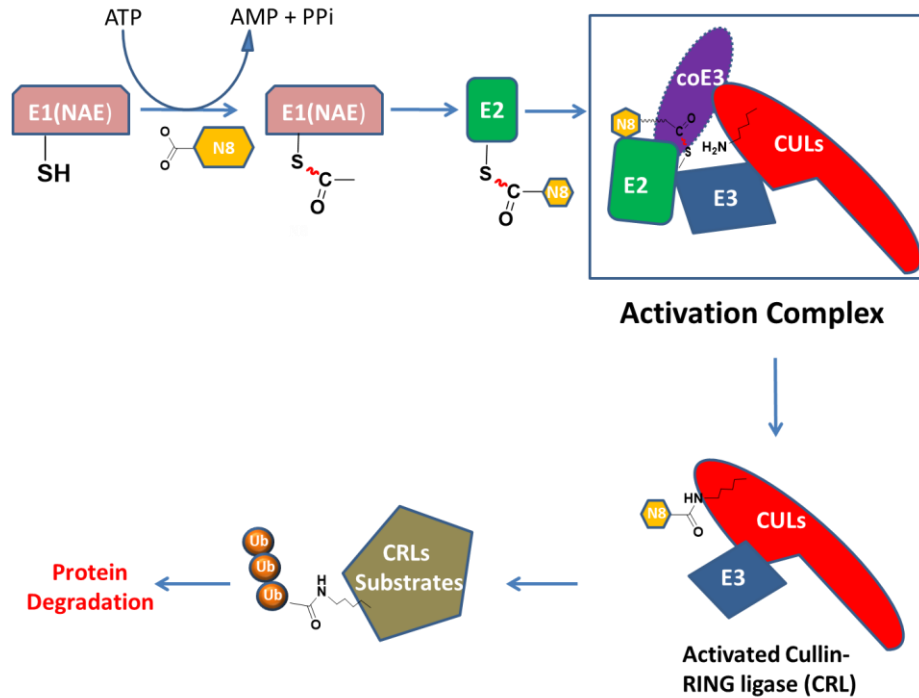
## Chapter 1 Introduction

### 1.1 Introduction to the Neddylaton Pathway

#### 1.1.1 Neddylaton Pathway

The programmed and selective degradation of proteins is regulated by the ubiquitin-proteasome system (UPS) which maintains cellular homeostasis.<sup>1</sup> The UPS plays an important role in many cellular regulation process, such as transcriptional regulation, signal transduction cell cycle regulation, and endocytosis.<sup>1</sup> The process of ubiquitin ligation to the substrate protein involves the sequential action of three enzymes: the E1 ubiquitin-activating enzyme, the E2 ubiquitin conjugating enzyme, and the E3 ubiquitin ligating enzyme.<sup>2</sup> The neddylation pathway is analogous to the ubiquitin pathway and involves involved covalent attachment of NEDD8 to the substrate protein.<sup>3</sup> However, the neddylation of a target protein will not directly result in protein degradation but in the modulation of the target protein activity. NEDD8 (neural precursor cell expressing developmentally downregulated-8) is a most studied ubiquitin-like protein with a highest sequence similarity to ubiquitin.<sup>4</sup> This neddylation cascade starts with the activation of NEDD8 by the NEDD8 activating enzyme (NAE), the E1 enzyme. The activated NEDD8 transfers to either UBC12 or UBE2F, the two E2 enzymes found in mammalian cells. In the final step, the E3 enzyme catalyzes the transfer of NEDD8 to the substrate cullins, resulting in the activation of a CRL.<sup>5</sup> (Figure 1.1)





**Figure 1.1 The neddylation pathway and the activation of CRLs**

### 1.1.2 Cullin-RING Ligase

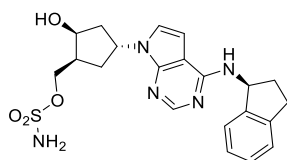
In mammalian cells, the Cullin-Ring ubiquitin ligases (CRL) are the most prominent family of E3 ligases in the UPS.<sup>6, 7</sup> There are seven cullin complexes (CUL1, CUL2, CUL3, CUL4A, CUL4B, CUL5 and CUL7) in human cells with an individual cullin protein being the central component of each CRL.<sup>6</sup> Activation of each CRL requires covalent attachment of NEDD8 to CULs. (Figure 1.1) Studies show that the neddylation of CULs regulates about 20% of all cellular ubiquitination.<sup>8</sup> CUL1 CRL, the most studied system, forms a complex with UBC12, DCN1, RBX1 and NEDD8 proteins for efficient transfer of NEDD8 from UBC12 to CUL1.<sup>9</sup>

### 1.1.3 Current strategies targeting the neddylation pathway

Evidence has shown that NEDD8 is over-expressed and the CRLs are highly activated in human cancer cells.<sup>10, 11</sup> Therefore, targeting the neddylation pathway is an attractive approach to the development of novel therapeutics.<sup>3, 12, 13</sup> **MLN4924**, an inhibitor of the neddylation activating enzyme, arose from the first attempt at targeting the neddylation pathway.<sup>8</sup> Because **MLN4924** inhibits the upstream protein, NAE in the neddylation cascade, it effectively causes broad inactivation of all CRLs. **MLN4924** causes apoptotic death of tumor cells by blocking the CRL-mediated protein turnover. **MLN4924** also suppresses tumor cell growth *in vivo*, and it is now in clinical trials for cancer treatment.<sup>14</sup>

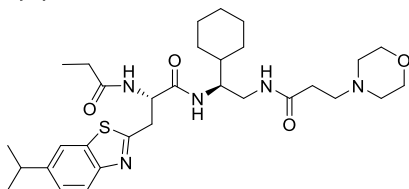
In 2014, the crystal structure of the most studied CUL1 CRL complex was solved.<sup>9</sup> In this complex, DCN1 interacts with both UBC12 and cullin 1 through two separate domains, enhancing cullin neddylation.<sup>9, 15</sup> These findings suggest that targeting the protein-protein interaction between DCN1 and UBC12 within the CRL complex could be an alternative strategy with which to target the neddylation pathway. In 2007, our group and Guy et al. published DCN1 inhibitors **DI-591** and **NAcM-OPT**, respectively.<sup>16, 17</sup> **DI-591** selectively inhibits cullin 3 neddylation and dose dependently increases the protein level of CUL3 CRL mediated substrates.<sup>16</sup> **NAcM-OPT** inhibits DCN1-dependent cellular neddylation, and cell proliferation in DCN1-amplified cell lines.<sup>17</sup>

**(A) NAE inhibitor**



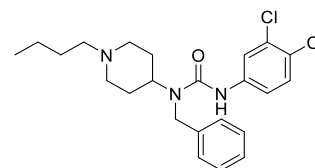
**MLN4924**

**(B) DCN1 inhibitor**



**DI-591**

DCN1  $K_i = 4.4$  nM



**NAcM-OPT**

DCN1 TR-FRET  $IC_{50} = 79$  nM

**Figure 1.2 Inhibitors targeting the neddylation pathway**

## 1.2 Introduction to MLL methyltransferase

### 1.2.1 MLL1 methyltransferase and its relevance to disease

In human cells, H3K4 methylation is catalyzed by the MLL family of methyltransferases, SET7/9. There are six members in the MLL family: MLL1, MLL2, MLL3, MLL4, SET1A and SET1B.<sup>18</sup> Although they have different substrate specificity, and subunit compositions, all MLL proteins share the same catalytic core complex.<sup>19</sup> The SET domain is the core catalytic center of the MLL which is located in the C-terminus of the protein. Other proteins, including the WD40 repeat protein 5 (WDR5), absent/small homebox 2 like protein (ASH2L), retinoblastoma-binding protein 5 (RBBP5) and DPY-30 in a complex with the SET domain compose the catalytic machinery.<sup>20</sup>

Rearrangements of the mixed-lineage leukemia (MLL) gene, also known as MLL1, have been detected in >70% of acute lymphoblastic leukemias (ALL) in infants and 5-10% of acute myelogenous leukemias (AML) in adults.<sup>21, 22</sup> AML patients carrying the MLL fusion gene (MLL leukemia) are unresponsive to current treatments and have a very poor prognosis.<sup>23</sup> It is noted that the MLL translocations invariably occur on only one MLL allele,

which means the remaining wild-type MLL allele still retains the H3K4 methyltransferase activity.<sup>24</sup> It has been shown that both the wild-type and fusion MLL proteins are required for MLL-AF9-induced leukemogenesis and maintenance of MLL-AF9-transformed cells.<sup>24</sup> Consequently, inhibition of the H3K4 methyltransferase activity of MLL has been proposed as a novel therapeutic strategy for the treatment of MLL leukemia.

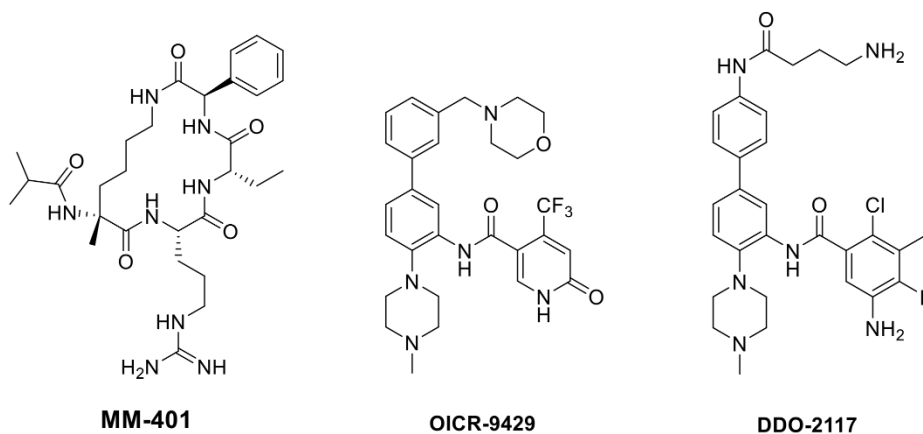
### 1.2.2 Current strategies targeting the MLL1 methyltransferase activity

As described in the previous section, the SET domain of MLL1 is responsible for the catalysis of methylation on H3K4. However, the MLL1 protein alone is not enough to exert the full catalytic activity.<sup>25</sup> Studies have shown that when MLL1 forms a complex with WDR5, RbBP5, and ASH2L, the methyltransferase activity is recovered.<sup>25</sup> Structural studies also showed that MLL1 alone adopts an open conformation and WDR5 protein causes the MLL1 SET domain to adopt an closed, active conformation.<sup>26</sup> Therefore, targeting the WDR5/MLL1 protein-protein interaction has become an emerging strategy for MLL1 HMT activity inhibition.

Our group developed a series of cyclized peptidomimetic compounds based on the sequence of the Win motif that binds to WDR5. Compound **MM-401** binds to WDR5 with a  $K_i$  value  $<1$  nM.<sup>27</sup> Biochemical studies show that **MM-401** selectively inhibits the MLL1 dependent H3k4 methyltransferase activity in the MLL protein family. **MM-401** also inhibits the proliferation of MV4;11 MLL fusion cell lines.<sup>27</sup> Another series of small molecule WDR5 inhibitors was developed by Al-awar et al.<sup>28, 29</sup> High-throughput screening followed by structure-based optimization led to the discovery of compound **OICR-9429** ( $K_d = 0.03$   $\mu$ M) which inhibits proliferation of AML cells.<sup>28, 29</sup> Further

optimization by another group yielded **DDO-2117** with a  $K_d$  value of 13.6 nM and an  $IC_{50}$  value of 7.6 nM in an HMT assay.<sup>30</sup>

At present, there has been no known potent inhibitor targeting the MLL1 SET domain reported in the literature.



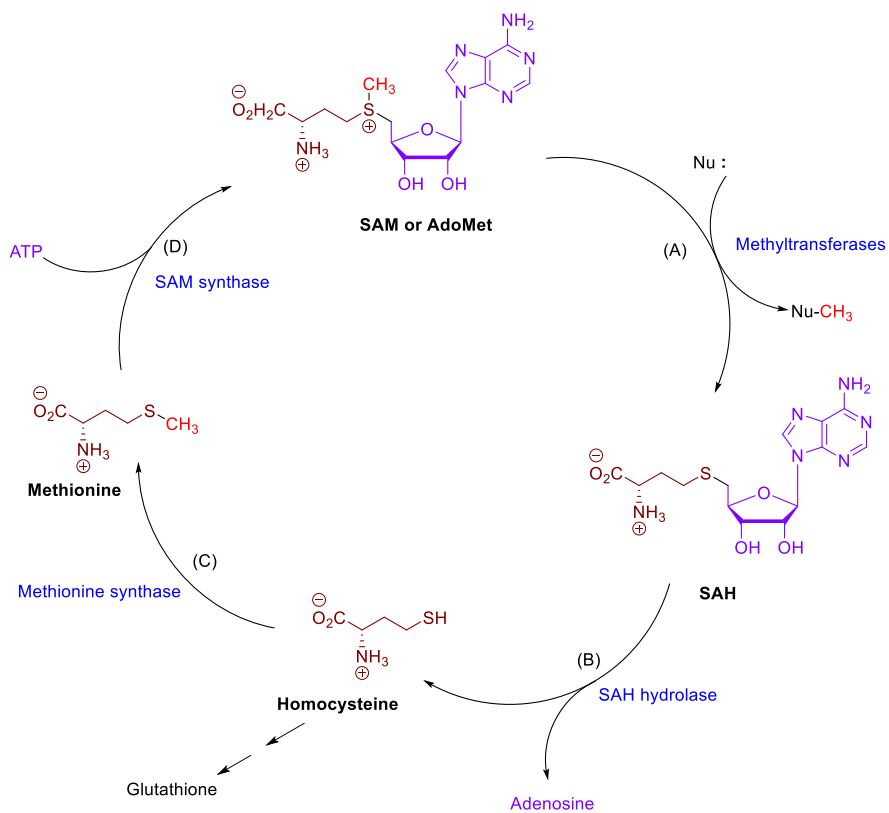
**Figure 1.3 WDR5 inhibitors**

### 1.2.3 SAM Analogs as methyltransferase inhibitors

#### 1.2.3.1 SAM cycle and its role in biology

S-adenosylmethionine (S-adenosyl-L-methionine, AdoMet or SAM) was discovered in 1952.<sup>31</sup> It is a molecule containing a sulfonium ion and generated from a single step coupling reaction of methionine with adenosine catalyzed by SAM synthase.<sup>32</sup> SAM is the second most ubiquitous cofactor after ATP. The major role of SAM is as a methyl donor for different nucleophiles (C-, N, O- or S-) catalyzed by various methyltransferases.<sup>33</sup> The methylation of substrates, including histone protein, RNA, DNA, and carbohydrate, are important for epigenetic regulation, protein degradation and

signal transduction. In general, the methylation reaction is carried out by an  $S_N2$  nucleophilic attack on the electrophilic methyl group of SAM.<sup>33</sup> The generated molecule including the methylated substrate and the by-product *S*-adenosyl-*L*-homocysteine (AdoHcy or SAH). Hydrolysis of SAH, followed by methylation of homocysteine catalyzed by methionine synthase will regenerate methionine.<sup>33</sup> The entire SAM cycle is illustrated in Figure 1.

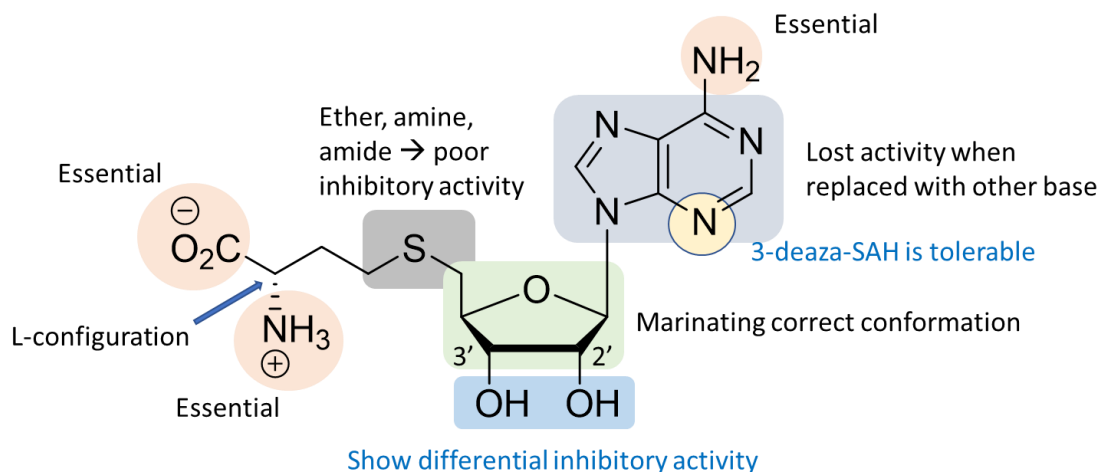


**Figure 1.4 The SAM cycle** (A) Methyltransferases catalyze the reaction of the methyl group of cofactor SAM with the nucleophile. (B) Hydrolysis of SAH generates homocysteine. (C) Subsequent methylation catalyzed by methionine synthase regenerates methionine. (D) SAH is regenerated by the reaction of ATP with methionine, catalyzed by SAM synthase.

### 1.2.3.2 Early studies on SAM analogs targeting methyltransferases

SAM/SAH has been recognized as a valuable chemical scaffold with which to investigate the function of methyltransferases. SAH is the product of demethylated SAM, and SAH itself features the inhibitory activity toward SAM-dependent methyltransferases.<sup>34</sup> This general feature suggests the potential for developing methyltransferase inhibitors based on SAH.<sup>35, 36</sup> In fact, much work has been done, for example by Coward and Wu since the 1970s to modify the structure of SAM/SAH in an attempt to target various methyltransferases.<sup>35-41</sup> However, due to the structural modifications focused mainly on SAM itself, those compounds have general lack of specificity and usually have similar or weaker inhibitory activity than SAH.<sup>35-41</sup> Wu et al. made systematic structural modifications to the base, amino acid, and sugar portion of SAH and tested the inhibitory activity of the products in COMT, PNMT, HMT and HIOMT.<sup>38-40</sup> This SAR study revealed that the 6-amino group of SAH is essential for inhibitory activity in COMT, PNMT, HMT and HIOMT.<sup>38</sup> Replacement of the adenine in SAH by uracil, cytosine, guanine, or hypoxanthine resulted in loss of inhibitory activity.<sup>38</sup> Replacing the ribose moiety of SAH by cyclopentyl, 2,3-dihydroxycyclopentyl, or a five-carbon cyclic system, resulted in a dramatic decrease of the inhibitory activity toward SAM-dependent methyltransferases.<sup>36</sup> Further investigation of the 2' hydroxyl and 3' hydroxyl group of the ribose moiety showed differential inhibition among COMT, PNMT, HMT, and HOIMT. For example, removal of either the 2'-hydroxyl or the 3'-hydroxyl of SAH is detrimental for HOIMT inhibition, while 3'-deoxy compounds are tolerable in PNMT.<sup>40</sup> Replacement of 5'-thioether with an ether, amine, or amide resulted in poor

inhibitory activity toward COMT and tRNA methylase.<sup>41</sup> The terminal amine and terminal carboxyl group on SAH are required for several methyltransferase inhibition reactions.<sup>39</sup>



**Figure 1.5 Summary of SAR on modifying SAH for methyltransferases**

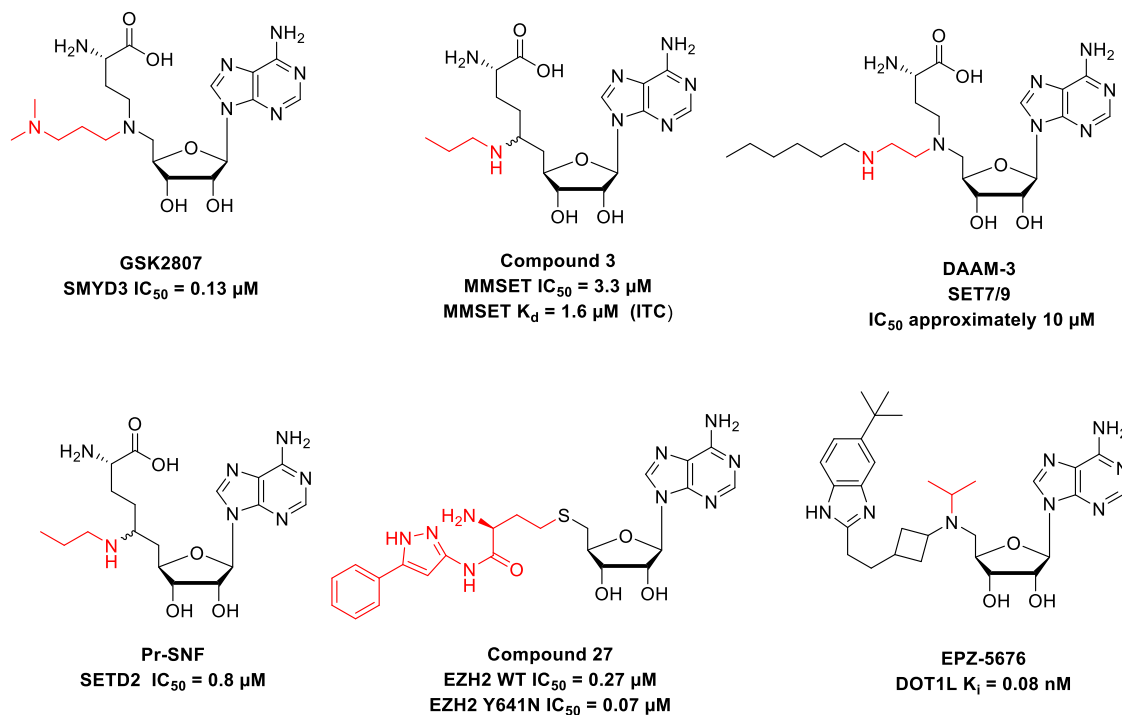
### 1.2.3.3 The bisubstrate strategy to design SAM based analogs targeting methyltransferases

The bisubstrate inhibitor is composed of two linked fragments that bind to different binding pocket of an enzyme and mimic the ternary complex of an enzymatic reaction. Compared to single-binding site inhibitors, the bisubstrate inhibitor has more extended interaction with the target enzyme leading to better binding affinity and specificity.<sup>42</sup> This design strategy has been utilized for a variety of targets, including kinases<sup>43</sup>, N-acetyltransferase<sup>44</sup>, phosphorylase<sup>45</sup> and methyltransferases.<sup>46, 47</sup>

As discussed in a previous section, the early modifications based on SAH resulted in even weaker activity and poor selectivity among methyltransferases. Currently, the SAM-based inhibitor design for methyltransferase mainly focused on bisubstrate strategy. This strategy has been successfully employed for the design of direct inhibitors for a



number of histone methyltransferases such as MMSET,<sup>48</sup> SET7/9,<sup>46, 47</sup> SMYD3,<sup>49</sup> SETD2,<sup>50</sup> EZH2,<sup>51</sup> and DOT1L<sup>52, 53</sup>.



**Figure 1.6 SAM analog inhibitors targeting histone methyltransferases**

A research group in GlaxoSmithKline discovered a novel SMYD3 inhibitor **GSK2807** with an IC<sub>50</sub> value of 0.13 μM.<sup>54</sup> **GSK2807** is selective for SMYD3 over 9 other tested histone methyltransferases. The co-crystal structure of **GSK2807** with SMYD3 shows that the propyl dimethylamine moiety in **GSK-2807** occupies the lysine tunnel in SMYD3. A different compound, **DAAM-3** shows moderate inhibitory activity toward SET7/9 and the selectivity was not reported. The crystal structure of **DAAM-3** with SET7/9 showed that its N-hexyl group is inserted into the narrow lysine tunnel.<sup>25</sup> **Compound 3** and **Pr-SNF** are actually the same compound but were claimed as MMSET and SETD2 inhibitors respectively in different publications.<sup>48, 50</sup> **Pr-SNF** shows moderate activity

toward SET7/9, CARM1 and PRMT1, but has no activity at concentrations up to 100  $\mu$ M with SET8, G9a and GLP.<sup>50</sup> **Compound 27** was designed based on a homology model of EZH2. This compound exhibited fairly good selectivity to 10 HMTs and it is even more potent for Y641N mutant EZH2 enzyme.<sup>51</sup> It is the only compound retaining the sulfur atom from SAM. **EPZ-5676** was developed by Epizyme and it is currently being tested in clinical trials. **EPZ-5676** is extremely potent toward DOT1L and it fails to inhibit a large panel of HMTs.<sup>52, 53</sup> The reason for its selectivity is that DOT1L is the only HMT without a SET domain. In addition, all SAM analog HMT inhibitors, except **EPZ-5676**, has no cellular data reported in the literature.

### 1.3 Overview of Thesis

In the following chapters, we demonstrate the development of chemical probes targeting the DCN family of proteins and the MLL1 methyltransferase. Chapter 2 presents the structure-based design of high-affinity peptidomimetic inhibitors targeting the DCN1-UBC12 protein-protein interaction. Chapter 3 describes the development of a BLI and an FP assay for DCN3, and is followed by focused-library screening and structure-based optimization for potent and selective first-in-class DCN3 inhibitors. Chapter 4 outlines the rational design of first-in-class MLL1 methyltransferase SET domain inhibitors using a bivalent approach. Through working at the interface between chemistry and biology, my dissertation inaugurates novel, first-in-class compounds which examine the neddylation system and the epigenetics mediated by MLL1 methyltransferase, opening a new path for innovations in understanding their roles in human disease.

## **Chapter 2 Development of Peptidomimetic Inhibitors of the DCN1-UBC12 Protein-Protein Interaction**

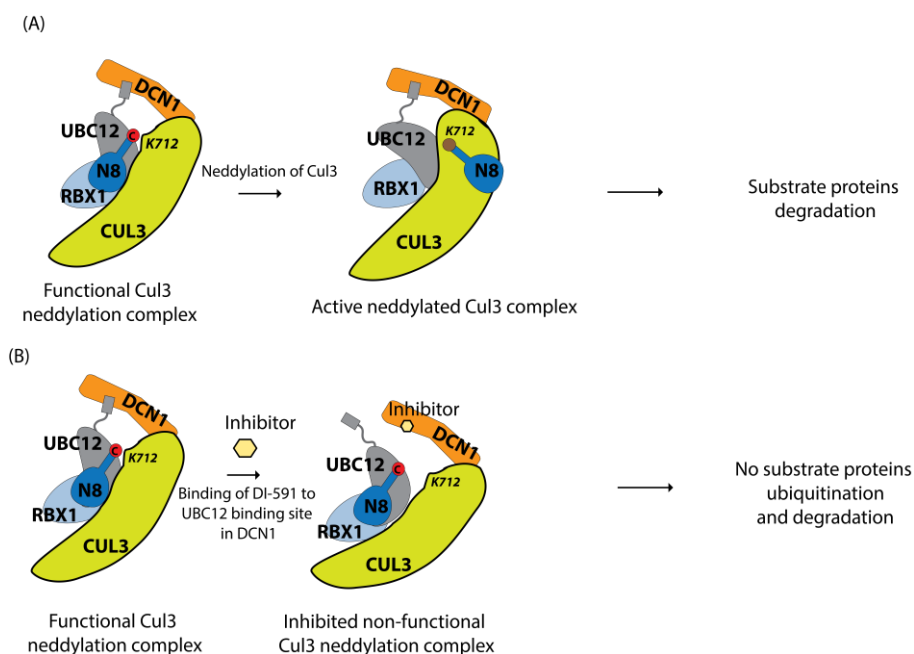
### **2.1 Introduction**

One major cellular mechanism to eliminate misfolded or damaged proteins is to target them to proteasomes for degradation through a ubiquitin-dependent process.<sup>55, 56</sup> Cullin-RING ubiquitin ligases (CRLs) are one of the largest family proteins responsible for degradation of approximately 20% of cellular proteins.<sup>6, 57-59</sup> In mammalian cells, there are eight cullin members, Cul1, Cul2, Cul3, Cul4A, Cul4B, Cul5, Cul7 and Cul9, each of which serves as the central component of the respective CRL.<sup>6, 60, 61</sup> Covalent attachment to a Neural precursor cell Expressed Developmentally Down-regulated protein 8 (NEDD8) to each cullin member, a process known as neddylation, is necessary for activation of an individual CRL<sup>62-65</sup>. Neddylation involves a three-step enzymatic cascade, in which NEDD8 is first activated and attached to the E1 neddylation enzyme UBA3-NAE1. A second step requires its attachment to one of the two E2 enzymes (UBC12 or UBE2F) and in a third step, to cullin. A small-molecule inhibitor, MLN4924, which forms a covalent bond with NEDD8 and blocks neddylation, is capable of achieving a broad and effective inhibition of all CRLs.<sup>8</sup>

Because each CRL modulates degradation of specific substrate proteins, compounds that can selectively regulate the activity of an individual CRL member are valuable research tools for further investigation of the roles of each CRL in biological

processes and human diseases, and such compounds may also have the potential to be developed as new therapies for human diseases. To this end, we have proposed a general strategy with which to target the protein-protein interaction network within the multi-protein neddylation E3 ligase complex as a means of inhibiting the activity of individual CRLs.<sup>16</sup>

Analysis of the x-ray crystallographic structure of the Cullin1-RBX1-UBC12-NEDD8-DCN1 complex<sup>9, 15</sup> suggests that the protein-protein interaction between DCN1 and UBC12 may permit the design of small-molecule inhibitors, hereafter called DCN1 inhibitors. We have recently published a structure-based design and extensive evaluation of DI-591 as a high-affinity, druglike, small-molecule inhibitor of DCN1.<sup>16</sup> The posed model of inhibitor mediated CUL3 CRL inactivation is proposed in Figure 2.1. Another recent independent study has demonstrated the structure-based design of high-affinity, non-peptide, small-molecule DCN1 inhibitors.<sup>17</sup>



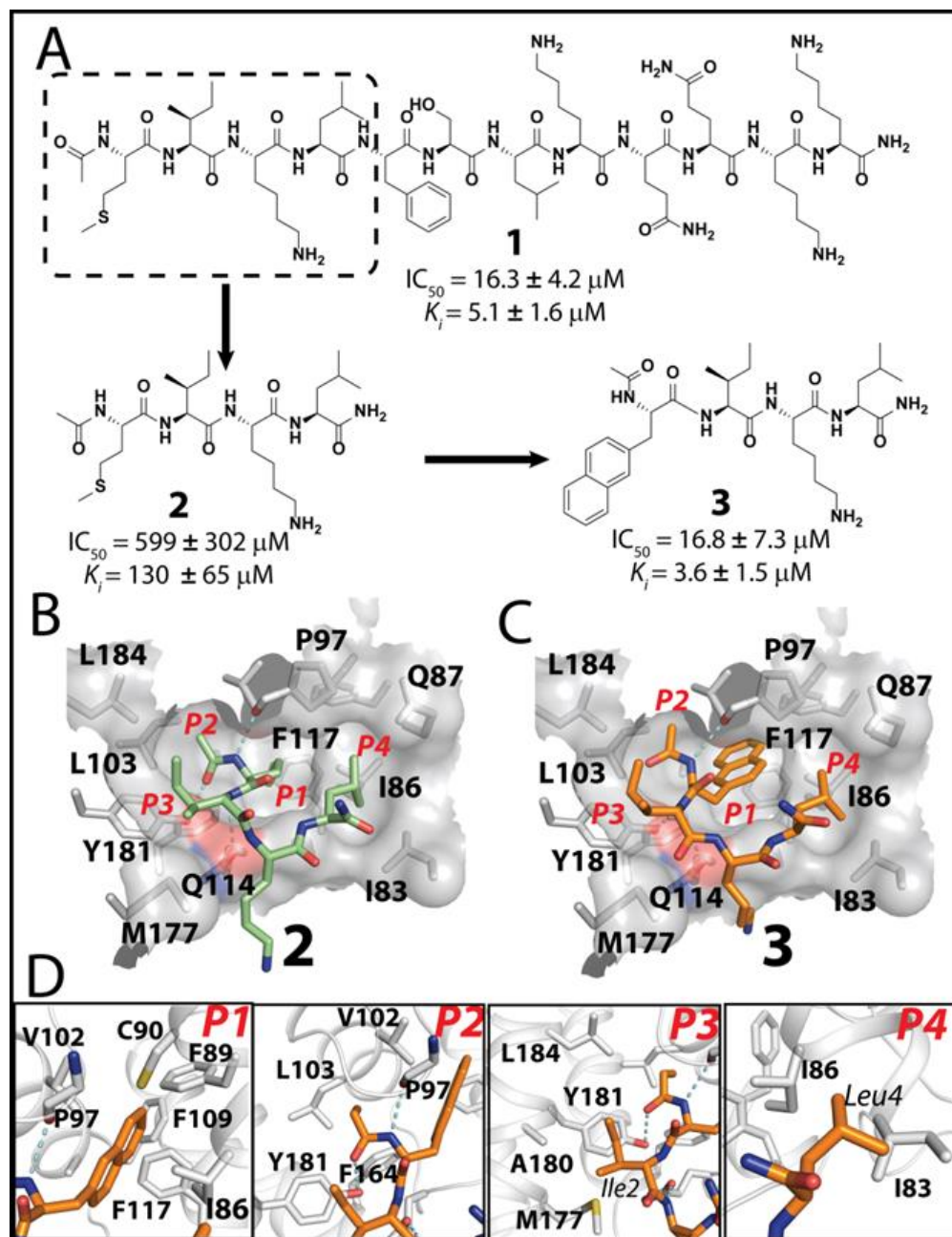
**Figure 2.1 Proposed model of activation of CUL3 CRL and the model of inactive CUL3 CRL blocked by inhibitor**

In an effort parallel to our development of DI-591, we have carried out the design, synthesis and evaluation of peptidomimetic inhibitors of DCN1 based upon a tetrapeptide derived from a 12-residue segment of the UBC12 peptide. This has resulted in the discovery of potent, cell-permeable peptidomimetic DCN1 inhibitors with  $K_D$  values <10 nM. The co-crystal structure of one such inhibitor in a complex with DCN1 provides the structural basis for their high-affinity interaction and for further optimization. Evaluation of one potent inhibitor (**36**, DI-404) shows that in cells, it engages the DCN1 protein and effectively blocks its interaction with the UBC12 protein. DI-404 also selectively inhibits the neddylation of cullin 3 CRL over cullin CRLs in cells.

## 2.2 Results and Discussion

### 2.2.1 Identification of a 4-residue lead peptide

As shown in Figure 2.2 A, our systematic truncation of the 12-residue UBC12 peptide identifies a tetrapeptide (**2**) derived from the *N*-terminus of UBC12 which retains a significant affinity to DCN1. Further modifications of this tetrapeptide, (**2**), leads to compound **3**, which binds to DCN1 with a  $K_i$  value of 3.6  $\mu$ M and is thus as potent as the 12-mer peptide **1** ( $K_i = 5.1 \mu$ M) (**Figure 2.2 A**). We employed this tetrapeptide (**3**) as the starting point for systematic optimization aimed at improving its binding affinity to DCN1.

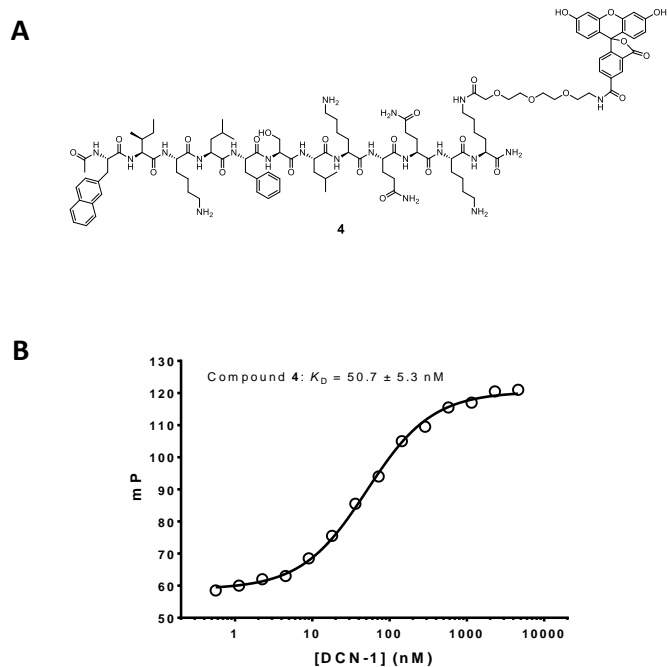


**Figure 2.2 Chemical structures of designed peptides and the structural analysis of peptides-DCN1 interaction** (A) Chemical structures of the 12-residue UBC12 peptide (1) and tetrapeptides 2 and 3. (B). Crystal structure of peptide 2 (green) with DCN1 extracted from the structure of peptide 1 with DCN1 (PDB ID: 3TDU). (C) Predicted binding model of compound 3 (orange) in a complex with DCN1. Hydrogen bonds are depicted as cyan dash lines. (D). Detailed views of compound 3 and interacting residues in the P1-P4 pockets.

Based upon the co-crystal structure of UBC12 peptide (**1**) complexed with DCN1 (Figure 2.2 B),<sup>9, 15</sup> we predicted a binding model for compound **3** (Figure 2.2 C), which shows that its naphthalene group occupies the P1 “core pocket” of DCN1, while the acetyl group, the side chain of Ile2 and the side chain of Leu4 interact with the P2, P3 and P4 pockets, respectively. We performed stepwise modifications of compound **3** with the goal to optimize its interactions with residues in each of these four pockets.

### **2.2.2 Development of a fluorescence polarization (FP)-based competitive binding assay**

To accurately evaluate the binding affinities of our designed compounds to DCN1, we designed and synthesized a fluorescently tagged tracer (**4**), (Figure 2.3 A). The  $K_D$  value of **4** with DCN1 was determined to be 50.7 nM (Figure 2.3 B). We developed and optimized the fluorescence polarization (FP)-based competitive binding assay using compound **4** and recombinant human DCN1 protein. With this assay, we determined the binding affinities of our designed compounds in this study. All  $IC_{50}$  and  $K_i$  values reported in the present study were obtained from at least three independent experiments.



**Figure 2.3** Chemical structure of compound 4. B) Binding isotherm of compound 4 to DCN1.

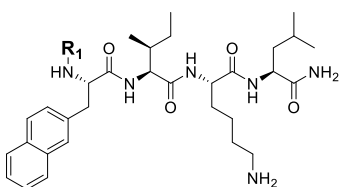
### 2.2.3 Modification of the N-terminal acetyl group, interacting with the P2 pocket

As shown in Figures 2.2B and 2.2C, the methyl of the *N*-terminal acetyl group is directed towards the P2 hydrophobic pocket and, the carbonyl of this acetyl group forms a hydrogen bond with the hydroxyl group of Y181. The NH group bearing the *N*-terminal acetamide group forms a hydrogen bond with the backbone carbonyl group of P97 in DCN1. To probe the interactions with the P2 pocket formed by V102, L103 and F164 (Figure 2.2D), the acetyl group in **3** was replaced by a formyl group giving **5** ( $K_i = 27.5 \mu\text{M}$ ), which is 8 times less potent than **3**. Removal of the acetyl group produced **6**, which has negligible affinity to DCN1 at 100  $\mu\text{M}$ . These data confirm the importance of the *N*-terminal acetyl group and are consistent with the previously reported binding affinities of 12-mer residue peptides.<sup>15</sup>



Our predicted binding model (Figure 2.2D) shows that the P2 pocket can accommodate a larger hydrophobic moiety, and accordingly, several small hydrophobic groups were tested, giving the results shown in Table 2.1. Substitution of the methyl group in **3** with an ethyl group gave compound **7** ( $K_i = 1.1 \mu\text{M}$ ), improving the binding affinity by a factor of 3. However, replacement of the ethyl group in **7** with a larger group, such as cyclopropyl (**8**,  $K_i = 5.4 \mu\text{M}$ ), isopropyl (**9**,  $K_i = 3.5 \mu\text{M}$ ), n-propyl (**10**,  $K_i = 8.0 \mu\text{M}$ ), or isobutyl (**11**,  $\text{IC}_{50} > 20 \mu\text{M}$ ) decreases the binding affinity. Binding affinity is also significantly reduced when the propionamide in **7** is replaced by a moiety that is similarly sized but less hydrophobic, such as methylurea (**12**,  $K_i = 4.0 \mu\text{M}$ ) or methyl carbamate (**13**,  $\text{IC}_{50} > 50 \mu\text{M}$ ). We concluded that the optimal N-terminal cap is a propionyl group, as in compound **7**.

**Table 2.1 Structures and binding affinities of compound 3 and its analogues designed to investigate the interactions with the P2 pocket in DCN1.**

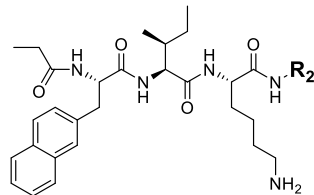


	<b>3</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>
$R_1$										
$\text{IC}_{50}$ ( $\mu\text{M}$ )	$16.8 \pm 7.3$	$128 \pm 59$	$> 100$	$5.2 \pm 1.1$	$25.1 \pm 11.5$	$16.5 \pm 1.7$	$37.0 \pm 23.6$	$> 20$	$18.5 \pm 8.7$	$> 50$
$K_i$ ( $\mu\text{M}$ )	$3.6 \pm 1.5$	$27.5 \pm 12.7$		$1.1 \pm 0.2$	$5.4 \pm 2.5$	$3.5 \pm 0.4$	$8.0 \pm 5.1$		$4.0 \pm 1.9$	

## 2.2.4 Modifications at the C-terminus probing interactions with the P4 pocket

Next we optimized the interactions of the C-terminal residue, Leu4 in **7**, with the P4 pocket of DCN1. The co-crystal structure of peptide **1** complexed with DCN1 (**Figure 2.2B**) and our modeling of **3** (**Figure 2.2D**) showed that the hydrophobic side chain of Leu4 in **3** is directed to the P4 hydrophobic pocket, while the carbamoyl group of Leu4 is oriented towards the solvent, away from the DCN1 protein and thus lacking any specific interaction with the protein. Based on these observations, we replaced Leu4 in **7** with various hydrophobic groups with the goal to improve the potency and at the same time to reduce the overall peptidic nature of the resulting compounds (Table 2.2).

**Table 2.2 Structures and binding affinities of compound 7 and its analogues, designed to investigate the P4 pocket in DCN1.**



	<b>7</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>	<b>21</b>
R <sub>2</sub>									
IC <sub>50</sub> (μM)	5.2 ± 1.1	34.9 ± 6.5	3.9 ± 0.3	10.8 ± 2.1	0.61 ± 0.19	0.91 ± 0.11	1.9 ± 0.3	0.55 ± 0.08	0.80 ± 0.11
K <sub>i</sub> (μM)	1.1 ± 0.2	7.5 ± 1.4	0.81 ± 0.07	2.5 ± 0.5	0.097 ± 0.042	0.16 ± 0.02	0.38 ± 0.05	0.084 ± 0.017	0.14 ± 0.02

Replacement of Leu4 in compound **7** with a phenyl group gave compound **14** ( $K_i = 7.5 \mu\text{M}$ ), which decreases the binding affinity by a factor of 7. Replacement of Leu4 by a benzyl group afforded **15** ( $K_i = 0.81 \mu\text{M}$ ), which maintains the binding affinity. Substitution of the benzyl group in **15** with 2-phenylethyl group yielded **16** ( $K_i = 2.5 \mu\text{M}$ ), which is 3-times less potent than **15**. We next incorporated a second phenyl ring at the  $\alpha$

position of the benzyl group in **15** to restrict the orientation of the first phenyl group. The resulting compound **17** binds to DCN1 with a  $K_i$  value of 0.097  $\mu\text{M}$  and is thus 8 times more potent than **15**, revealing the benefit of restricting the conformation of the phenyl group. As shown in the predicted binding model of **17** with DCN1 (Figure 2.4), one of the phenyl rings of the benzhydryl group in **17** interacts with the P4 pocket and the other phenyl ring is oriented towards the solvent. In addition to enhancing the binding affinity by restricting the conformational freedom, replacement of the primary amide with a phenyl ring also reduces the overall peptidic nature of compound **17**.

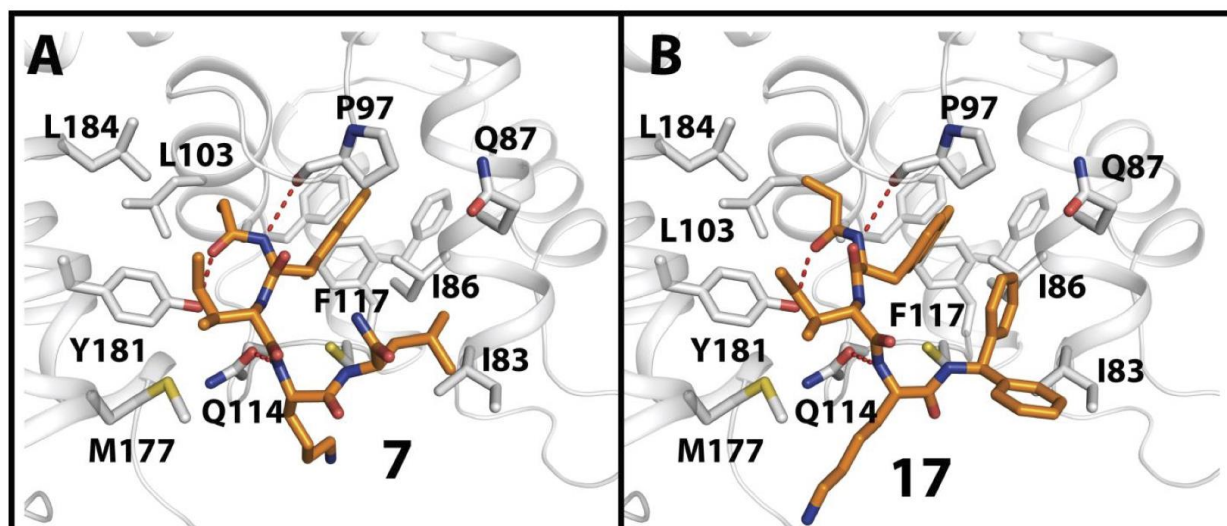


Figure 2.4 Binding mode of compound **7** and **17** in DCN1

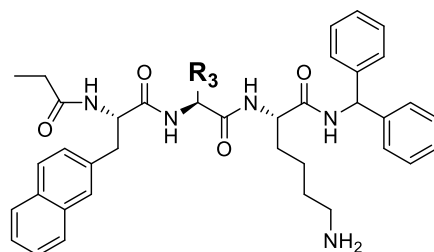
Encouraged by the binding data for **17**, we employed a bicyclic ring system to control the orientation of the phenyl ring in **15**.<sup>66-68</sup> As shown in Table 2, replacement of the benzyl group in compound **15** by (*R*)-2,3-dihydro-1*H*-inden-1-yl yielded **18**, by (*S*)-2,3-dihydro-1*H*-inden-1-yl (**19**), by (*R*)-1,2,3,4-tetrahydronaphthalen-1-yl (**20**) and by (*S*)-1,2,3,4-tetrahydronaphthalen-1-yl (**21**). Compounds **18**, **19**, **20** and **21** bind to DCN1 with  $K_i$  values of 0.16  $\mu\text{M}$ , 0.38  $\mu\text{M}$ , 0.084  $\mu\text{M}$  and 0.14  $\mu\text{M}$ , respectively, showing that the 6-6

bicyclic ring system in **20** and **21** enhances the binding more effectively than the 5-6 bicyclic ring system in **18** and **19**. Further, the (*R*)-configuration in **18** and **20** appears to be the preferred conformation for the chiral center because it places the phenyl ring in a favorable position for binding to DCN1, consistent with the *L*-conformation of Leu4 in **7**. Notably, compound **20** with the (*R*)-1,2,3,4-tetrahydronaphthalen-1-yl group is as potent as **17**. In addition to the enhanced binding affinity, substitution of Leu4 in **7** with diphenylmethyl (**17**) or (*R*)-1,2,3,4-tetrahydronaphthalen-1-yl (**20**) also reduces the peptidic characteristic of the resulting compound by removal of one amide group.

### 2.2.5 Modifications of Ile2 to probe interactions with the P3 pocket

As shown in Figure 2.2 D, the Ile2 residue in **3** interacts with the P3 shallow hydrophobic pocket in DCN1. We modified the Ile2 in compound **17** with residues of different size, shape and hydrophobicity to explore the interactions of this residue with the P3 pocket in DCN1. Compounds **22**, **24** and **25** with a cyclopentylglycine, cyclohexylalanine or a *t*-butylalanine group bind to DCN1 with  $K_i$  values of 36.3, 44.3 nM and 42.9 nM. All these compounds are modestly more potent than **17** ( $K_i = 96.7$  nM). On the other hand, compound **23**, containing cyclohexylglycine binds to DCN1 with  $K_i$  values of 97.3 nM and is similarly potent as **17**. However, compound **26** with 2,2,2-trifluoromethylalanine ( $K_i = 672$  nM) and **27** with thiophen-2-ylalanine ( $K_i = 174$  nM) are both weaker than **17**.

**Table 2.3 Structures and binding affinities of compound 17 and its analogues designed to probe the interactions with the P3 pocket in DCN1.**



	<b>17</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>
<b>R<sub>3</sub></b>							
<b>IC<sub>50</sub> (nM)</b>	611 ± 193	330 ± 95	613 ± 385	368 ± 110	361 ± 95	3266 ± 490	967 ± 122
<b>K<sub>i</sub> (nM)</b>	96.7 ± 41.6	36.3 ± 20.6	97.3 ± 83.0	44.3 ± 23.8	42.9 ± 11.3	672 ± 101	174 ± 22

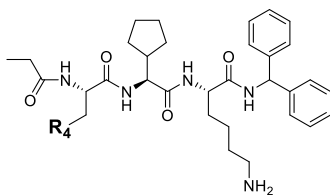
### 2.2.6 Investigations of the interactions with the P1 core pocket

As shown in Figure 2.2 D, the naphthyl group of compound **3** is inserted deeply into the P1 core pocket in DCN1, and so we next focused our modifications on the naphthyl group in compound **22**. After extensive exploration, the benzothiazole ring was identified as a good replacement for the naphthalene group. Compound **28** containing a benzothiazole binds to DCN1 with a  $K_i$  value of 50.7 nM and is thus as potent as **22** (Table 4). To further optimize the interactions of compound **28** with DCN1, different substituents on the benzothiazole ring were examined and the results are summarized in Table 4. Replacement of the benzothiazol-2-yl alanine in **28** with 4-fluorobenzothiazol-2-yl alanine or 4-fluorobenzothiazol-2-yl alanine yielded compounds **29** and **30**, respectively, which are 20- and 4-times less potent than **28**, respectively. Replacement of the benzothiazol-2-yl alanine with 6-fluorobenzothiazol-2-yl alanine gave **31** ( $K_i = 54.4$  nM), which is equipotent with **28**. We next explored the influence of chloro substitution of the

benzothiazole and found that a 5-chlorobenzothiazol-2-yl alanine group (**32**,  $K_i = 88.6$  nM) has a decreased binding affinity while compound **33** with a 6-chlorobenzothiazol-2-yl alanine group ( $IC_{50} = 132$  nM,  $K_i < 50$  nM) is more potent than **28** ( $IC_{50} = 397$  nM,  $K_i = 50.7$  nM).

In view of the superior potency of compound **33**, we employed a label-free Bio-Layer Interferometric (BLI) method to determine the dissociation constant ( $K_D$ ) of **33** and several other compounds. In this BLI assay, compound **33** binds to DCN1 with a  $K_D$  value of 8.1 nM, and compounds **22**, **28** and **31** have  $K_D$  values of 56.4 nM, 31.8 nM and 37.0 nM, respectively (Table 2.4). Thus, compound **33** is 4 times more potent than compound **28**. We next used the BLI assay to evaluate the binding affinities of new analogues of compound **33** to DCN1.

**Table 2.4 Structures and binding affinities of compound 22 and its analogues designed to investigate the P1 core pocket in DCN1.**

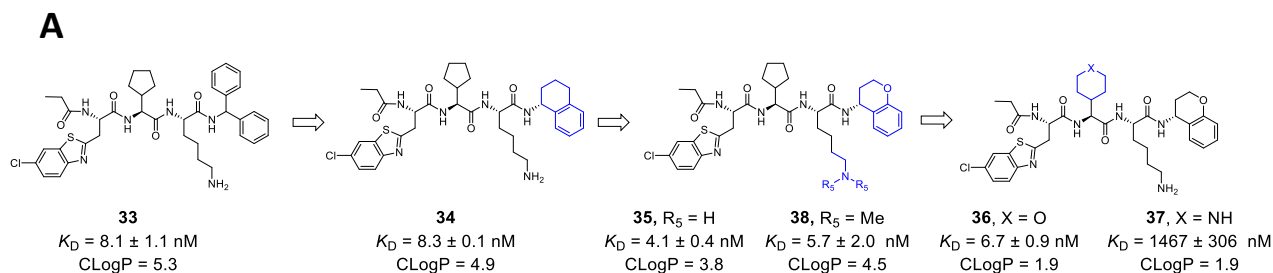


	<b>22</b>	<b>28</b>	<b>29</b>	<b>30</b>	<b>31</b>	<b>32</b>	<b>33</b>
$IC_{50}$ (nM)	330 ± 95	397 ± 96	4917 ± 2596	1197 ± 653	413 ± 132	572 ± 129	132 ± 53
$K_i$ (nM)	36.3 ± 20.6	50.7 ± 12.3	1029 ± 543	224 ± 122	54.4 ± 17.4	88.6 ± 20.0	< 50
$K_D$ (nM) OctetRED	56.4 ± 6.3	31.8 ± 8.2	not tested	not tested	37.0 ± 14.9	not tested	8.1 ± 1.1

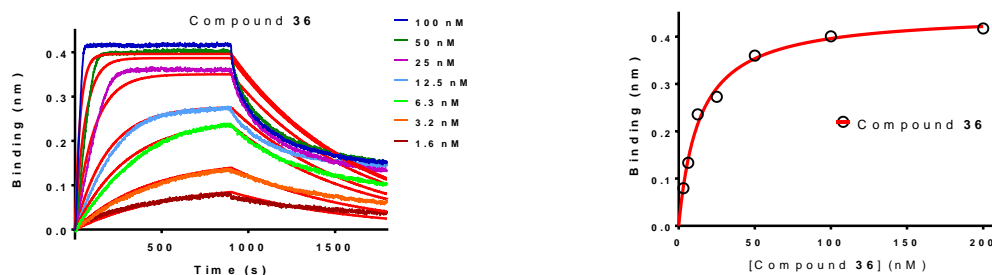
### 2.2.7 Optimization of the physicochemical properties of compound 33

Compound **33** binds potently to DCN1 ( $K_D = 8.1$  nM), but it is quite hydrophobic. We next focused our optimization on improving the solubility of compound **33**, while

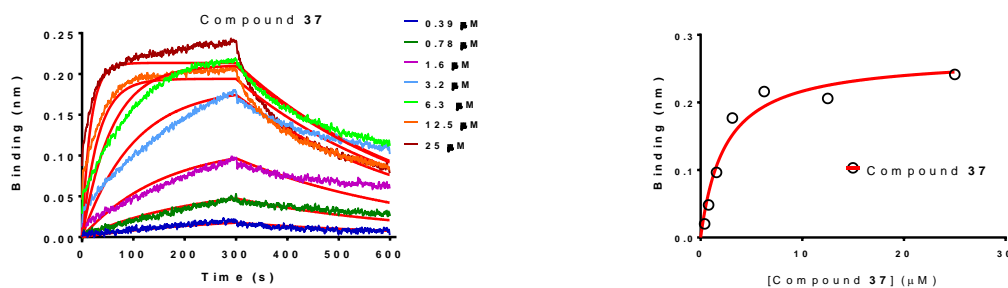
retaining its high binding affinity to DCN1 (Figure 2.5 A). Replacement of the diphenylmethanamide in **33** with (*R*)-1,2,3,4-tetrahydronaphthalen-1-amine gave compound **34** ( $K_D = 8.3$  nM in the BLI assay), which binds to DCN1 as potently as **33**. In compound **34**, the aromatic ring of (*R*)-1,2,3,4-tetrahydronaphthalen-1-amine is predicted to penetrate into the P4 pocket while the aliphatic ring portion of the bicyclic ring system is exposed to solvent. This provides an opportunity of insertion of a polar atom into this aliphatic ring to improve the overall physicochemical properties without negatively affecting the binding affinity of the resulting compounds. Accordingly, compound **35** was designed by replacing the C1 atom in the (*R*)-1,2,3,4-tetrahydronaphthalen-1-amine ring with an oxygen atom. The resulting compound **35** binds to DCN1 with a  $K_D$  value of 4.1 nM in the BLI assay, and its calculated ClogP value is decreased to 3.8 from the calculated ClogP value of 5.0 for compound **33**. We also synthesized compound **38**, in which the primary amine group of **35** is dimethylated. Compound **38** binds to DCN1 with a  $K_D$  value of 5.7 nM in the BLI assay. To further reduce the hydrophobicity of **35**, an oxygen or nitrogen atom was inserted into the cyclopentyl ring in **35** to yield **36** (DI-404) and **37**, respectively, which have calculated CLogP values of 1.9. Compound **36** binds to DCN1 with a  $K_D$  value of 6.7 nM (Figure 2.5 B) and **37** has a  $K_D$  value of 1467 nM (Figure 2.5 C) in the BLI assay. The much higher affinity for compound **36** than for **37** is almost entirely due to the much faster on-rate for **36** (Figure 2.5 B) than for **37** (Figure 2.5 C) since these two compounds have essentially the same off-rates.



**B** Compound **36**:  $K_D = 6.7 \pm 0.9$  nM,  $k_{on} = 2.9 \pm 1.0 \times 10^5$  M<sup>-1</sup>s<sup>-1</sup>,  $k_{off} = 2.0 \pm 1.0 \times 10^{-3}$  s<sup>-1</sup>



**C** Compound **37**:  $K_D = 1467 \pm 306$  nM,  $k_{on} = 2.4 \pm 0.2 \times 10^3$  M<sup>-1</sup>s<sup>-1</sup>,  $k_{off} = 2.6 \pm 0.3 \times 10^{-3}$  s<sup>-1</sup>



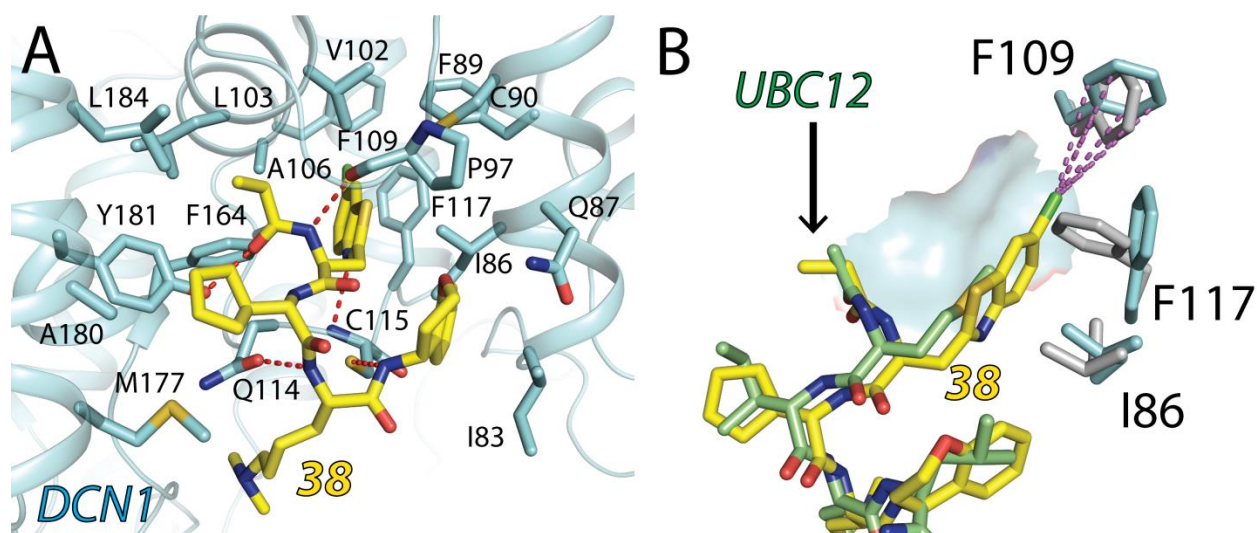
**Figure 2.5 Optimization of compound 33 and the Kinetic binding sensorgrams**

(A) Optimization of compound **33** leading to compound **36**. CLogP values were calculated using ChemBioDraw Ultra (<http://www.perkinelmer.com>). (B) Kinetic binding sensorgrams of **36** were determined with concentrations ranging from 1.6 nM to 100 nM. (C) Kinetic binding sensorgrams of **37** with concentrations ranging from 0.39 to 25 μM. In both (B) and (C), the steady state fitting of the equilibrium responses vs compound concentrations is based on 1:1 binding model.



## 2.2.8 Determination of the crystal structure of compound **38** in a complex with DCN1

To gain structural insights for their high binding affinities, we attempted to determine co-crystal structures for a number of compounds in a complex with DCN1. We have successfully determined the co-crystal structure of compound **38** in complex with DCN1 at a resolution of 2.16 Å (Figure 2.6 A, PDB ID 6B5Q).



**Figure 2.6** Co-crystal structure of DCN1 complexed with compound **38**. (A) Co-crystal structure of DCN1 (cyan) complexed with compound **38** (yellow, PDB ID 6B5Q). (B) Overlay of the DCN1-**38** complex (PDB ID 6B5Q) with DCN1-UBC12 (green) peptide (PDB ID 3TDU). Major conformational differences at F109 and F117 at the P1 pocket were found between these two complex structures. Hydrogen bonds are shown as red dashed lines. Halogen-aromatic interactions are depicted by purple dashed lines.

Consistent with our modeling prediction, the 6-chlorobenzothiazol-2-yl group of **38** is buried deeply in the P1 hydrophobic pocket formed by I86, F89, C90, F109, F117, V102, A106 and F164 (Figure 2.6 A) and this deep penetration in the P1 pocket by **38** induces major conformational changes at F109 and F117 to make room to accommodate **38** (Figure 2.6 B). Additionally, chlorine-aromatic interactions are observed between **38** and

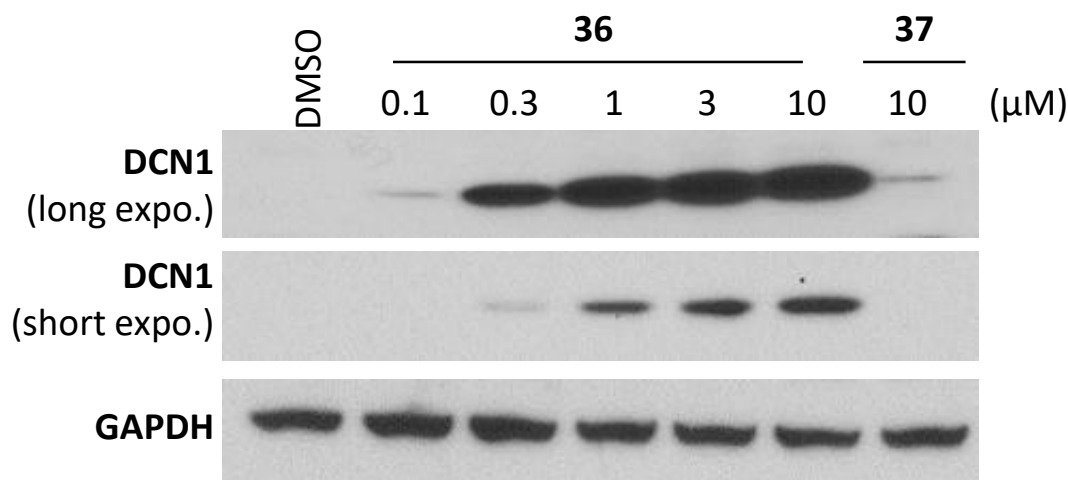
F109 of DCN1.<sup>69, 70</sup> The distance from the Cl atom in compound **38** to the center of aromatic ring of F109 is 4.96 Å, suggesting a Cl- $\pi$  interaction. The Cl atom in compound **38** also interacts with F117 with a distance of 3.4-4.0 Å to the edge of the phenyl ring. The combination of the Cl- $\pi$  interaction of compound **33** with F109 in DCN1 and the hydrophobic interaction with F117 in DCN1 underlies a ~4-fold increase in potency of compound **33** as compared to compounds **28**, **31** and **32** (Table 4). Compared to the acetyl group in **1**, the additional methyl of the propionyl group in compound **38** fits well into the small hydrophobic pocket formed by L103, L184, Y181, A106 and V102. Similar to those in the co-crystal structure of the UBC12 peptide **1**-DCN1 complex, the propionyl group of **38** forms hydrogen bonds with P97 and Y181 of DCN1 while the backbone amide of K3 in **38** forms a hydrogen bond with Q114 of DCN1. The phenyl of the (*R*)-chroman group in **38** projects into the hydrophobic pocket lined by I83, I86, Q87, P97 and C115, while the saturated portion of this bicyclic system is oriented towards the solvent. The cyclopentyl group in **38** is half-buried in the hydrophobic region formed by M177, A180, L184 and Y181. The repulsion between this hydrophobic pocket and the positive charged of piperidine in **37** prevents the burial of piperidine in **37** in DCN1, while the uncharged tetrahydro-2H-pyran in **36** is favorable to be buried in this pocket, which may explain, at least in part, the much higher affinity for compound **36** than for **37**.

The co-crystal structure of compound **38** complexed with DCN1 thus provides structural insights for the high binding affinity of compound **38** and its close analogues such as compound **36** to DCN1.

### 2.2.9 Cellular evaluation of compound 36 (DI-404)

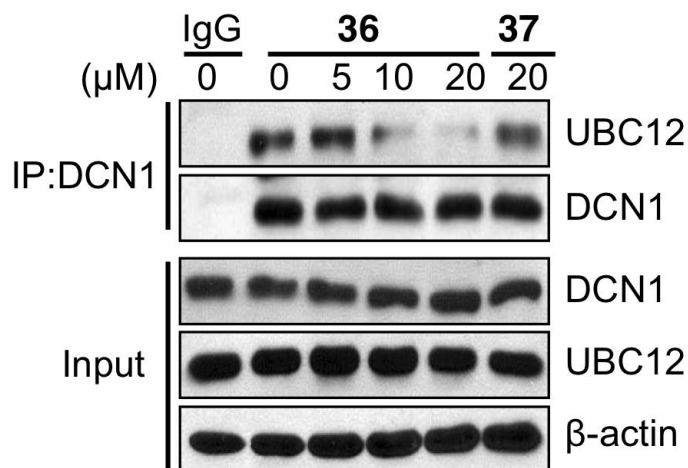
DCN1 is overexpressed in squamous cell carcinoma (SCC) cell lines.<sup>71</sup> We evaluated compound **36** in two SCC lung cancer cell lines H2170 and SK-MES-1 with compound **37** included as the control compound.

The cellular thermal shift assay (CETSA)<sup>72</sup> was employed to investigate the target engagement of **36** in cells. As shown in Figure 2.7, cellular DCN1 protein was denatured at 53 °C in H2170 cells by treatment with DMSO, but the thermal stability of DCN1 protein was clearly enhanced, in a dose-dependent manner by **36**. Compound **36** was effective in enhancing the thermal stability of cellular DCN1 protein at concentrations as low as 0.3 μM. In comparison, the much less potent control compound **37** has a minimal effect on the thermal stability of cellular DCN1 protein at concentrations as high as 10 μM (Figure 2.7).



**Figure 2.7 Engagement of compound 36 with DCN1 in cells.** Compound **36** engages cellular DCN1. H2170 cells were treated by **36** or **37** for 1 h, and then the cells were heated at 53 °C for 3 min, followed by frozen in liquid nitrogen and thawed through 3 cycles. DCN1 protein level in whole cell lysates was examined by western blotting analysis. GAPDH was used as a loading control.

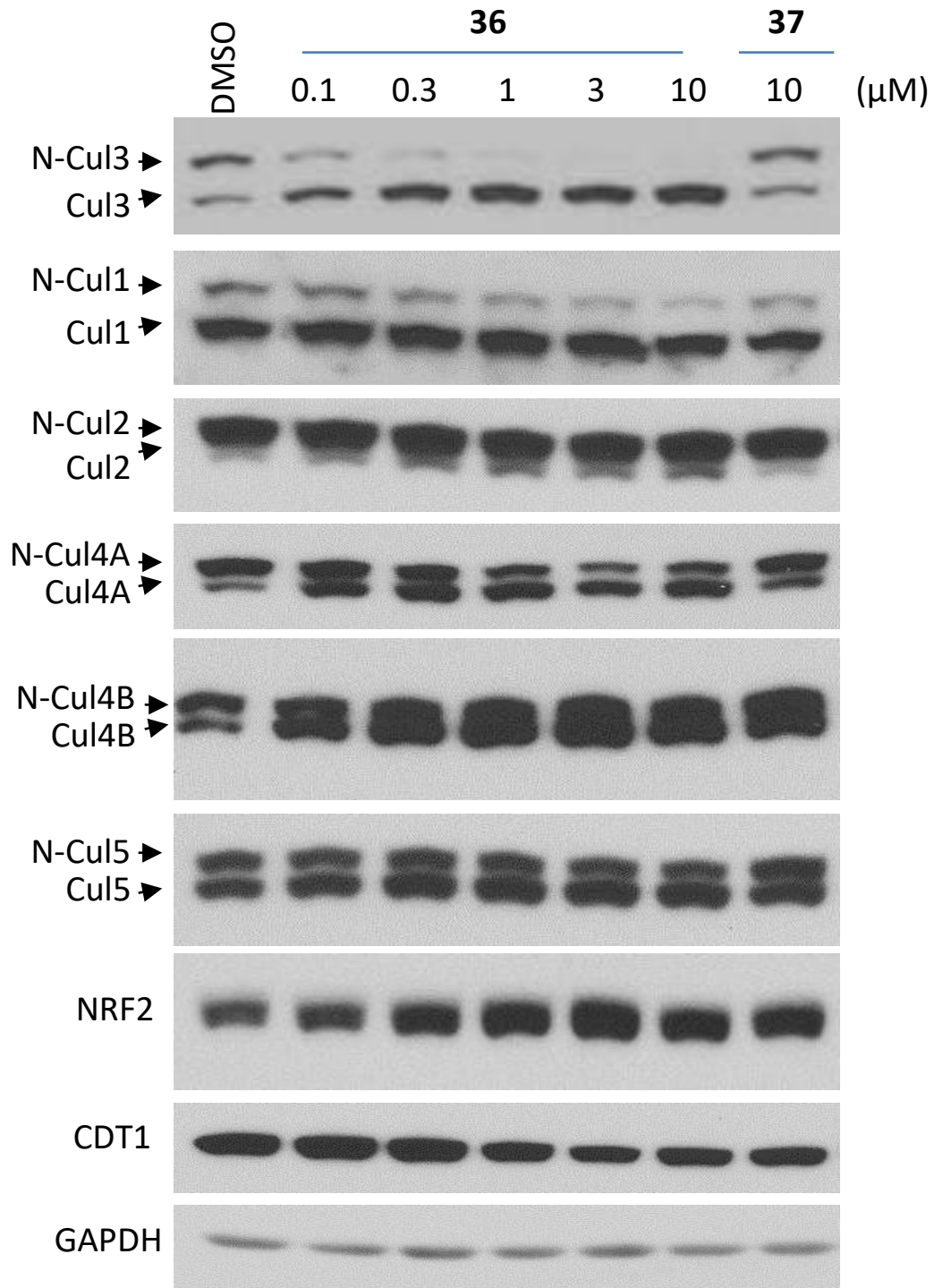
To determine if **36** disrupts the association of DCN1 protein with UBC12 protein in cells, a co-immunoprecipitation study was carried out on the lysate from H2170 cells treated with compound **36** or **37**. Compound **36** decreases the association of cellular DCN1 protein with cellular UBC12 protein in a dose-dependent manner but the control compound (**37**) had no such effect (Figure 2.8). Furthermore, compound **36** has no effect on the levels of DCN1 and UBC12 proteins (Figure 2.8). These data suggest that compound **36** effectively inhibits the DCN1-UBC12 protein-protein interaction in cells.



**Figure 2.8 Effect of compound 36 on DCN1-UBC12 protein-protein interaction in H2170 cells**

Since DCN1 protein is an integral part of the neddylation activation complex, we next investigated if compound **36** can effectively inhibit the neddylation of an individual cullin. We treated H2170 cells with compound **36**, with compound **37** included as a control. Our data showed that compound **36** inhibits the neddylation of cullin 3 in a dose-dependent manner and is effective at concentrations as low as 0.3 μM (Figure 2.9). Treatment with **36** converts cullin 3 into a predominantly unneddylated form at 1 μM in

H2170 cells (Figure 2.9).



**Figure 2.9** Effect of cullin neddylation of compound 36 in H2170 cells. Compound 36 selectively inhibits the neddylation of cullin 3 over that of other cullin members. H2170 cells were treated with 36 or 37 for 16 h. Treated cells were

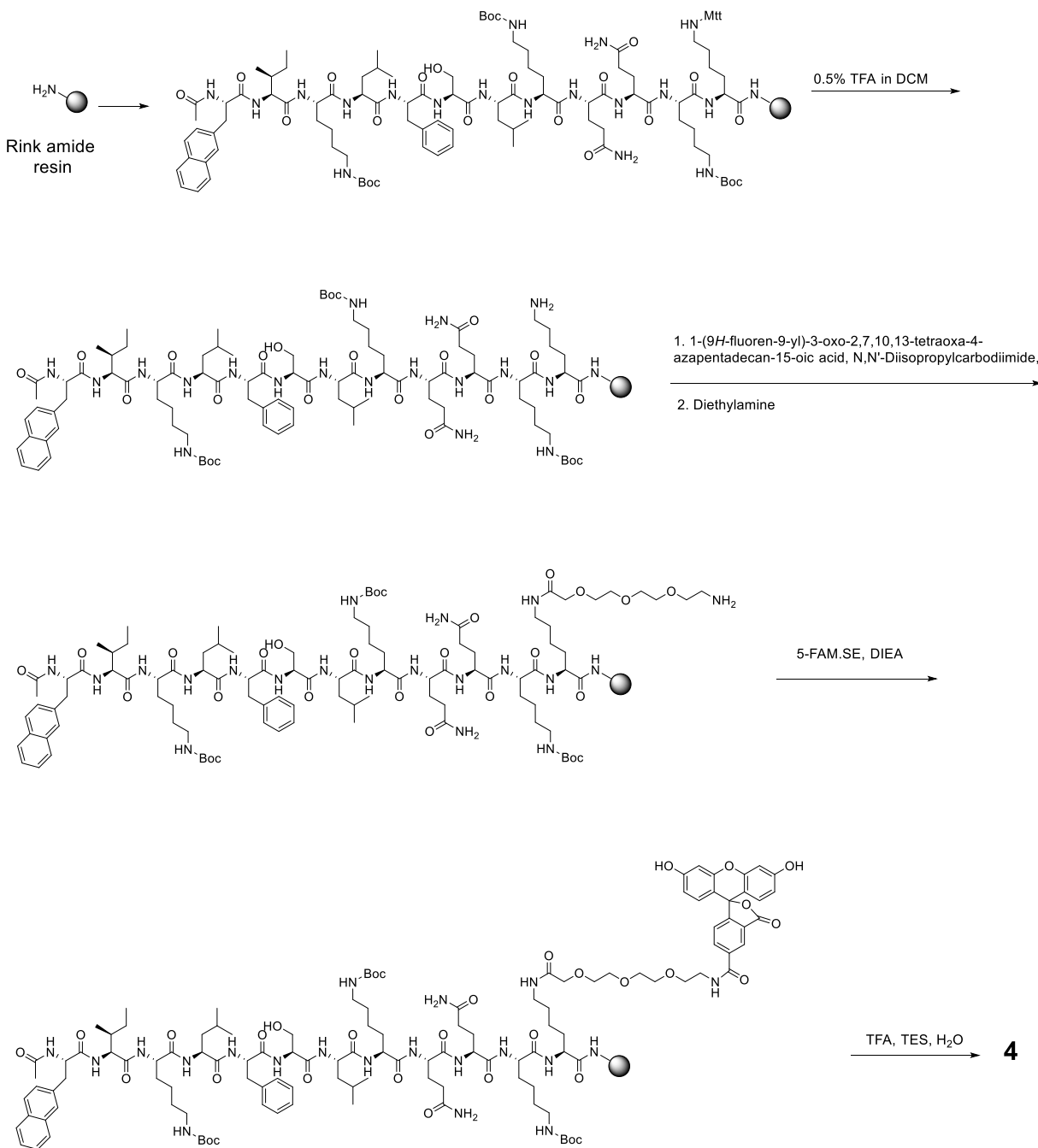
lysed. The protein levels of neddylated cullin (N-Cul) and un-neddylated cullin (Cul) proteins and two substrates of cullins, NRF2 and CDT1, in whole cell lysates were examined by western blotting analysis. GAPDH was used as a loading control.

While compound **36** effectively and potently inhibits the neddylation of cullin 3, it is much less effective in inhibition of the neddylation of cullin 1 and other cullins (Figure 2.9). In fact, it has only a modest effect in inhibition of the neddylation of cullin 1 at concentrations as high as 10  $\mu$ M and has no obvious effect on neddylation of cullins 2, 4A, 4B and 5 (Figure 2.9). Significantly, the much less potent control compound **37** has no effect on the neddylation of all cullin members examined, showing that the effect by compound **36** is specific to cullin 3 (Figure 2.9). As a consequence of selective inhibition of cullin 3 neddylation, compound **36** effectively increases the level of NRF2 protein, a substrate of cullin 3 CRL (CRL3), but has no effect on CDT1, a substrate of cullin 4A CRL (CRL4A) (Figure 2.9).

### 2.3 Chemistry

As shown in Scheme 1, a fluorescent labeled tracer (**4**) was produced by solid-phase peptide synthesis. 5-Carboxyfluorescein succinimide ester (5-FAM, SE) was attached to the side chain of the Lys12 of the linear peptide *via* a linker.

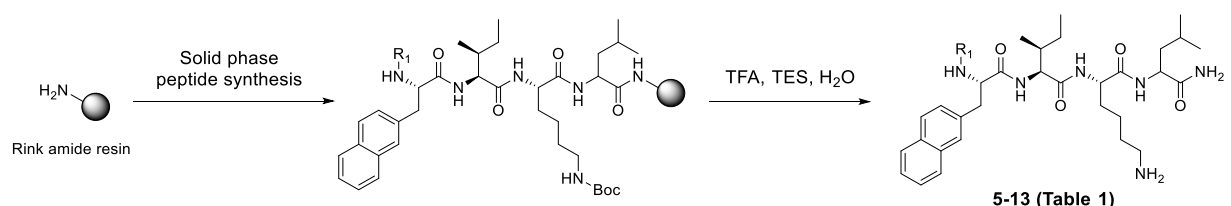
## Scheme 2.1 Synthesis of compound 4.



Peptides **5-13** were synthesized by solid-phase peptide synthesis using Fmoc chemistry as shown in Scheme 2. Rink amide resin was used as the solid support and the coupling reagent was diisopropylcarbodiimide. The crude peptides were cleaved from the resin by a TFA:TES:H<sub>2</sub>O (36:1:2) cleavage cocktail, which also removed the

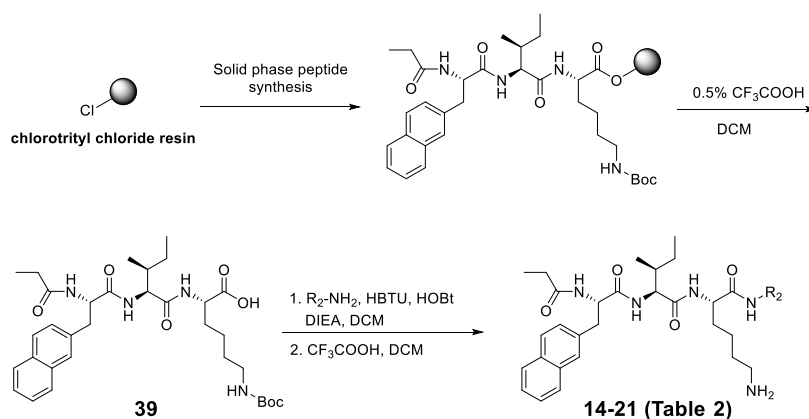
protecting groups. After the cleavage, the solvents were evaporated and the residue was purified by RP-HPLC to give compounds **5-13**.

### Scheme 2.2 Synthesis of compounds 5-13.



Compounds **14-21** were synthesized as shown in Scheme 3. The intermediate (**39**) was synthesized with 2-chlorotrityl chloride resin as the solid support. Fmoc-Lys(Boc)-OH (**40**) was first loaded onto the 2-chlorotrityl chloride resin and Fmoc chemistry was then used to carry out the chain elongation. The crude carboxylic acid (**39**) was cleaved from the resin with 0.5% TFA in DCM and purified by HPLC. It then was coupled to various amines and this was followed by removal of the Boc protecting group, providing compounds **14-21**.

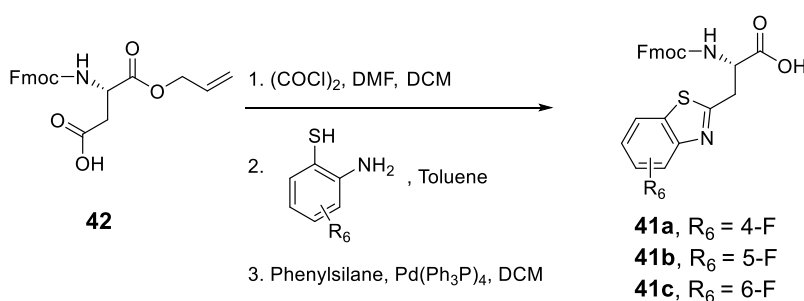
### Scheme 2.3 Synthesis of compounds 14-21.





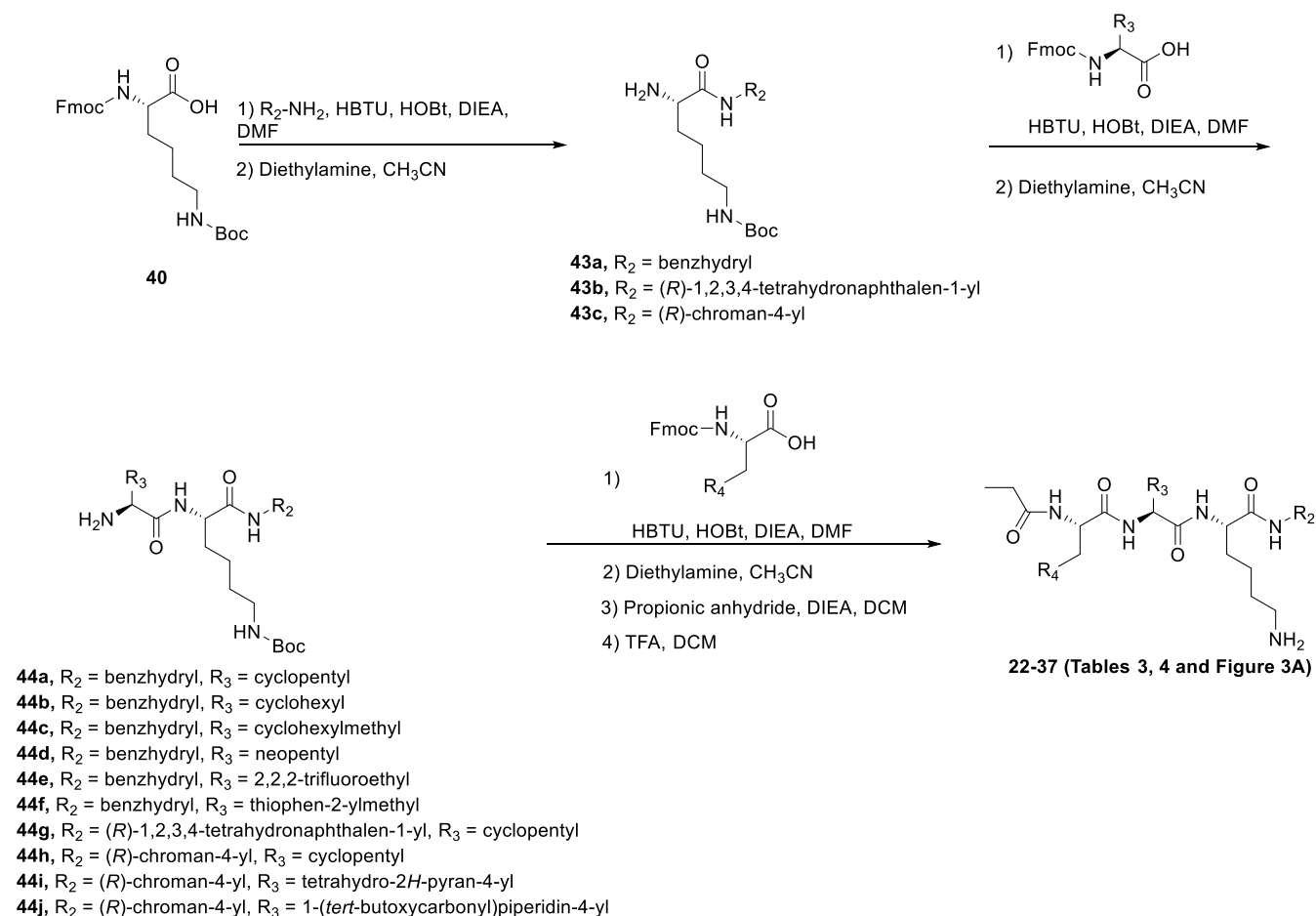
The unnatural amino acids (**41**) were synthesized as shown in Scheme 4. The acid group of Fmoc-Asp-O-allyl (**42**) was converted to different substituted benzothiazoles by treatment with oxalyl chloride, followed by the appropriately substituted 2-aminobenzenethiol.<sup>16, 73</sup> The O-allyl protecting group was removed in the presence of phenylsilane and Pd(Ph<sub>3</sub>P)<sub>4</sub>, giving compounds **41a-41c**.

**Scheme 2.4 General synthetic route for amino acids 41.**



The general method for the synthesis of compounds **22-37** is shown in Scheme 5. Compound **43** is produced by coupling Fmoc-Lys(Boc)-OH (**40**) to various amines (R<sub>2</sub>-NH<sub>2</sub>) with HBTU as coupling reagent and this is followed by removal of the Fmoc group by treatment with diethylamine. Employing a similar procedure, condensation of **43** with various Fmoc protected amino acids and subsequent deprotection of the Fmoc group gives compounds **44**. Next, various Fmoc protected amino acids are attached to **44** and a propionyl group is introduced as the N-terminal capping group. Compounds **22-37** are generated by deprotection and purification. Compound **38** is afforded by treating **35** with formaldehyde and Na(AcO)<sub>3</sub>BH.

## Scheme 2.5 General synthetic route of compounds 22-37.



d

## 2.4 Summary

Starting from a tetrapeptide (**2**) derived from the UBC12 protein, we performed systematic modifications to optimize its interactions with four pockets (P1, P2, P3 and P4) in the protein DCN1, and its physiochemical properties. Our efforts have yielded a series of potent peptidomimetic compounds, exemplified by compound **36** (DI-404), which binds to DCN1 with a  $K_D$  value of 6.7 nM. Our cellular studies demonstrate that **36** engages cellular DCN1 protein and inhibits the association of cellular DCN1 protein with the UBC12

protein. The cellular data further show that while **36** inhibits the neddylation of cullin 3 in a dose-dependent manner, it is much less effective in inhibition of the neddylation of cullin 1 and has no effect on the neddylation of other cullin members that were examined. Determination of a co-crystal structure of compound **38**, a potent and close analogue of **36**, provides structural insights for their high-affinity interactions with DCN1 and a structural basis for further design and optimization. Optimization of **36** (DI-404) may ultimately yield new therapies for the treatment of human conditions and diseases in which inhibition of cullin 3 CRL may provide a benefit.

## Chapter 3 Discovery of Potent and Selective DCN3 Inhibitors

### 3.1 Introduction

Protein degradation is tightly regulated by the ubiquitin-proteasome system (UPS) to maintain cellular homeostasis.<sup>1</sup> In mammalian cells, the Cullin-Ring ubiquitin ligases (CRLs) are the most prominent family of E3 ligases in the UPS.<sup>6, 7</sup> There are seven cullin complexes (CUL1, CUL2, CUL3, CUL4A, CUL4B, CUL5 and CUL7) in human cells with an individual cullin protein being the central component of each CRL. Activation of each CRL requires covalent attachment of NEDD8 (neural precursor cell expressing developmentally downregulated-8), a process known as neddylation. This neddylation cascade starts with the activation of NEDD8 by the NEDD8 activating enzyme (NAE), the E1 enzyme. The activated NEDD8 transfers to either UBC12 or UBE2F, the two E2 enzymes found in mammalian cells. In the final step, the E3 enzyme catalyzes the transfer of NEDD8 to the substrate cullins, resulting in the activation of a CRL.

Because CRLs are responsible for degradation of some 20% of mammalian proteins, they play key roles in the regulation of a large number of biological processes.<sup>7</sup> The dysregulation of CRLs has been linked to several human diseases, including cancer.<sup>60, 74</sup> Consequently, targeting CRLs represents a new strategy for the development of cancer therapeutics. The pioneering work in targeting CRLs by Millenium Pharmaceuticals Inc. has led to the discovery and clinical development of MLN4924,

which is a covalent inhibitor of the NAE E1 enzyme and is potent and effective in inhibiting all CRLs.<sup>8</sup> However, selective targeting of individual CRLs has been very challenging. Recently, we proposed designing small-molecule inhibitors to target protein-protein interactions within the multi-component E3 ligase complexes as a strategy to target individual CRLs selectively.<sup>75, 76</sup> Our efforts led to the discovery of small-molecule inhibitors of the DCN1-UBC12 protein-protein interaction, which selectively inhibits neddylation of Cullin 3.<sup>9, 15</sup> These recent findings therefore suggest that targeting other protein-protein interactions within the multi-component E3 ligase complexes may be an attractive approach to selectively target individual CRLs.

There are 5 family members of DCN proteins (DCN1-5) in human cells,<sup>77-79</sup> and each DCN protein has a different N-terminal domain and a conserved C-terminal potentiating neddylation (PONY) domain. While the function of DCN1 has been extensively investigated, only limited studies on DCN3 protein have been reported.<sup>79,80, 81</sup> Biophysical data and x-ray crystallographic data show that while DCN1 binds to both UBE2F and UBC12 with similar affinities, DCN3 protein binds more strongly to UBE2F than to UBC12, and DCN1 and DCN3 have some differences in their binding residues that interact with UBE2F and UBC12. Therefore, it may be possible to design small-molecule inhibitors, hereafter called DCN3 inhibitors that are capable of selectively binding to DCN3 over DCN1 and blocking the interactions of DCN3 with its binding partners.<sup>82</sup>

In the present study, we report our discovery of first-in-class potent, selective and cell-permeable small-molecule DCN3 inhibitors.

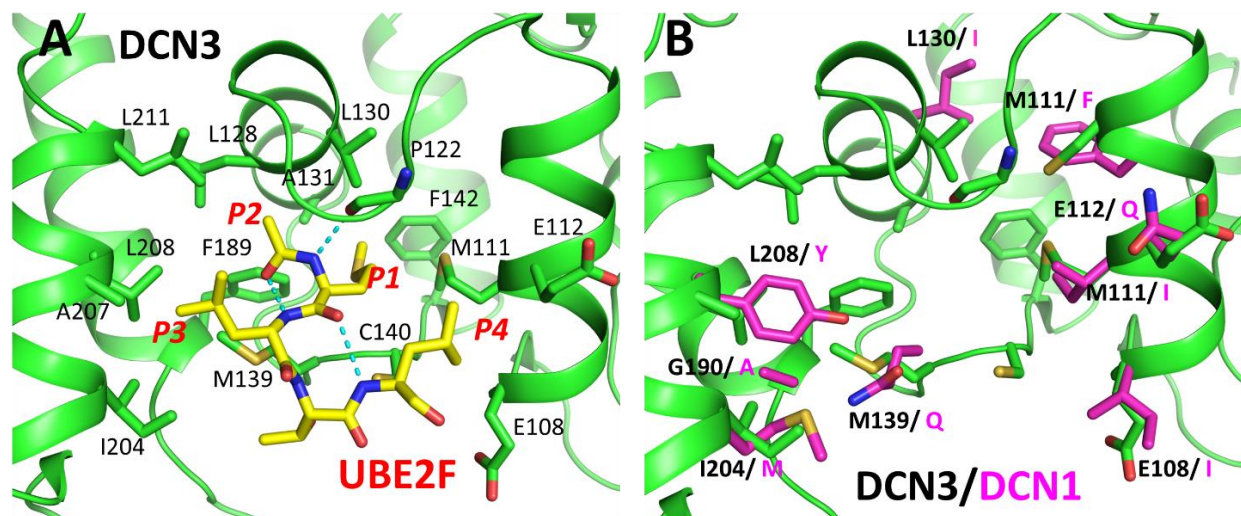
## 3.2 RESULTS AND DISCUSSION

### 3.2.1 Analysis of the binding pockets in DCN1 and DCN3

The co-crystal structure of DCN1 in a complex with its binding partner UBC12 and the co-crystal structure of DCN3 in a complex with an acetylated N-terminal UBE2F peptide have been reported.<sup>82</sup> We analyzed the binding pockets in DCN1 and DCN3 to facilitate the design of inhibitors selective for DCN3 over DCN1.

The co-crystal structures revealed that both DCN1 and DCN3 proteins interact primarily with the acetylated N-terminal tetrapeptide segment of either UBC12 or UBE2F. For our design, we defined the DCN1 and DCN3 binding sites into four pockets, P1-P4 (**Figure 3.1A**). The methionine residues in UBC12 and UBE2F penetrate deeply into the conserved, hydrophobic P1 pocket in both DCN1 and DCN3, respectively. The hydrophobic residues forming this P1 pocket are the same in DCN1 and DCN3 with the exception of two hydrophobic residues: Leu130 in DCN1 becomes Ile105 in DCN3, and Ile86 in DCN1 is replaced by Met111 in DCN3. The P2 pocket in DCN1 and DCN3 interacts with the N-terminal acetyl group in both UBC12 and UBE2F and is conserved with a single residue change: Tyr181 in DCN1 becomes Leu208 in DCN3. The P3 pocket in both DCN1 and DCN3 is hydrophobic in nature and partly solvent exposed and interacts with isoleucine and leucine residue in UBC12 and UBE2F, respectively. The P3 pocket in DCN1 and DCN3 differs with two residues, a change of tyrosine (Tyr181) in DCN1 to leucine (Leu208) in DCN3, and a change of alanine (Ala98) in DCN1 to threonine (Thr123) in DCN3. The P4 pocket in DCN1 and DCN3 is also partially solvent exposed

and several residues involved in this pocket in DCN1 and DCN3 are different (**Figure 3.1B**).



**Figure 3.1 Crystal structure analysis of DCN1 and DCN3.** (A) Crystal structure of DCN3 with the UBE2F peptide (PDBID: 4GBA). Key binding residues in DCN3 and the UBE2F peptide are shown in stick. Hydrogen bonds are depicted in cyan dashed lines. (B) Superposition of crystal structures of DCN3:UBE2F (PDBID: 4GBA) and DCN1:UBC12 (PDBID: 3TDU). Key residues differences in the binding site of DCN1 and DCN3 are labeled and colored in pink and green, respectively.

We propose that these differences in binding site residues between DCN1 and DCN3 can be exploited for the design of selective inhibitors for either DCN1 or DCN3. Indeed, in our previous study, we have designed DI-591 as a small-molecule inhibitor which is highly selective for DCN1 over DCN3.<sup>75</sup>

### 3.2.2 Identification of initial DCN3 inhibitors by screening our DCN1 inhibitor library

In our previous studies, we designed and synthesized a series of small molecules as inhibitors of DCN1.<sup>75, 76</sup> Since DCN1 and DCN3 share similar binding sites, we reasoned that some of our previously synthesized DCN1 inhibitors may display significant binding affinities to DCN3.

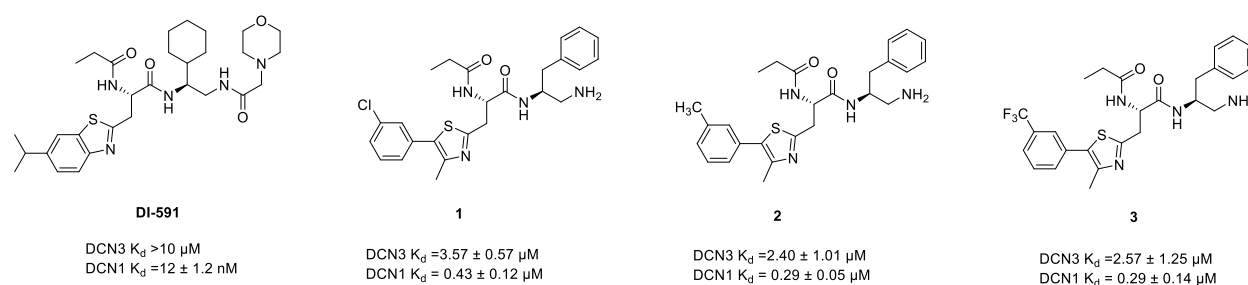
We employed a label-free, biolayer interferometry (BLI) method to screen our previously synthesized small-molecule DCN1 inhibitors for their binding affinities to recombinant, human DCN3 protein. To validate this BLI assay, we tested the binding affinity of a known 25-residue UBE2F<sup>NAC1-25</sup> peptide for DCN3. Our BLI experiments determined that the 25-residue UBE2F<sup>NAC1-25</sup> peptide binds to DCN3 with  $K_d = 3.20 \mu\text{M}$  (Table 3.1), close to its previously reported  $K_d$  value.<sup>82</sup> We also synthesized and used the BLI assay to test a number of truncated UBE2F peptides, with 5 to 9 residues. Our data, in Table 3.1 show that UBE2F<sup>NAC1-9</sup> (first 9 residues) has a binding affinity very similar to that of UBE2F<sup>NAC1-25</sup> ( $K_d = 2.73 \mu\text{M}$  vs  $3.20 \mu\text{M}$ ), but a peptide with the first 7 residues, UBE2F<sup>NAC1-7</sup> is >30-times weaker than UBE2F<sup>NAC1-9</sup> in binding to DCN3, and still shorter peptides UBE2F<sup>NAC1-5</sup> and UBE2F<sup>NAC1-6</sup> have no appreciable binding to DCN3 at concentrations up to  $100 \mu\text{M}$ .



**Table 3.1 Binding affinities of UBE2F peptides to DCN3 as determined in a direct BLI binding assay.**

Peptide name	Sequence	$K_d$ ( $\mu\text{M}$ )
UBE2F <sup>NAC 1-25</sup>	Ac-MLTLASKLKRDDGLKGSRTAATASD-NH <sub>2</sub>	3.20 $\pm$ 0.30
UBE2F <sup>NAC 1-9</sup>	Ac-MLTLASKLK-NH <sub>2</sub>	2.73 $\pm$ 0.15
UBE2F <sup>NAC 1-7</sup>	Ac-MLTLASK-NH <sub>2</sub>	103 $\pm$ 85
UBE2F <sup>NAC 1-6</sup>	Ac-MLTLAS-NH <sub>2</sub>	> 100
UBE2F <sup>NAC 1-5</sup>	Ac-MLTLA-NH <sub>2</sub>	> 100

Using the BLI assay, we evaluated a total of 67 small molecules from our collection of previously synthesized DCN1 inhibitors for their binding affinities to DCN3. While 64 out of 67 these compounds had no obvious binding to DCN3 at concentrations up to 10  $\mu\text{M}$ , we identified three compounds (**1-3**), with very similar chemical structures, which display  $K_d$  values of 2-4  $\mu\text{M}$  (Figure 3.2).

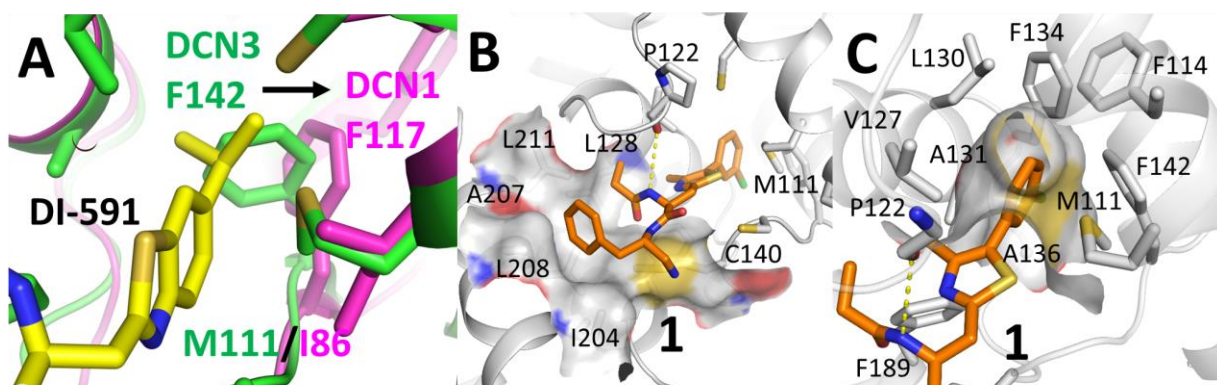


**Figure 3.2 Chemical structures and binding affinities of previously synthesized DCN1 inhibitors. Compounds 1-3 display significant binding affinities to DCN3.**

### 3.2.3 Computational prediction of the binding of lead compounds in complexes with DCN3

To guide our optimization efforts, we performed computational modeling to predict the binding models of these initial three lead compounds to DCN3. In the co-crystal structure of DI-591 in a complex with DCN1,<sup>75</sup> we observed that the F117 of DCN1 in the P1 pocket is rotated as compared to F117 in the co-crystal structure of DCN1 in complex with the UBC12 peptide and in this way a larger binding pocket to accommodate DI-591 is created. Our docking simulations indicated that the corresponding conserved residue in DCN3 would have to rotate in order to accommodate ligands such as compounds **1-3**. Therefore, we applied the rotation of F117 in DCN1 that was observed in the co-crystal structure of DI-591 in a complex with DCN1 (PDB ID: 5UF1) to a phenylalanine, F142 in DCN3 for molecular docking simulations in this study (Figure 3.3A). As described above, the major structural difference in the P1 pocket is that Met111 in DCN3 has been replaced by Ile86 in DCN1. Consistently, we observed that DCN3 has a narrower opening to the P1 pocket and prevents entry of the planar phenylthiazole ring of DI-591 (Figure 3.3A). This finding can explain why most of the compounds such as DI-591 with a fused-aromatic moiety in our DCN1 inhibitor library show no appreciable binding to DCN3 in the BLI assay.<sup>75, 76</sup> On the other hand, the docking model for compound **1** in complex with DCN3 suggests that compound **1** has a shape that permits its entry into the P1 pocket (Figure 3.3B). The phenyl ring and the thiazole ring are in planes that have a dihedral angle of ~45° between them. The thiazole ring forms an edge- $\pi$  interaction with F189. The propionamide side chain fits well into the hydrophobic pocket formed by L128, L211, and

F189, and the phenylaniline fragment in **1** enjoys hydrophobic interactions with L128, L211, A207, L208 of the P3 pocket. The junction between the m-chlorophenyl ring and the thiazole ring in **1** is surrounded by V127 and M111 (Figure 3.3C), and the m-chlorophenyl ring is surrounded by multiple hydrophobic residues in the deep P1 pocket of DCN3, including A131, L130, F134, F114, and F142 (Figure 3.3C).



**Figure 3.3 Modeling of DCN inhibitors on DCN1 and DCN3 proteins.** (A) Rotation of F117 observed in the crystal structure of DI-591 in complex with DCN1 (PDBID: 5UFI) was applied to F142 in DCN3 to obtain docked poses of identified lead compounds **1-3** with DCN3. (B) Docking models of compound **1** with DCN3. Hydrogen bonds are depicted in yellow dashed lines. (C) Detailed view of the interaction of compound **1** with the P1 pocket of DCN3.

### 3.2.4 Development of a fluorescence polarization-based, competitive binding assay for DCN3

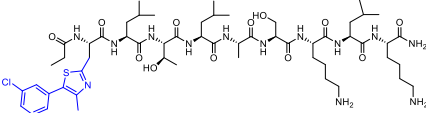
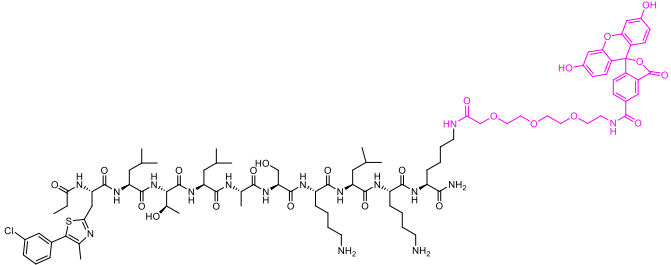
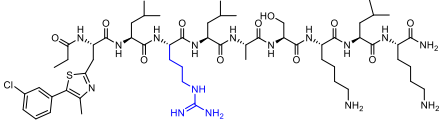
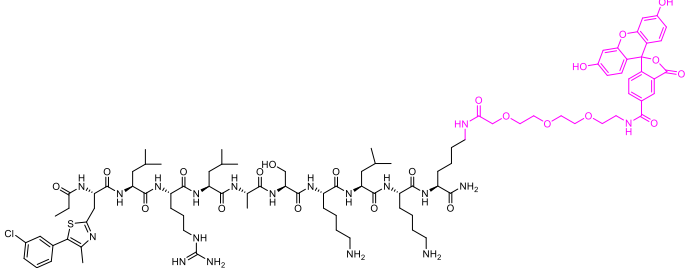
The ligand-free, BLI method is time-consuming and expensive and we developed a fluorescence polarization (FP)-based, competitive binding assay for DCN3 which supports rapid and accurate determination of the binding affinities of our synthesized DCN3 inhibitors.<sup>7,8</sup>

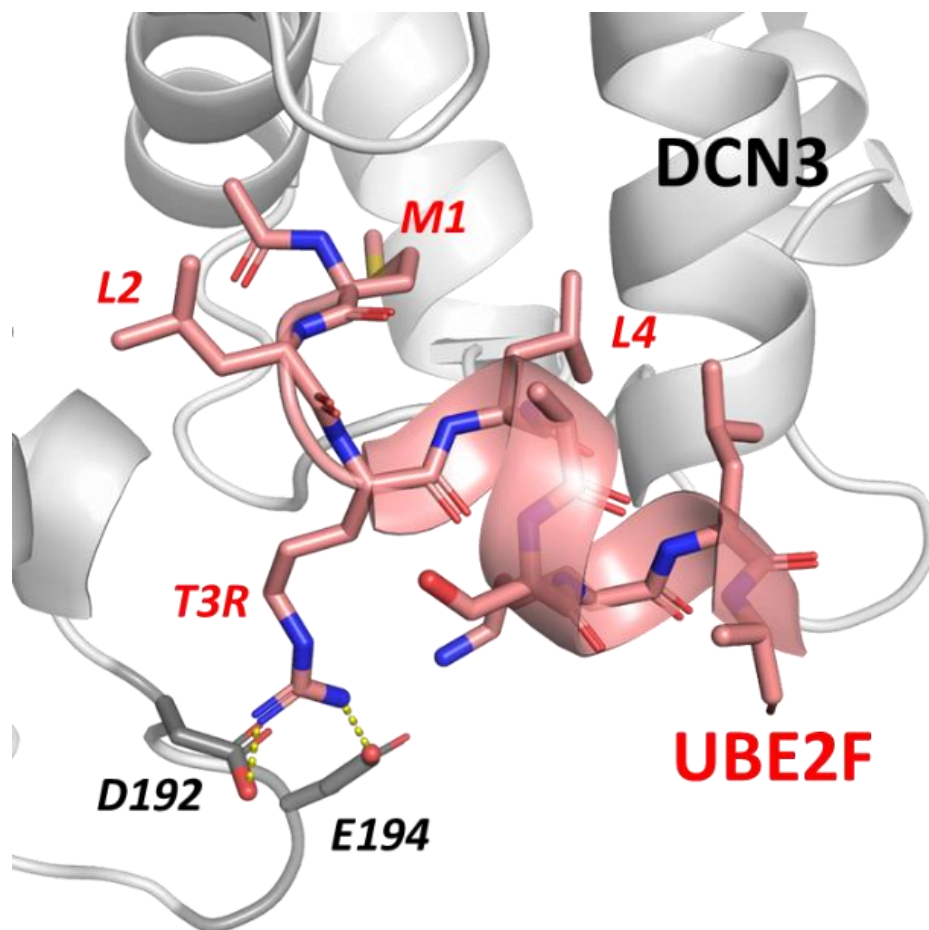
Our truncation analysis shows that in the BLI assay, the 9-residue UBE2F<sup>NAC1-9</sup> peptide binds to DCN3 with a  $K_d$  value of 2.73  $\mu$ M (Table 3.1), which is comparable to that of the 25-residue UBE2F<sup>NAC1-25</sup> peptide ( $K_d = 3.2 \mu$ M). After further shortening of the length of the 9-residue peptide producing the UBE2F<sup>NAC1-7</sup> peptide, a dramatic decrease of binding affinity to DCN3 is observed (Table 3.1). We therefore employed the 9-residue UBE2F<sup>NAC1-9</sup> peptide as the template for the development of a FP assay.

In our modeled structures of these three initial DCN3 lead compounds, the substituted phenylthiazole group occupies the P1 binding pocket (Figure 3.3B). To obtain a high-affinity ligand for DCN3, we replaced the methionine residue side chain in the UBE2F<sup>NAC1-9</sup> peptide with a 3-chlorophenyl-4-methylthiazole group. The resulting compound, TC-6038, has a  $K_d$  value of 75.7 nM, and is 36 times more potent than the UBE2F<sup>NAC1-9</sup> peptide. Based on TC-6038, we added a lysine residue at the C-terminus to tether with 5-carboxyfluorescein (5-FAM), which led to the design of a fluorescently tagged peptide tracer (DCN3-Flu0). Unfortunately, the tracer DCN3-Flu0 has a  $K_d$  value of 484 nM in the BLI assay, which is adequate but not ideal for the development of a sensitive and accurate FP assay. In the crystal structure of DCN3 in complex with UBE2F, the Thr3 residue of UBE2F does not interact directly with DCN3, but there are two negatively charged residues, D192 and E194 approximately 7.5 Å away from the T3 residue. Modeling suggested that mutation of the Thr3 residue to arginine would present additional hydrogen bonding interactions with D192 and E194. (Figure 3.4) Accordingly, we replaced the threonine residue in TC-6038 with arginine and the resulting compound, TC-6052, was found to have a  $K_d$  value of 45.7 nM, 2-times more potent than TC-6038. Based upon TC-6052, we synthesized DC3-Flu1 as a new tracer, which was determined

to have a  $K_d$  value of 184 nM in the BLI assay (Table 3.2). Our titration experiments determined that DC3-Flu1 has a  $K_d$  value of 139 nM to DCN3 (Figure 3.5), consistent with the  $K_d$  value obtained from the BLI binding assay. Importantly, the  $\Delta mP$  value for DC3-Flu1 is 120, which provides a large signal-to-noise ratio. Consequently, we have developed a competitive FP assay using DC3-Flu1 for determination of the binding affinities of our designed DCN3 inhibitors.

**Table 3.2 Binding Affinity of Mutated UBE2F Peptides and the Tracers Determined in BLI Binding Assay**

Compound	Structure	$K_d$ (nM)
TC-6038		$85.8 \pm 11.6$
DCN3-Flu0		$484 \pm 175$
TC-6052		$46 \pm 11$
DCN3-Flu1		$184 \pm 42$



**Figure 3.4 Modeled truncated UBE2F with T3R mutation with DCN3** Modeled structure of the T3R mutated UBE2F peptide with DCN3 based upon the co-crystal structure of DCN3 in complex with a UBE2F peptide (PDBID: 4GBA). DCN3 and the mutated UBE2F peptide are labeled in black and red respectively. Hydrogen bonds are depicted in yellow dashed lines.

We evaluated the binding affinities of compounds **1-3** to DCN3 in the competitive FP assay. Compounds **1**, **2** and **3** have  $IC_{50}$  values of 3.2, 4.5 and 5.2  $\mu M$ , respectively, and have calculated  $K_i$  values of 0.67, 0.43 and 0.53  $\mu M$ , respectively. We employed compounds **1-3** as our initial leads for further optimization.

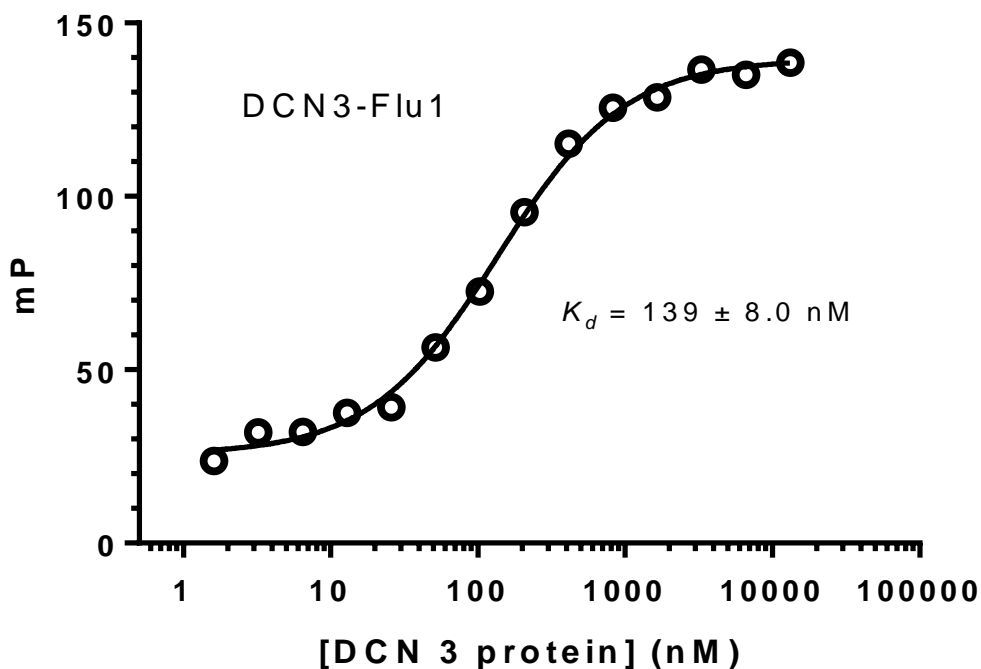


Figure 3.5 Binding isotherm of DCN3-Flu1 to DCN3 protein.

### 3.2.5 Structure-activity relationship studies of the P1 binding pocket

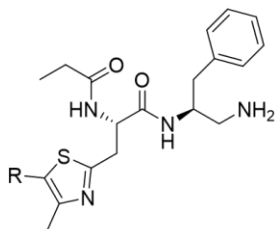
In our predicted binding model for compound **1** in a complex with DCN3, the *m*-chloro substituent on the phenyl ring fits into the hydrophobic P1 pocket formed by A136, A131, F134 and F142 (Figure 3.3C). To further probe the importance of the *meta*-substitution, we synthesized a series of compounds.

Removal of the *m*-chloro substituent from the phenyl ring in compound **1** gave compound **4**, which shows no significant binding up to 10  $\mu$ M. Changing the position of the chlorine substituent from the *meta* to the *ortho*- or the *para*-position yielded compounds **5** and **6** respectively, neither of which show any appreciable binding to DCN3 at concentrations up to 10  $\mu$ M. Replacement of the substituted phenyl in compounds **1-3**



with a cyclohexyl group resulted in compound **7**, which also has no binding affinity to DCN3 up to 10  $\mu\text{M}$  (Table 3.2).

**Table 3.3 Structure-activity relationship studies to probe the P1 binding pocket.**

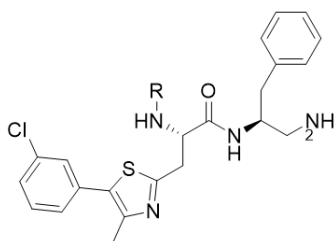


ID		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
R								
DCN3	IC <sub>50</sub> ( $\mu\text{M}$ )	3.2 $\pm$ 2.3	4.6 $\pm$ 0.8	5.2 $\pm$ 0.2	>10	>10	>10	>10
	K <sub>i</sub> ( $\mu\text{M}$ )	0.94 $\pm$ 0.76	1.4 $\pm$ 0.2	1.6 $\pm$ 0.1				
DCN1	IC <sub>50</sub> ( $\mu\text{M}$ )	0.67 $\pm$ 0.03	0.43 $\pm$ 0.05	0.53 $\pm$ 0.02	5.2 $\pm$ 0.2	>10	2.3 $\pm$ 0.3	>10
	K <sub>i</sub> ( $\mu\text{M}$ )	0.019 $\pm$ 0.01	0.07 $\pm$ 0.01	0.95 $\pm$ 0.01	1.6 $\pm$ 0.02		0.29 $\pm$ 0.03	

### 3.2.6 Structure-activity relationship studies of the P2 binding pocket

In our design of DCN1 inhibitors, we showed that the propionamide group in DI-591, which interacts with the P2 pocket, is optimal for binding to DCN1.<sup>75, 76</sup> Our analysis of the crystal structures showed that DCN1 and DCN3 are significantly different in their P2 pockets (Figure 3.3A). Accordingly, we made extensive modifications to our DCN3 inhibitors with the objective of improving the binding affinity to DCN3 and at the same time reducing the binding affinity to DCN1 to obtain selective DCN3 inhibitors. The obtained SAR data are summarized in Table 3.4.

**Table 3.4 Structure-activity relationship studies to probe the P2 binding pocket.**



ID		<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>
R							
DCN3	IC <sub>50</sub> (μM)	>10	>10	>10	1.1 ± 0.2	1.7 ± 0.3	>10
	K <sub>i</sub> (μM)				0.28±0.05	0.47 ± 0.11	
DCN1	IC <sub>50</sub> (μM)	2.36±0.39	1.32±0.16	>10	7.69±1.15	>10	>10
	K <sub>i</sub> (μM)	0.75±0.14	0.38±0.08		2.49±0.38		
ID		<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>
R							
DCN3	IC <sub>50</sub> (μM)	3.85 ± 1.51	0.92 ± 0.07	4.61±0.35	0.56± 0.06	0.84 ±0.07	0.80 ± 0.03
	K <sub>i</sub> (μM)	1.17 ± 0.48	0.23 ± 0.02	1.41 ±0.11	0.11 ±0.02	0.20 ±0.02	0.19 ±0.01
DCN1	IC <sub>50</sub> (μM)	>10	>10	>10	5.34 ±1.08	>10	>10
	K <sub>i</sub> (μM)				1.70±0.36		

Changing the propionamide group to acetamide resulted in compound **8**, which has no appreciable binding to DCN3 up to 10 μM but still binds to DCN1 with an IC<sub>50</sub> value of 2.4 μM. Replacing the propionamide group with an isobutyramide group yielded compound **9**, which shows no binding to DCN3 up to 10 μM but binds to DCN1 with an IC<sub>50</sub> value of 1.3 μM. Changing the propionamide group to benzamide yielded compound **10**, which fails to bind to either DCN3 or DCN1 up to 10 μM.

In the co-crystal structure of DCN1 in a complex with DI-591, the carbonyl group of the propionamide group in DI-591 forms a hydrogen bond with the hydroxyl group of the Tyr181 side chain in DCN1. (Figure 3.3B) The Tyr181 residue in DCN1 is replaced by a phenylalanine residue (Phe189) in DCN3. Consequently, we have synthesized compound **11**, in which the carbonyl group in compound **10** has been removed. Compound **11** binds to DCN3 with an IC<sub>50</sub> value of 1.1 μM (K<sub>i</sub> = 276 nM) and binds to DCN1 with an IC<sub>50</sub> value of 7.7 μM (K<sub>i</sub> = 2.4 μM). Hence, compound **11** is a moderately potent DCN3 inhibitor which, based upon the K<sub>i</sub> values shows a 9-fold selectivity over DCN1.

We installed a chlorine atom at the *meta*, *ortho*, or *para*-position of the phenyl ring in **11**, obtaining compounds **12**, **13** and **14**, respectively. Compound **12** with a *m*-chlorine and **14** with a *p*-chlorine bind to DCN3 with IC<sub>50</sub> values of 1.6 μM and 3.9 μM, respectively. However, **13** with an *o*-chlorine shows no binding to DCN3 up to 10 μM. We then placed a fluorine atom at the *meta*, *ortho*, or *para*-position of the phenyl ring in **12** and obtained **15**, **16** and **17**, respectively. Compounds **15**, **16** and **17** bind to DCN3 with IC<sub>50</sub> values of 0.92 μM, 4.6 μM and 0.56 μM, respectively. Encouraged by these results, we synthesized the 3,4-difluoro compound **18** and the 3,5-difluoro compound **19** and found that they bind to DCN3 with IC<sub>50</sub> values of 0.83 and 0.79 μM, respectively.

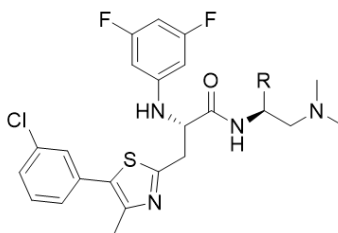
We tested the binding affinities of these potent DCN3 inhibitors (**15 – 19**) to DCN1 to determine their selectivity. All these compounds have IC<sub>50</sub> value >10 μM, except **17** which has IC<sub>50</sub> value of 5.38 μM. While **15**, **16** and **18** show partial inhibition of DCN1 at 10 μM, **19** has no appreciable binding to DCN1 up to 10 μM. Hence, compound **19** is a

potent ( $K_i = 190$  nM) and selective DCN3 inhibitor over DCN1. We next performed further optimization on compound **19**.

### **3.2.7 Structure-activity relationship studies of the P3 binding pocket**

In both DCN1 and DCN3, the P3 pocket is a shallow, hydrophobic pocket, partially exposed to solvent. While the P3 pocket in DCN1 interacts with an isoleucine residue in UBC12, the P3 pocket in DCN3 interacts with a leucine residue in UBE2F. Comparison of the DCN1 and DCN3 P3 pockets showed that DCN3 has a slightly larger P3 pocket than DCN1. We performed modifications of **19** to probe the P3 pocket in DCN3, obtaining the results in Table 3.5.

**Table 3.5 Structure-activity relationship studies to probe the P3 binding pocket**



ID		<b>20</b>	<b>21</b>	<b>22</b>	<b>23</b>	<b>24</b>
R			H <sub>3</sub> C <sub>2</sub>			
DCN3	IC <sub>50</sub> (μM)	0.72 ± 0.02	>10	1.18 ± 0.37	0.37 ± 0.03	0.32 ± 0.06
	K <sub>i</sub> (μM)	0.15 ± 0.02	>10	0.31 ± 0.12	0.053 ± 0.014	0.034 ± 0.019
DCN1	IC <sub>50</sub> (μM)	>10	>10	>10	>10	>10
ID		<b>25</b>	<b>26</b>	<b>27</b>	<b>28</b>	<b>29</b>
R						
DCN3	IC <sub>50</sub> (μM)	0.32 ± 0.06	2.99 ± 0.66	1.25 ± 0.45	0.70 ± 0.25	0.32 ± 0.06
	K <sub>i</sub> (μM)	0.034 ± 0.019	0.89 ± 0.21	0.33 ± 0.14	0.16 ± 0.08	0.034 ± 0.019
DCN1	IC <sub>50</sub> (μM)	>10	>10	>10	>10	>10
ID		<b>30</b>	<b>31</b>	<b>32</b>	<b>33</b>	<b>34</b>
R						
DCN3	IC <sub>50</sub> (μM)	0.32 ± 0.01	3.90 ± 0.01	>10	>10	>10
	K <sub>i</sub> (μM)	0.035 ± 0.002	1.19 ± 0.01			
DCN1	IC <sub>50</sub> (μM)	>30	>10	>10	>10	>10

For modifications of this site, we changed the primary amine in compound **19**, which is directed towards the solvent exposed area to a dimethylamine with the objective of improving the cell permeability<sup>83</sup>. This yielded compound **20**, which binds to DCN3 with an IC<sub>50</sub> value of 0.72 μM (K<sub>i</sub> = 150 nM), similar to **19**. Compound **20** shows no appreciable

binding to DCN1 up to 10  $\mu\text{M}$ . Hence, **20** has a good affinity to DCN3 and an excellent binding selectivity over DCN1, so we employed **20** as the new template for subsequent modification.

To assess the importance of the hydrophobic interactions of the benzyl group in compound **19**, we synthesized **21** in which the benzyl group has been replaced by a methyl group. Compound **21** binds to DCN3 with an  $\text{IC}_{50}$  value  $>10 \mu\text{M}$ , confirming the importance of the hydrophobic interactions with the P3 pocket. Replacing the methyl group in **21** with isobutyl resulted in **22**, which has an  $\text{IC}_{50}$  value of 1.2  $\mu\text{M}$  to DCN3 and only a very weak affinity for DCN1 ( $\text{IC}_{50} >10 \mu\text{M}$ ). Changing the methyl group in **21** to a cyclohexyl yielded **23**, which has a good affinity for DCN3 ( $\text{IC}_{50} = 368 \text{ nM}$ ) and a poor affinity for DCN1 ( $\text{IC}_{50} > 10 \mu\text{M}$ ). Replacement of the methyl group in **21** with a methylcyclohexyl group led to **24**, which binds to DCN3 with an  $\text{IC}_{50}$  value of 316 nM and has a weak affinity to DCN1 ( $\text{IC}_{50} > 10 \mu\text{M}$ ).

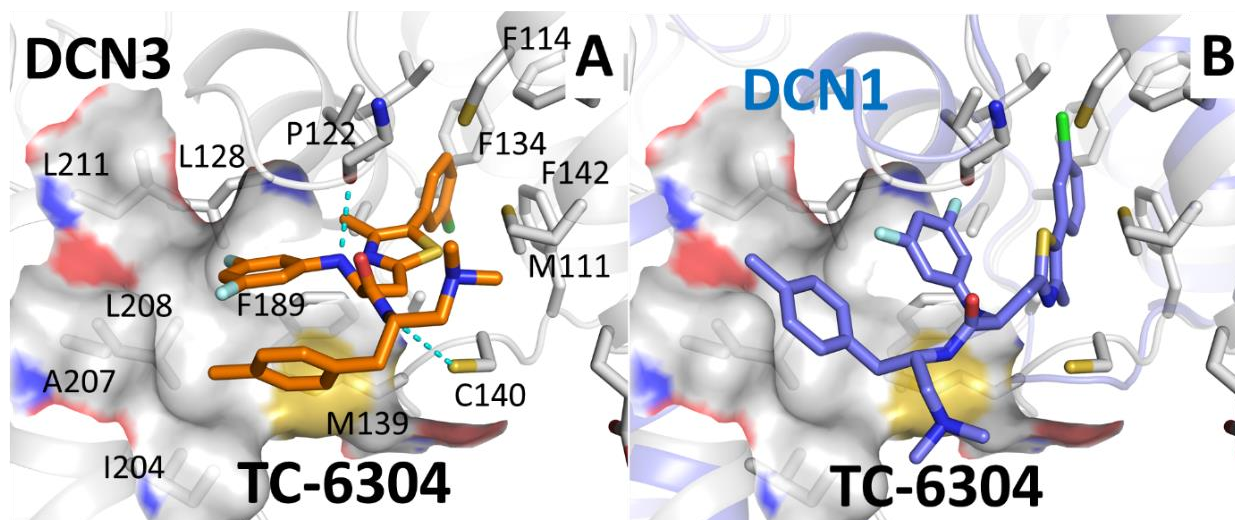
We next performed modifications of the phenyl group in **20** by installation of a chlorine substituent at the *ortho*-, *meta*- or *para*-position, which yielded **25**, **26** and **27**, respectively. Compounds **25**, **26** and **27** bind to DCN3 with  $\text{IC}_{50}$  values of 2.99, 1.25 and 0.70  $\mu\text{M}$ , respectively and have no appreciable binding to DCN1 up to 10  $\mu\text{M}$ . Because **27** with a *p*-chloro substituent has a better affinity to DCN3 than **25** and **26** with *o*-chloro or *m*-chloro substituents, we evaluated other groups at the *para*-position. We synthesized compound **28** with a *p*-fluoro, **29** with a *p*-trifluoromethyl, **30** with a *p*-methyl, **31** with a *p*-ethyl and **32** with a *p*-hydroxyl. Compounds **28**, **29**, **30**, **31**, and **32** bind to DCN3 with  $\text{IC}_{50}$  values of 1.2, 6.5, 0.31, 3.9 and  $>10 \mu\text{M}$ , respectively. These five compounds show no appreciable binding to DCN1 up to 10  $\mu\text{M}$ . In fact, **30** (TC-6304) with a *p*-methyl

substituent on the phenyl ring only shows 15% inhibition of DCN1 in the competitive FP assay at 30  $\mu\text{M}$ . Hence, **30** (TC-6304) with a *p*-methyl substituent on the phenyl ring demonstrates a good binding affinity to DCN3 ( $\text{IC}_{50} = 0.31 \mu\text{M}$  and  $K_i = 35 \text{ nM}$ ) and a binding selectivity of > 100-fold over DCN1.

We further probed this P3 pocket by synthesizing compounds **33** and **34** which both contain a larger, hydrophobic group. Compound **33** with a 2,3-dihydro-1*H*-indene and **34**, containing a 3-methyl-1*H*-indole group have  $\text{IC}_{50}$  values of >10  $\mu\text{M}$  to DCN3, confirming that these groups are too large to be accommodated in the P3 pocket.

### 3.2.8 Computational docking study of (TC-6304)

TC-6304 binds to DCN3 with a high binding affinity and demonstrates an excellent binding selectivity over DCN1. To gain an insight into the high binding affinity of TC-6304 to DCN3, we performed a computational docking study of TC-6304 and DCN3.



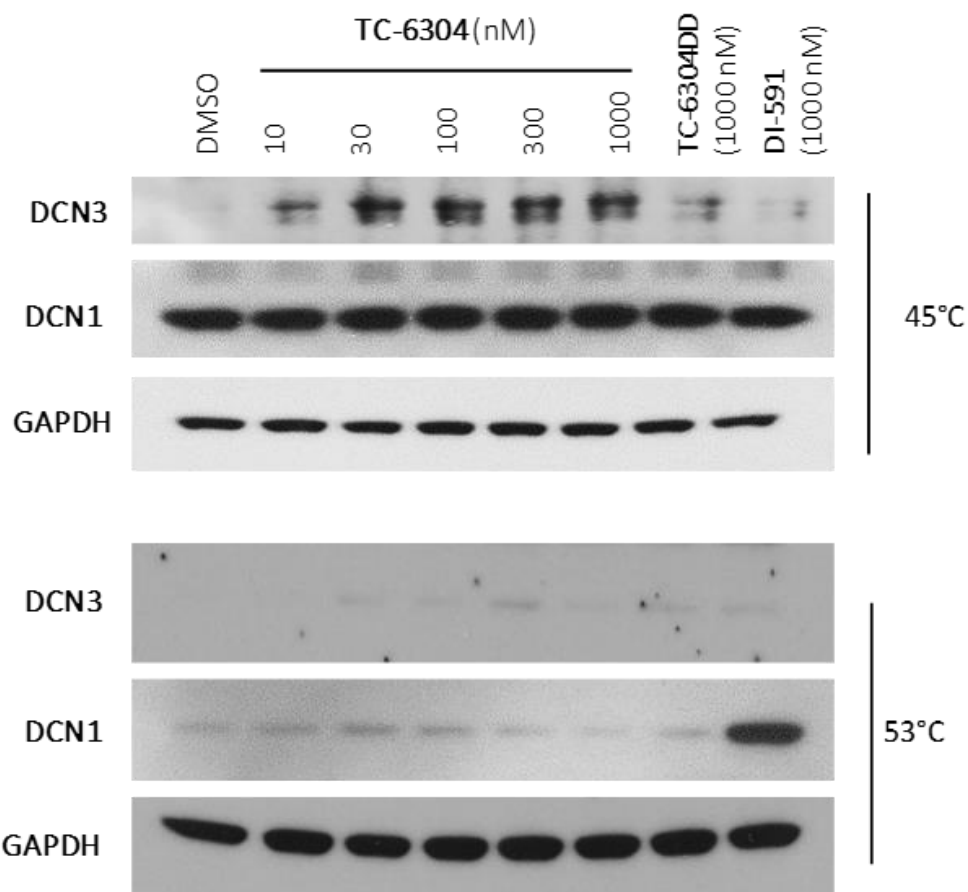
**Figure 3.6 Computational Docking of TC-6304 with DCN1/3** (A) Docking model of TC- 6304 with DCN3. Hydrogen bonds are depicted in cyan dashed lines. (B) Alignment of the docking model of TC-6304 with DCN1 (purple color) to DCN3.

The predicted binding model for TC-6304 in a complex with DCN3 (**Figure 3.6**) shows that the *m*-chlorophenyl thiazole moiety interacts with the residues forming the P1 pocket in a highly complementary manner. The *m*-chlorophenyl and thiazole moieties adopt a twisted relationship with a dihedral angle of 45° in order to fit into the P1 pocket. A hydrogen bond is observed between the backbone carbonyl of P122 of DCN3 and the NH moiety of the 3,5-difluoroaniline group of TC-6304. The 3,5-difluoroaniline group fits nicely into the P2 pocket and has close hydrophobic contacts with the L208, L211 and L128 residues of DCN3. The 4-methyl-benzyl group interacts with residues F189, L208, A207, I204 and M139 which form the P3 pocket. The dimethylamino group is exposed to solvent and has no specific interactions with DCN3.



### 3.2.9 Evaluation of the interaction of TC-6304 with DCN3 protein in cells

The cellular thermal shift assay (CETSA)<sup>72</sup> is a powerful assay with which to determine if a small molecule ligand interacts with the intended target in cells by examining enhancement of the thermal stability of the targeted protein. We employed a CETSA experiment to evaluate the effect of compound **30** (TC-6304) on the thermal stability of both DCN3 and DCN1 proteins in cells (Figure 3.7).



**Figure 3.7 The cellular thermal shift assay.** Examination of the thermal stability of DCN3 and DCN1 proteins in cells treated with a selective DCN3 inhibitor TC-6304, its enantiomeric control compound TC-6304DD and our previously reported

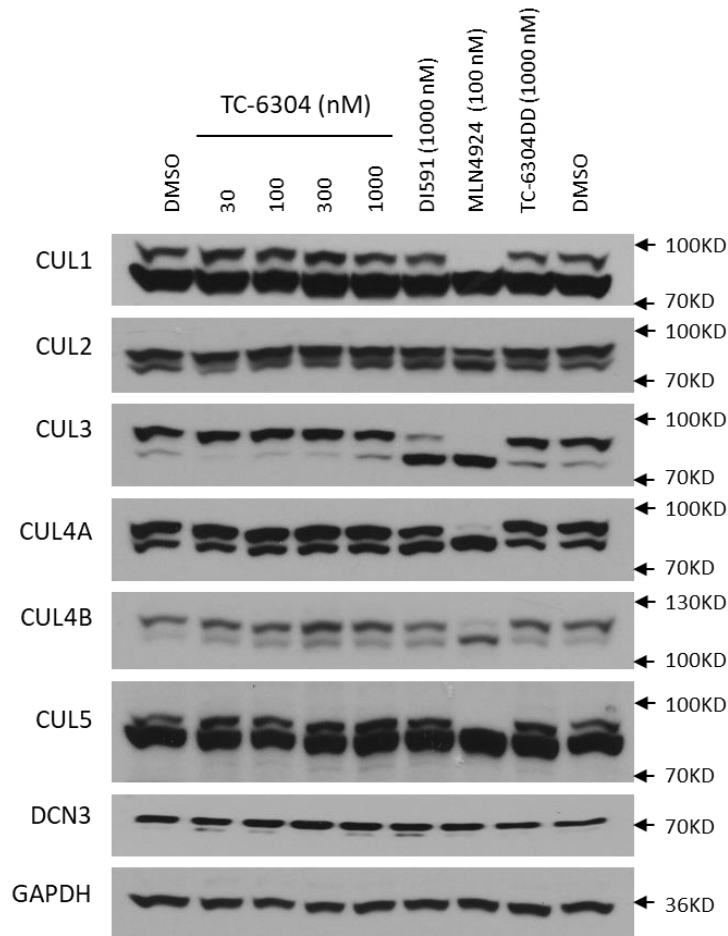
selective DCN1 inhibitor DI-591. A549 cells were treated for individual compound at indicated concentrations for 1 h. The protein level of DCN3 or DCN1 was examined by western blotting analysis, with GAPDH used as the loading control.

Our results show that TC-6304 effectively enhances the thermal stability of cellular DCN3 protein in a dose-dependent manner and has no significant effect on the thermal stability of cellular DCN1 protein. In contrast, DI-591, our previously reported selective DCN1 inhibitor effectively stabilizes cellular DCN1 protein at 1  $\mu$ M but has no effect on the thermal stability of DCN3 protein. The CETSA experiment provides direct evidence that TC-6304 interacts with DCN3 in cells.

### **3.2.10 Evaluation of TC-6304 for its effect on neddylation of cullin E3 ligases in cells**

In our previous study,<sup>7</sup> we demonstrated that DI-591, a DCN1 inhibitor selectively inhibits neddylation of cullin 3 and has a minimal effect on the neddylation of other cullin members. We evaluated TC-6304 for its effect on neddylation of different cullin E3 ligases in cells, obtaining the results shown in **Figure 3.8**.

As expected, MLN4924, an E1 inhibitor, effectively inhibits neddylation of all cullins. Consistent with our previous data, DI-591 effectively and selectively inhibits neddylation of cullin 3. However, TC-6304 has no effect on the neddylation of all cullins at all concentrations tested. Although more studies are clearly needed to further investigate the biological consequence of selective binding to DCN3 by TC-6304, our data clearly show that DCN3 and DCN1 proteins have very different roles in regulation of neddylation activity of cullin E3 ligases.

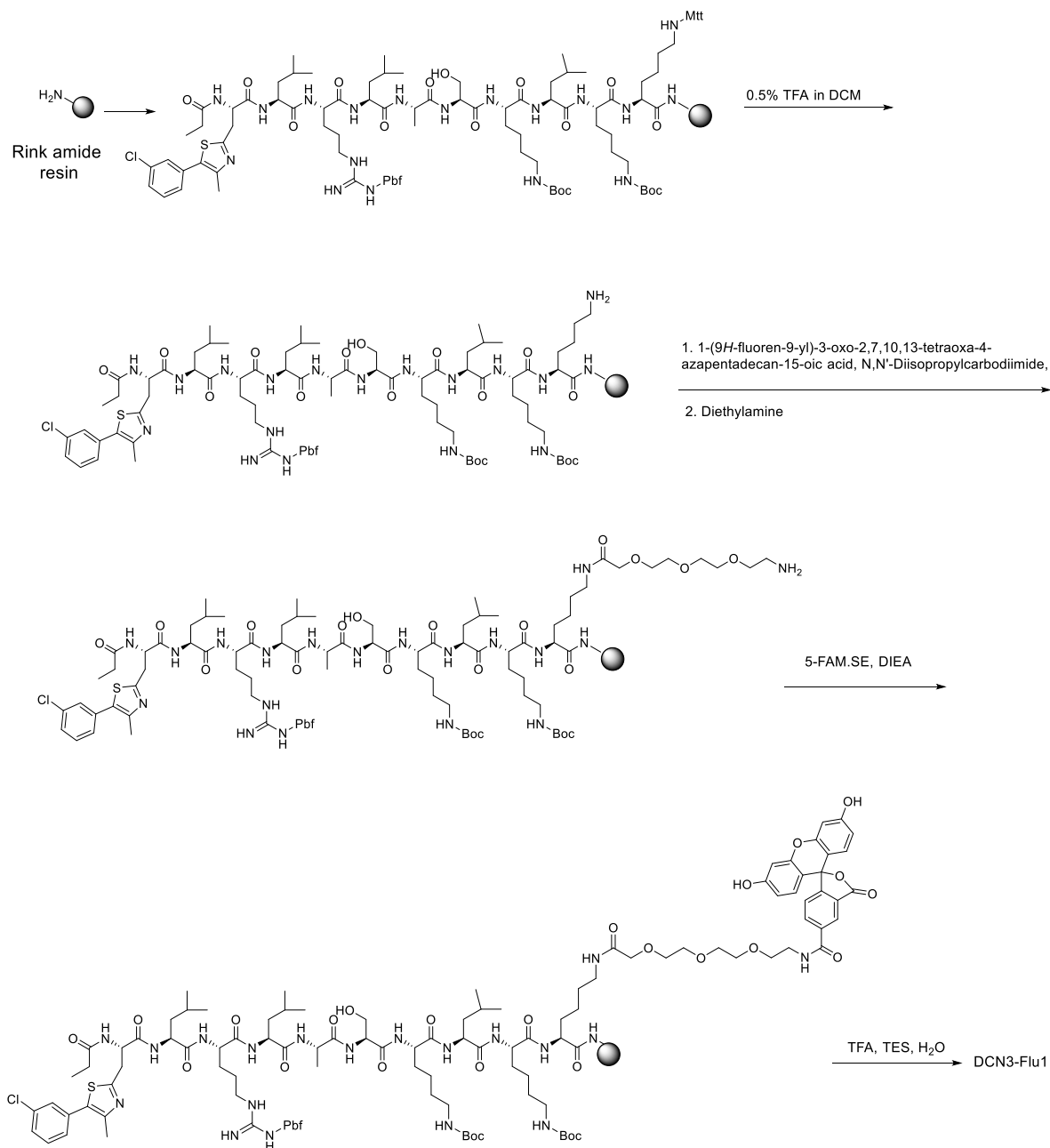


**Figure 3.8 . Evaluation of the effect of a selective DCN3 inhibitor TG-6304 on neddylation of different cullins in cell.** Lung cancer A549 cells were treated with TG-6304, DCN1 inhibitor DI591, or an inactive control TG-6304DD and an E1 inhibitor MLN4924 for 24 h. Protein levels of neddylated (active) and un-neddylated (inactive) different cullin proteins, as well as DCN3 and CDT1, were examined by western blotting analysis with GAPDH used as the loading control.

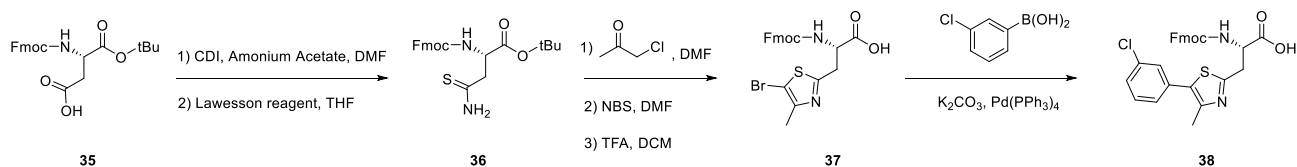
### 3.3 Chemistry

As shown in Scheme 1, a fluorescent labeled tracer (DCN3-Flu1) was produced by solid-phase peptide synthesis. 5-Carboxyfluorescein succinimide ester (5-FAM, SE) was attached to the side chain of the Lys of the linear peptide *via* a linker.<sup>8</sup>

**Scheme 3.1 Synthesis of tracer DCN3-Flu1.**

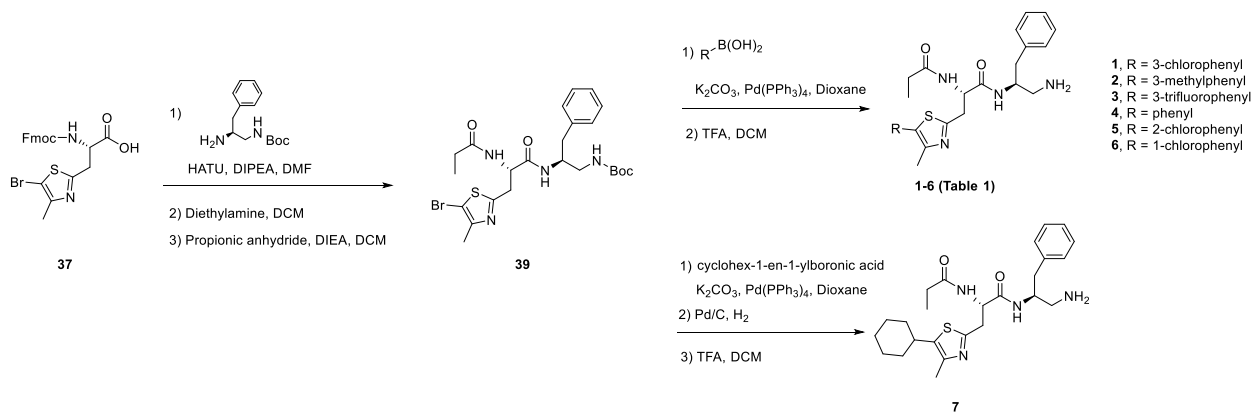


### Scheme 3.2 General Synthetic Route for Amino Acid 38



The unnatural amino acid (38) was synthesized as shown in Scheme 2. The acid group of Fmoc-Asp-O-t-Bu (35) was converted to an acetamide then transformed to a thioamide using Lawesson's reagent, giving compound 36. Cyclization of methylthiazole by the reaction of chloroacetone with 36 followed by bromination using NBS, and deprotection of the BOC by TFA, provided 37. Suzuki coupling of 37 with 3-chlorophenylboronic acid led to compound 38.

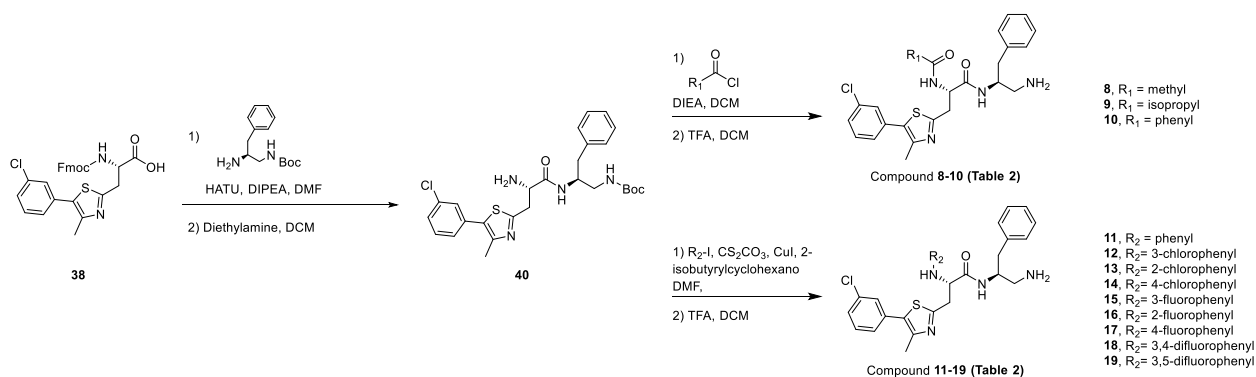
### Scheme 3.3 Synthesis of compounds 1-7.



The general method for the synthesis of compounds 1-7 is shown in Scheme 3. Compound 39 was produced by coupling 37 with *tert*-butyl (2-amino-3-phenylpropyl) carbamate using HATU as the coupling reagent, and was followed by removal of the Fmoc protecting group by treatment with diethylamine to allow the installation of a propionyl

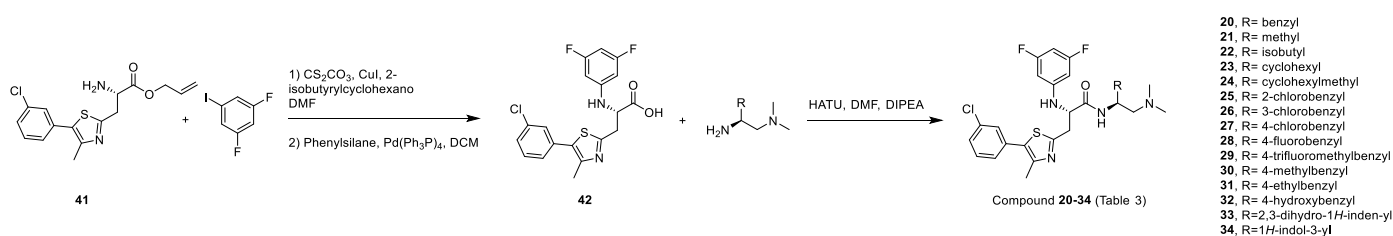
group to the amine. Suzuki coupling of **39** with various R-substituted boronic acid gives compounds **1-7**.

### Scheme 3.4 Synthesis of compounds 8-19.



The general method for the synthesis of compounds **8-19** is shown in Scheme 4. The common intermediate (**40**) was produced by coupling **38** to *tert*-butyl (2-amino-3-phenylpropyl)carbamate with HATU as coupling reagent, and this was followed by removal of the Fmoc protecting group by treatment with diethylamine. The primary amine (**40**) was reacted with acyl chloride to produce compounds **8-10**. The Ullman-type N-arylation coupling<sup>84</sup> was carried out to install various aromatic rings on the N-terminus of **40**, giving compounds **11-19**.

### Scheme 3.5 Synthesis of compounds 20-34.



The general method for the synthesis of compounds **8-19** is shown in Scheme 5. Compound **41** was synthesized using the method described in Scheme 2. N-arylation of **41** was performed using the Ullman coupling reaction with 1,3-difluoro-5-iodobenzene to give compound **42**. Compounds **20-34** were synthesized by coupling of **42** with different R-substituted amines (Scheme 5) with HATU as coupling agent.

### 3.4 SUMMARY

In this study, we report our design, synthesis and evaluation of non-peptide, small molecule ligands targeting the protein-protein interactions of DCN3 with UBC2E and UBC12, both of which are DCN3 binding partners in the cullin E3 ligase complexes. By screening our in-house chemical library of small-molecule DCN1 inhibitors, we identified three initial lead compounds, which belong to the same chemical class. Our subsequent structure-activity relationship studies, guided by structure-based design, have yielded a series of potent and highly selective DCN3 inhibitors, exemplified by TC-6304, which binds to DCN3 with a  $K_i = 35$  nM and shows minimal binding to DCN1 up to 30  $\mu$ M. Our CETSA data provides direct evidence that TC-6304 is cell-permeable, interacts with DCN3 protein in cells, and is capable of enhancing the thermal stability of DCN3 at low nanomolar concentrations. Furthermore, TC-6304 has no effect on the thermal stability of cellular DCN1 protein, whereas our previously reported selective DCN1 inhibitor DI-591 effectively enhances the thermal stability of DCN1 protein but not of the DCN3 protein. While DI-591 effectively and selectively inhibits the neddylation of cullin 3, TC-6304 has no obvious effect on neddylation of any of the cullin members examined, indicating that DCN3 and DCN1 play very different roles in regulation of the neddylation of cullin proteins

and the activation of cullin E3 ligase complexes. This study has yielded first-in-class, potent, selective and cell-permeable DCN3 inhibitors with TC-6304 being the best example. Further studies are ongoing to further investigate the biological consequences of selective binding to DCN3 and blocking the interactions with its binding partners using TC-6304 and the results will be reported in due course.



## Chapter 4 Discovery of Potent Small-Molecule Inhibitors of Mixed-Lineage Leukemia (MLL) Methyltransferase

### 4.1 Introduction

Rearrangements of the mixed-lineage leukemia (MLL) gene, also known as MLL1, have been detected in >70% of acute lymphoblastic leukemias (ALL) in infants and 5-10% of acute myelogenous leukemias (AML) in adults.<sup>21, 22</sup> AML patients carrying the MLL fusion gene have MLL leukemia, expect a very poor prognosis and are unresponsive to current treatments. There is therefore an urgent need to develop new therapeutic strategies for MLL leukemia.

MLL protein is a lysine methyltransferase, responsible for Histone 3 Lysine 4 (H3K4) methylation. MLL translocations invariably occur on only one MLL allele. Because the MLL fusion proteins contain no SET [Su(var)3-9, Enhancer-of-zeste and Triithorax] domain, which is the catalytic unit of MLL, MLL translocations lead to a loss of the H3K4 methyltransferase activity, but the remaining wild-type MLL allele retains the H3K4 methyltransferase activity. It has been shown that both the wild-type and fusion MLL proteins are required for MLL-AF9-induced leukemogenesis and maintenance of MLL-AF9-transformed cells. Therefore, inhibition of the H3K4 methyltransferase activity of MLL has been proposed as an attractive therapeutic strategy for the treatment of MLL leukemia.

Although MLL protein contains a catalytic SET domain, the protein itself has very low histone methyltransferase (HMT) activity.<sup>25</sup> When MLL protein forms a complex with WDR5, RbBP5 and ASH2L, its HMT activity is dramatically enhanced.<sup>25</sup> Therefore, targeting the protein-protein interactions within the MLL complex, in particular the WRD5-MLL protein-protein interaction, has been pursued as a strategy to inhibit the MLL HMT activity. In the last few years, we have reported the discovery of highly potent and cell-permeable peptidomimetics of the WDR5-MLL protein-protein interaction.<sup>27</sup> Potent, non-peptidic, small-molecule inhibitors of the WDR5-MLL interaction have also been reported.<sup>28, 29</sup>

Because MLL is a methyltransferase, an alternative therapeutic strategy is to develop small-molecule inhibitors that bind to its active site and directly block its HMT enzymatic activity. This strategy has been successfully employed for the design of direct inhibitors for a number of histone methyltransferases including MMSET,<sup>48</sup> SET7/9,<sup>46, 47</sup> SMYD3,<sup>49</sup> SETD2,<sup>50</sup> EZH2,<sup>51</sup> and DOT1L<sup>52, 53</sup> but, to date, no direct MLL inhibitor has been reported.

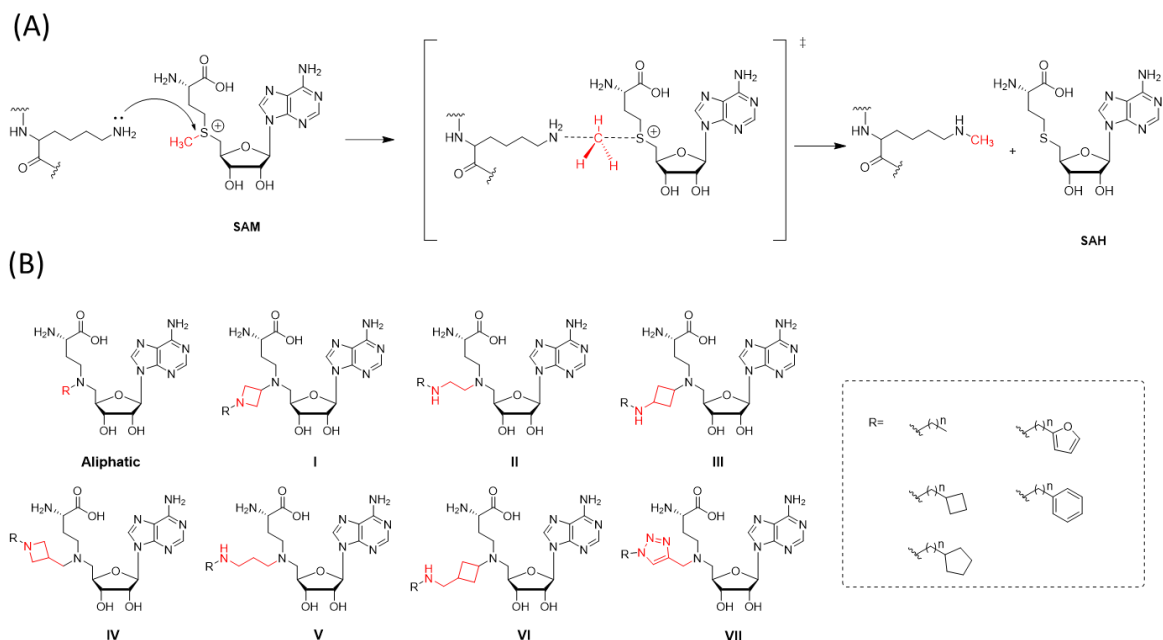
Like other methyltransferases, MLL employs S-adenosylmethionine (SAM or AdoMet) as a major methyl donor which engineers the transfer of a methyl group to histone lysine residues.<sup>33, 85</sup> In the present study, we describe the discovery of first-in-class, potent, direct MLL inhibitors, through the design, synthesis and evaluation of a SAM-based focused chemical library.

## 4.2 Results and Discussion

### 4.2.1 Design of a Focused Chemical Library Based upon SAM

Our strategy was to synthesize small molecules that mimic the transition state of the methylation chemical reaction of the lysine substrate by incorporation of the cofactor SAM with a portion of the lysine substrate as illustrated in Figure 4.1 (A).

In our design, the adenosine and the 2-aminobutanoic acid moieties were retained because they are structural moieties necessary for binding to the cofactor binding site. The positively charged sulfur atom in the transition state structure was replaced with a nitrogen atom. To identify optimal moieties that mimic the transition state structure, we systematically explored various primary, secondary or tertiary amines, as well the linker length between the two nitrogen atoms in the potential inhibitor. We further attached either an aliphatic or an aromatic hydrophobic fragment to the nitrogen to explore the chemical space around the lysine tunnel that is known to be responsible for the selectivity between different methyltransferases. This focused chemical library was categorized, based upon the nature of the group attached to the nitrogen atom mimicking the charged sulfur atom, into the following eight Groups: pure aliphatic, cyclobutene amine, cyclobutene methylene amine, triazole, azetidine, ethylamine, propylamine and methylene azetidine (Figure 4.1 (B)).

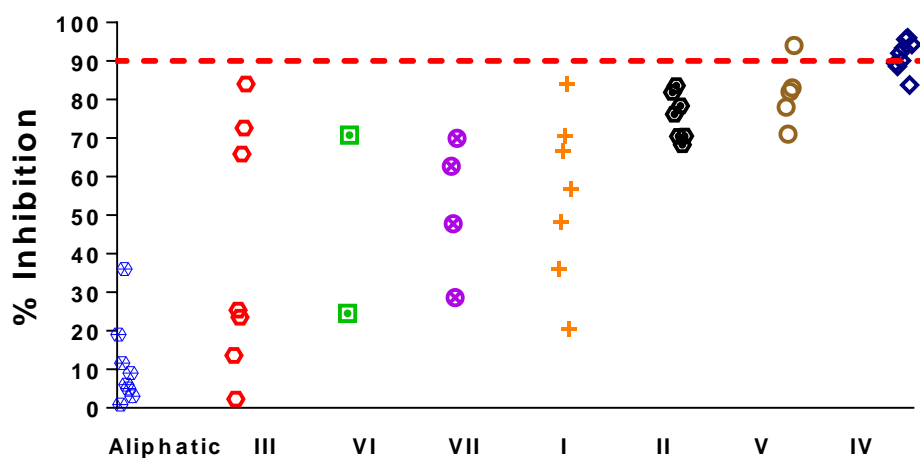


**Figure 4.1 Mechanism of lysine methylation and the Designed focused library.** Design of a focused SAM-based library. (A) Proposed mechanism of lysine methylation catalyzed by SAM-dependent lysine methyltransferases. (B) Designed SAM analog library based on the proposed transition state structure. The library contains 8 Groups: aliphatic, I (azetidene), II (ethylamine), III (cyclobutane amine), IV (methylene azetidene), V (propyl amine), VI (cyclobutane methylene amine) and VII (triazole). The R functional groups include either an aliphatic or an aromatic hydrophobic fragment.

#### 4.2.2 Identification of MLL Inhibitors by Screening the Focused Chemical Library

We screened this focused compound library for their ability to inhibit MLL methyltransferase at a single concentration of 10  $\mu$ M and obtained the results shown in Figure 2. This screening led to the identification of 5 compounds, each of which achieved >90% inhibition of the MLL HMT activity. Additionally, 5 other compounds were found to achieve 70-80% inhibition of the MLL HMT activity.

Among the compounds tested, 4 out of the 5 compounds containing a methylene azetidine (Group IV) inhibit the MLL activity by >90%. In contrast, all the compounds in the aliphatic subset exhibit less than 40% inhibition.



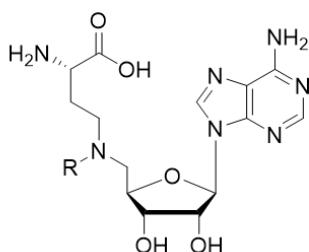
**Figure 4.2 Result of Single Dose Screening against MLL1 Methyltransferase.** Categorical scatterplot representation of the results of focused library screened against MLL methyltransferase using an HMT assay at a single concentration of 10  $\mu$ M. Red-dashed line represents 90% inhibition cutoff.

### 4.2.3 Validation of the Screening Hits

To validate the screening hits, we determined the  $IC_{50}$  values for all compounds which achieved >85% inhibition in the screening using the Alpha LISA assay that we have developed for MLL.<sup>86</sup> For Groups VI and VII, in which no single compound inhibited the MLL activity by >85%, we selected a representative compound with the highest inhibition and evaluated it in the MLL Alpha LISA assay. We included SAH (S-Adenosyl-L-homocysteine), a known pan-HMT inhibitor, as a positive control.<sup>34</sup> The results of these validation experiments are summarized in Table 4.1.

In the MLL Alpha LISA assay, SAH has an IC<sub>50</sub> value of 724 nM. The most potent inhibitors in Groups I-II are 2-3 times more potent than SAH. The most potent inhibitor in Group III has an IC<sub>50</sub> of 127 nM and is thus 5-times more than potent than SAH. Four compounds in Group IV, including compound **5**, achieve the highest potencies with IC<sub>50</sub> values of 21-72 nM, and are >10-times more potent than SAH. The most potent compound in Group V has an IC<sub>50</sub> value of 201 nM and is 3-times more potent than SAH. The most potent compound in Group VI has an IC<sub>50</sub> value of 415 nM, similar to that of SAH. The most potent compound in Group VII has an IC<sub>50</sub> value of 132 nM and is thus 5-times more potent than SAH.

**Table 4.1 Results of hit validation using Alpha LISA assay for MLL1 methyltransferase.**



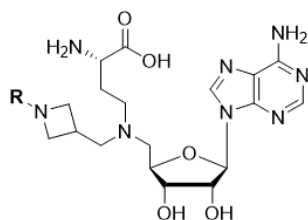
ID	Sub-set	R	IC 50 (nM)
TC-3241	I		187.00 ± 52.33
TC-3116	II		504.00 ± 147.08
TC-3246	III		136.00 ± 26.87
TC-3233	IV		77.50 ± 2.12
TC-3235	IV		27.50 ± 10.61
TC-3236	IV		28
TC-3232	IV		25.67 ± 6.66
TC-3263	V		224
TC-3261	VI		497.50 ± 24.75
TC-3209	VII		140.00 ± 36.77
SAH			724.00 ± 96.17

#### 4.2.4 Further Modifications of the Most Potent Compound 5

Compound **5**, with an  $IC_{50}$  value of 21.8 nM is the most potent MLL inhibitor identified in the focused chemical library. We performed further modifications of the azetidine in **5** to gain a further understanding of the SAR for this site with the results summarized in Table 4.2.

Substitution of the azetidine NH in **5** with a benzyl group resulted in compound **11**, which has an  $IC_{50}$  value of 47 nM and is thus 2-times less potent than **5**. Substitution of a chlorine atom onto each of three different positions of the phenyl ring in **11** led to a modest decrease in inhibitory potency (2-3 times). Replacing the phenyl group in **11** with 2-thiophenyl or 3-thiophenyl yielded **16** and **15**, respectively. While **16** has an  $IC_{50}$  value of 67 nM, which is 3-times less potent than **5**, compound **15** is equipotent with **5**.

**Table 4.2. SAR of subset IV compounds.**

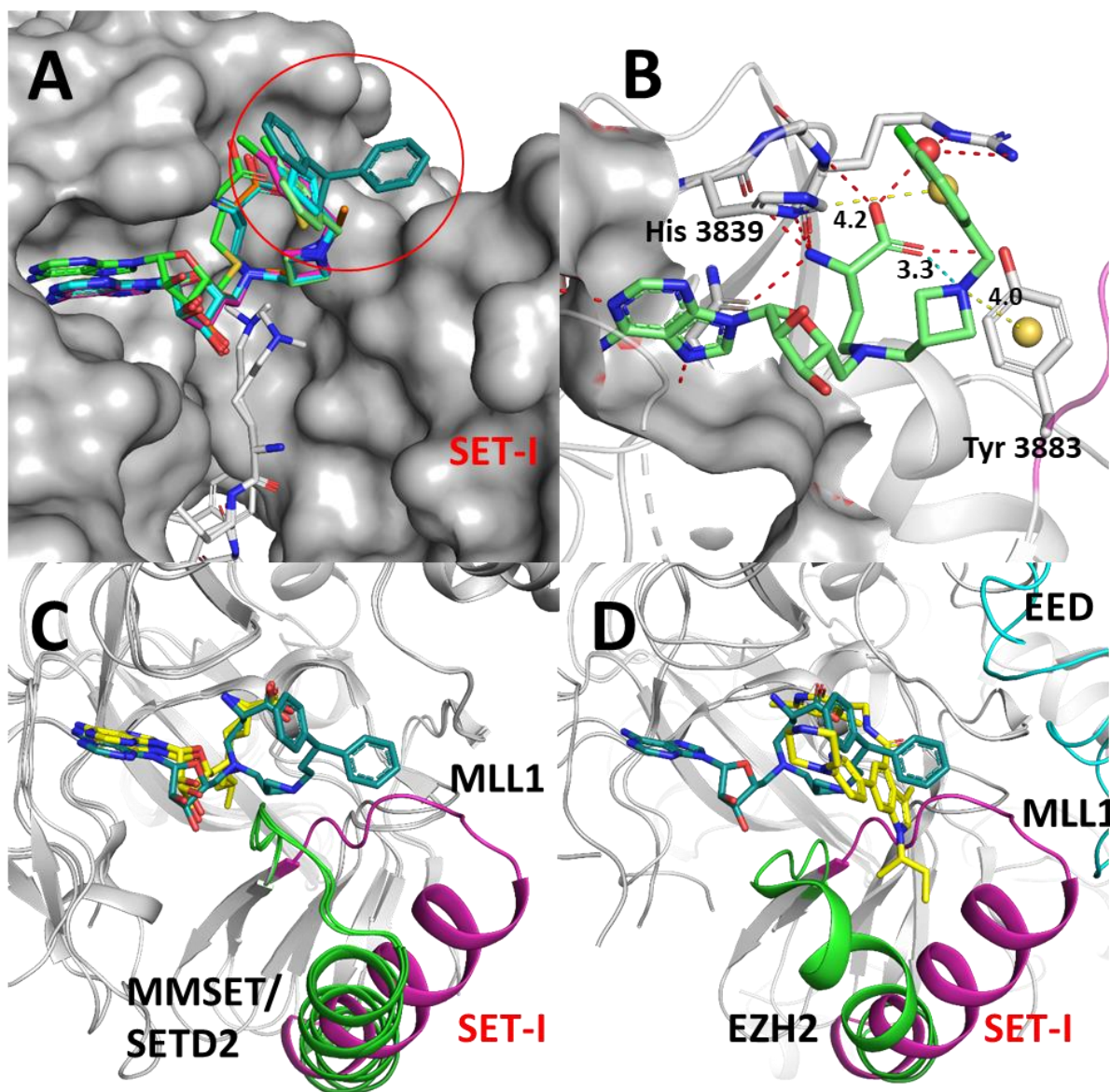


ID	R	$IC_{50}$ (nM)	ID	R	$IC_{50}$ (nM)
TC-5102		$46.8 \pm 3.9$	TC-5115		$15.2 \pm 0.8$
TC-5140		$99.5 \pm 8.4$	TC-5105		$77.3 \pm 5.7$
TC-5141		$150.2 \pm 11.5$	TC-5108		$119.0 \pm 18.4$
TC-5139		$55.6 \pm 2.1$	TC-5153		$138.4 \pm 11.3$



#### 4.2.5 Determination of Co-Crystal Structures of MLL Inhibitors in a complex with MLL1 SET Domain

To understand the structural basis for their potent inhibition of MLL activity, we determined co-crystal structures for 4 inhibitors: compound **16** (TC-5109), **14** (TC-5139), **12** (TC-5140) and **18** (TC-5153) in a complex with the MLL SET domain.



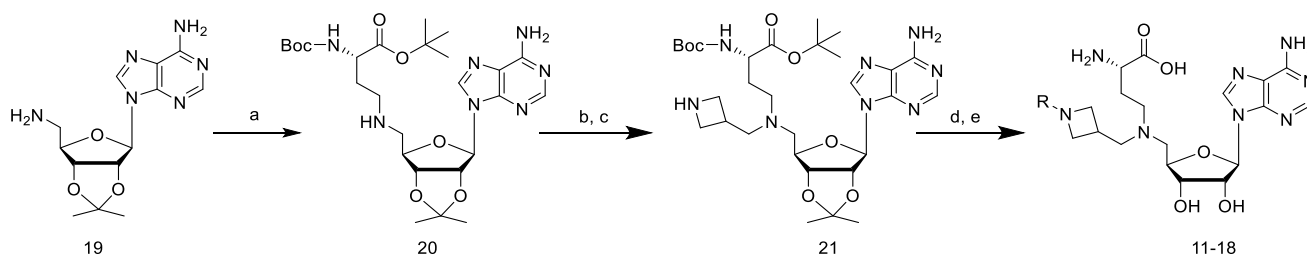
**Figure 4.3 Crystal structures of inhibitor and MLL1 methyltransferase.** (A) Crystal complex of N-azetidine SAM analogs with MLL SET domain, SAH and the H3K4 substrate (PDB ID: 2W5Z). (B) Detailed view of compound TC-5140 in a complex with MLL. Yellow spheres are the calculated coordinate of the center of mass of the phenyl ring. The red dashed lines show the conserved hydrogen bond network. The yellow dashed line is cation- $\pi$  interaction. The purple line is an edge-face  $\pi$ - $\pi$  interaction. The cyan dashed line shows intramolecular charged-charged interaction. (C) Superimposed structure of MLL, compound **18** (TC-5153) complex and MMSET/SETD2 and N-alkyl Sinefungin complex (PDB ID: 5LSY), SET-I domain colored purple and green in MLL and MMSET/SETD2, respectively. (D) Superimposed structure of MLL, TC-5153 complex and EZH2 and GSK126 complex (PDB ID: 5WG6), SET-I domain colored purple and green in MLL and EZH2, respectively.

These co-crystal structures showed that the binding modes of adenosine and the  $\alpha$ -amino-acid moieties of these 4 inhibitors align closely to the same moieties of SAH in its co-crystal structure with the MLL SET domain (PDB: 2W5Z). Unexpectedly, the methylene azetidine moiety in these inhibitors binds not to the lysine tunnel but is directed towards the solvent exposed site (Figure 3 A). The methylene azetidine moiety also forms an intramolecular salt bridge with the  $\alpha$ -amino acid of the inhibitors and engages in a cation-aromatic interaction with Tyr3883 as shown in the MLL/TC-5140 structure (Figure 3B). We also observed an edge-to-face  $\pi$ - $\pi$  interaction between the chlorophenyl of TC-5140 and His3839 at a distance of 4.2 Å. The aromatic-His3839 interaction was also found in the co-crystal structures of MLL in complex with TC-5109 and TC5139 containing a thiophene and a diphenyl group, respectively (Figure 3A). These crystal structures suggest that the azetidine group in our inhibitors provides the conformational orientation to permit additional interactions between the tail groups of the inhibitors and the MLL SET domain.

Previous structural studies indicated that MLL has a more spacious active site and a more open lysine binding groove than other histone methyltransferases.<sup>87, 88</sup> Superimposition of our MLL SET domain structures with three closely aligned lysine methyltransferases (MMSET, SETDB2, EZH2) shows that the orientation of SET-I of MLL is the only significant variation as it moves away from the C-flanking region<sup>88</sup> (Figures 3 C, 3 D). The alignment further indicates that the loop preceding SET-I in MMSET, SETDB2 and EZH2 would sterically clash with the *N*-azetidine moiety in our inhibitors. Previous studies also suggest that the SET-I must be reoriented to adopt a “closed” conformation in order to exert full catalytic activity on SAH.<sup>88, 89</sup> Our co-crystal structures therefore suggest that our designed MLL inhibitors may lock the MLL SET domain into an “open” and consequently inactive conformation. Taken together, our co-crystal structures provide a structural basis for the high affinity interactions between these MLL inhibitors with MLL and for further structure-based optimization.

### 4.3 Chemistry

**Scheme 4.1 Synthesis of compounds 11-18**



(a) *tert*-butyl (S)-2-((*tert*-butoxycarbonyl)amino)-4-oxobutanoate, NaHB(OAc)<sub>3</sub>, DCE, rt, 4 h; (b) benzyl 3-formylazetidine-1-carboxylate, NaHB(OAc)<sub>3</sub>, DCE, rt, 3 h; (c) H<sub>2</sub>, Pd/C, MeOH, rt, 1 h; (d) RCHO, NaHB(OAc)<sub>3</sub>, DCE, rt, 2 h; (e) TFA, DCM, rt, 1 h.

The synthesis of compounds **11-18** containing a substituted N-azetidine series is illustrated in Scheme 1. Reductive amination of compound **19** with *tert*-butyl (S)-2-((*tert*-butoxycarbonyl)-amino)-4-oxobutanoate produced compound **20**. Compound **21** was prepared by reductive amination of compound **20** with benzyl 3-formylazetidine-1-carboxylate, NaHB(OAc)<sub>3</sub> and DCE followed by removal of the protecting group by hydrogenation. Reductive amination of various aldehydes (RCHO) with compound **21** and the subsequent deprotection provided the final compounds **11-18**.

#### 4.4 Summary

In summary, we designed a library focused on SAM-analogs using a bisubstrate strategy intended to profile different methyltransferases. We use this chemical library to screen for inhibitors of MLL methyltransferase and found several compounds with potent inhibitory activity. Among them, compound **16** (TC-5115) exhibited an IC<sub>50</sub> value of 16 nM. Co-crystal structures of MLL in complex with 4 potent inhibitors were determined, which revealed that these inhibitors may lock the MLL SET domain in an open, inactive conformation.

## Chapter 5 Conclusion and Future directions

### 5.1 Conclusion

The goal of my work presented in this dissertation was to fill the unmet need for compounds that can be used to interrogate the biology of the Cullin-RING ubiquitin ligases (CRLs) and the MLL1 methyltransferase. We conducted structure-based design, organic synthesis, and biochemical/biophysical assays to discover potent inhibitors targeting DCN1, DCN3 and, MLL1 methyltransferase.

In **Chapter 2** of this dissertation, we have described the importance of the ubiquitin-proteasome system for maintaining the homeostasis of the cell and its relevance to disease. CRLs are the largest family of the E3 ligase, and its activation requires neddylation on the cullin protein. After thoroughly investigating the structural topology and the interface among all components of the Cullin1-CRL complex, we hypothesized that we could achieve individual CRL inhibition by targeting the DCN1-UBC12 protein-protein interaction. We performed systematic truncation on the N-terminus of UBC12 and identified a tetrapeptide lead compound. The co-crystal structure of this tetrapeptide with DCN1 revealed identical binding mode as the DCN1-UBC12 complex. We next optimized its interaction with P1-P4 hydrophobic pocket on DCN1, as well as the physical-chemical properties, yielded **DI-404** with  $K_i$  value of 6.7 nM. The cellular thermal shift assay demonstrated that **DI-404** engage DCN1 in cells. In western blot experiments, we

observed that **DI-404** selectively inhibit the neddylation of cullin 3. Also, a well-known substrate of cullin 3 CRL, NRF2 protein, was accumulated. We envision this tool compound could be used to investigate the biology of cullin 3 CRL. Further optimization based on this compound may ultimately yield new therapeutics for diseases related to cullin 3 CRL.

The proof-of-concept study in **Chapter 2** showed that selective cullin 3 CRL inhibition can be achieved by targeting protein-protein interaction within the CRL complex, specifically the interface of DCN1/UBC12. Encouraged by this finding, we reasoned that targeting other DCN family protein / E2 enzyme protein-protein interaction might inhibit other cullin family CRL functions. Through the literature survey, our analysis suggested that targeting DCN3 and UBE2F protein-protein interaction could be an alternative means for manipulating other CRLs. In the DCN1 project, we have made a significant number of small molecules during the iteration of the optimization process. Given that DCN1 and DCN3 shared similar binding sites, it is possible to find moderately active compounds that bind DCN3. Therefore, in **Chapter 3** of the dissertation, we described reversing the selectivity profile from DCN1 to DCN3 by structure-based design.

Since this DCN1-inhibitor library is highly tailored for DCN1, we only found three lead compounds with  $IC_{50}$  value less than 10  $\mu$ M against DCN3 in the label-free, biolayer interferometry assay. Based on the chemical scaffold of the lead compounds, and our truncation analysis of UBE2F peptide, we designed and synthesized a fluorescently tagged peptide tracer **DCN3-Flu1** ( $K_d = 138$  nM) for fluorescence polarization assay development. Through extensive optimization for three hydrophobic pockets of DCN3, we discovered **TC-6304** with a  $K_i$  value of 35 nM showing no binding up to 30  $\mu$ M. In the

cellular thermal shift assay, we did observe a dose-dependent engagement of **TC-6304** with DCN3 in the cell. However, we did not find any significant change in cullin-neddylation in all cullins tested. To our knowledge, our effort in **Chapter 3** is considered to be the first documented report of potent DCN3 inhibitor.

**Chapter 4** is a collaborative project with the laboratory of Dr. Yali Dou. We aimed to develop tool compounds to address the question about the general function of H3K4 methylation regulated by *direct targeting* the MLL SET domain. We first designed a focused chemical library based upon SAM to extensively explore the chemical space in the lysine tunnel in the MLL1 SET domain active site. Among this focused library, we identified a group of compounds with methylene azetidine moiety exhibited >90% inhibitory activity against MLL1 in a single-dose preliminary screening. The hits are validated by MLL Alpha LISA assay showing this group of compounds did outperform other groups of inhibitors. Further modifications yielded compound **TC-5115** with IC<sub>50</sub> value of 15 nM for MLL1, which is 48 times more potent than the pan-methyltransferase inhibitor, SAH. Given that there is yet no reported potent MLL1 SET domain inhibitor, our compound is the first documented effort for direct targeting MLL1 SET domain.

Collaborating with the laboratory of Dr. Jeanne Stuckey, we obtained four cocrystal structures of our inhibitor with MLL1 SET domain. All inhibitors share the same binding mode with SAH, and a similar molecular interaction between SAH and MLL1 was observed. The additional interaction between MLL1 and inhibitors explained the gain of the potency of Group IV compounds. Interestingly, we also observed a unique binding mode which is not seen in literature reported SAM-based methyltransferase inhibitors. The moiety that was designed to target the substrate-binding site bent over to the solvent-

exposed site and interact with the intrinsically flexible SET-I domain (motif) of MLL1. These crystal structures provide the solid starting point for further optimization, and I will discuss in detail in the next section.

## 5.2 Future Directions

### 5.2.1 The DCN project

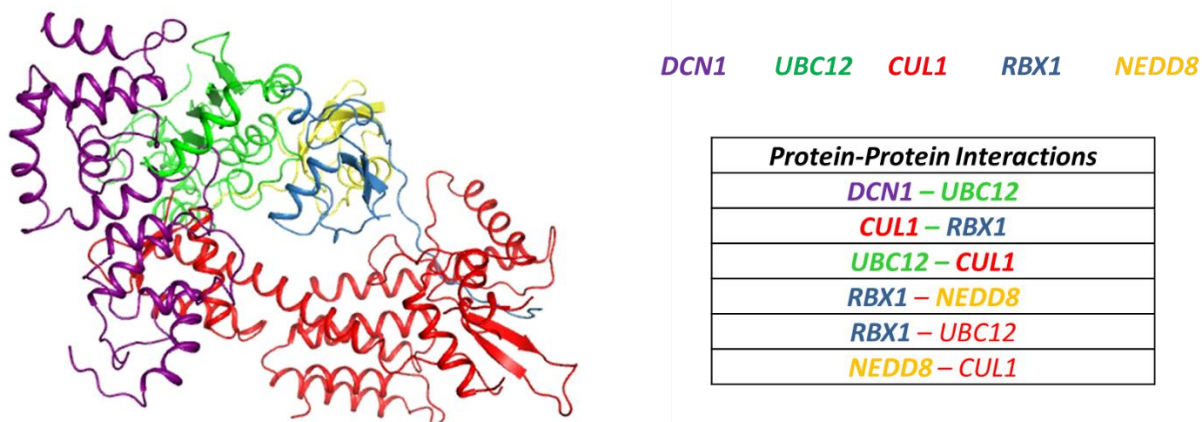
#### 5.2.1.1 *Screening against other proteins in the DCN family*

Our work described in **Chapter 2** demonstrates the proof-of-concept that we can manipulate the individual CRL activity through targeting the protein-protein interaction of DCN1/UBC12 within the multi-component CRL complexes. In **Chapter 3**, although our potent DCN3 inhibitor did not show any significant effect on cullin neddylation, we demonstrated that it is possible to design other member of DCN family proteins through a structure-based design approach. In the DCN1 and DCN3 projects together, our efforts in medicinal chemistry work accumulated >350 compounds. Because DCN family proteins are homologous proteins, we could still utilize this in-house DCN1/3 inhibitor library for screening purposes. One challenge for this direction was that no crystal structure is available for DCN2, DCN4 and DCN5 in complex with the corresponding E2 enzyme. Therefore, a well-refined homology model of those DCN proteins will be required to guide the optimization process.



### 5.2.1.2 Targeting other protein-protein integrations among different components in the CRL complex

The primary goal for the DCN projects is to achieve selective inhibition of CRL. In the analysis of the protein-protein interactions among components in the CUL1 CRL complex (RBX1~UBC12~NEDD8~CUL1~DCN1),<sup>9</sup> we found that there are other potential PPI sites that could support inhibitor design. These potential interaction sites include CUL1-RBX1, UBC12-CUL1, RBX1-NEDD8, RBX1-UBC12, and NEDD8-CUL1 (Figure 5.1). However, those protein-protein interactions are not equally “druggable”. Further computational prediction tools such as SiteMap (Schrodinger, New York, NY)<sup>90</sup> could be utilized to determine the best potential site for inhibitor design.



**Figure 5.1 Analysis of the protein-protein interactions in the CUL1 CRL complex**

## 5.2.2 The MLL1 methyltransferase project

### 5.2.2.1 Structural modifications of current MLL1 inhibitors

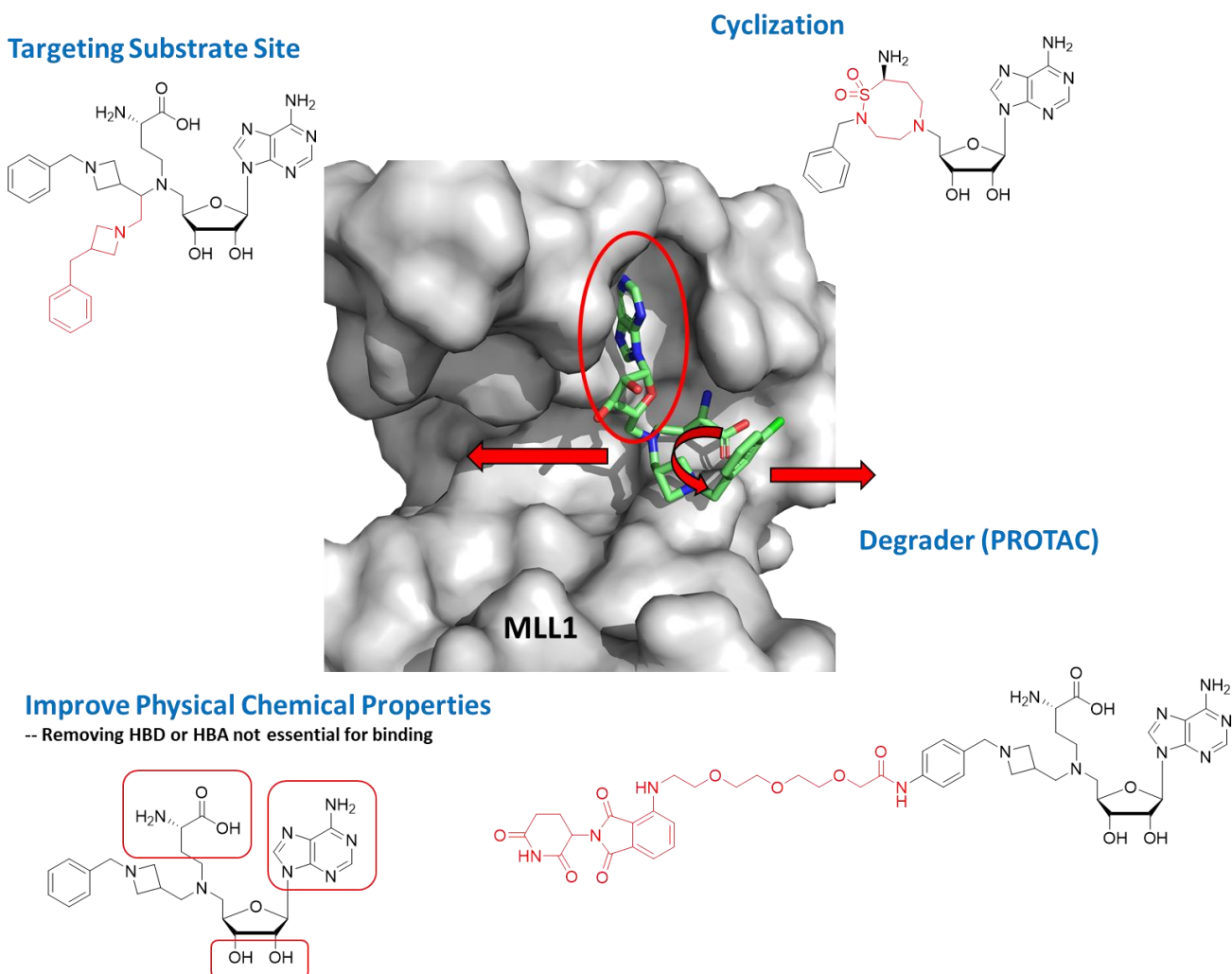


Figure 5.2 Further modifications of MLL1 inhibitors

#### 5.2.2.1.1 Improving the physical chemical properties

In our design of SAM analogs that target MLL1 methyltransferase, we applied the bisubstrate approach. We replaced the sulfur atom of SAH with nitrogen and attached the amino acid to this nitrogen. Several fragments were attached to the amine to explore the

chemical space of the MLL1 SET domain. Although our compound achieved 2-digit nanomolar inhibitory activity *in vitro*, activity in cells was not observed. From the literature, we found that it is common for SAM-based methyltransferase inhibitors to show potent inhibitory activity in biochemical assays but no cellular data is adduced. It has been suggested that adenine and the amino acid portion of SAM analogs are too polar. Interestingly, publications have shown that SAM itself can exert anticancer activity in various cancer cell lines.<sup>91, 92</sup> One study showed that SAM alters the methylation landscape and has a dose-dependent effect on cell growth and invasion in the SKhep1 cell line.<sup>93</sup>

A modification to improve the physical chemical properties is shown in Figure 5.2. The modification is focused on the amino acid, the adenine, and the 2'- or 3'-hydroxyl moieties. The goal for the modifications is to reduce the hydrophilicity of the inhibitors. Earlier studies of the modifications on the SAH itself have been summarized in Chapter 1. Briefly, most of the SAH modifications were either detrimental or resulted in weaker inhibitory activity. Removing the 2'-hydroxyl or 3'-hydroxyl moieties would be a first step, because MLL1 is one of the few HMTs that lacks hydrogen bond interactions. Replacement of adenine could be achieved by scaffold-hopping with computational simulation. Finally, the amino acid moiety would be very challenging because of its hydrogen bond network between HMTs.

#### **5.2.2.1.2 Targeting the lysine tunnel and the substrate binding site**

As was discussed in **Chapter 4**, we found from co-crystal structures, that our inhibitors do not occupy the lysine tunnel and the substrate binding site as molecular docking studies suggested. Instead, the azetidine moiety and the hydrophobic tail are

directed towards the solvent exposed site. In addition, all four crystal structures share the same binding mode, suggesting that current inhibitors could serve as an anchor with which to extend to the substrate binding site. A proposed compound is illustrated in Figure 5.2. The main purpose for this type of modification is to achieve the selectivity by optimizing the substrate binding pocket for MLL1.

#### **5.2.2.1.3 Design of MLL1 degraders (PROTAC)**

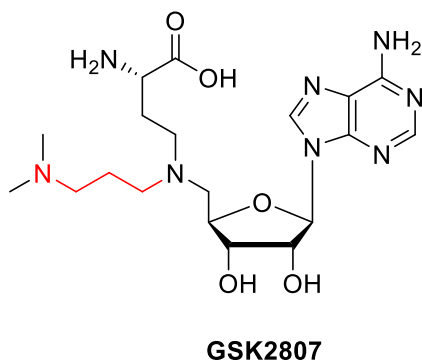
The PROTAC technology is an emerging approach to both therapeutic intervention and biological studies.<sup>94</sup> PROTAC molecules are composed of a target protein inhibitor, a E3 binding ligand, and a linker between the two binders. PROTACs will bind simultaneously with the E3 ubiquitin ligase and the target protein, resulted in promotion of the ubiquitination and subsequent degradation of the target protein.<sup>94</sup> To design a PROTAC, it is therefore necessary to determine which part of the inhibitor is directed towards the solvent exposed side of the protein. From the crystal structure, it is clear that an E3 ligase ligand and an appropriate linker could be attached the N-azetidine moiety as shown in Figure 5.2. One benefit of designing PROTAC structures is that we could attach a linker and an E3 ligase ligand to balance the hydrophilicity of SAM analog-based inhibitors. It is noted that MLL1 degraders will affect not only the catalytic function of SET domain. Other domains such as bromodomain, PHD domain might also be affected.

#### **5.2.2.1.4 Conformation restriction**

In all the crystal structures examined, we observed an intramolecular charge-charge interaction between the carboxylic acid and the N-azetidine moiety. By covalently linking these two moieties while maintaining the bound conformation, lower entropy could be achieved and this could result in lower binding free energy.

### 5.2.2.2 Screening of current SAM-based library for other methyltransferases

In this study, we discovered that the methylene azetidone (group IV) among the SAM-based library has binds preferentially to MLL1. In our effort to develop an MLL1 methyltransferase inhibitor, we have prepared >150 SAM analogs in this library. We can utilize this in-house library to screen against other methyltransferases of interest. In fact, after three years since we initiated the MLL1 project, a research group in GlaxoSmithKline discovered a novel SMYD3 inhibitor **GSK2807** with an  $IC_{50}$  value of  $0.13 \mu M$ .<sup>54</sup> **GSK2807** is selective for SMYD3 over other 9 tested histone methyltransferases and is also a SAM analog (Figure 5.3). We also made this compound in our SAM library in group V (propyl amine). This finding not only demonstrates that the bi-substrate design concept is a practical strategy for developing methyltransferase inhibitors but also suggests that SAM analog inhibitors can exhibit good selectivity for certain HMTs.



**Figure 5.3** An SMYD3 inhibitor, **GSK2807**

### 5.2.2.3 Test current MLL1 SET domain inhibitors against a panel of HMTs.

We currently have several potent MLL1 SET domain inhibitors, including **TC-5115** with the lowest  $IC_{50}$  value of 15 nM. In **Chapter 4** we discussed the co-crystal structure analysis which suggested that our compound could have selectivity over other HMTs. In

our preliminary study, one of the Group IV (methylene azetidine) compounds, **TC-3232** does not bind to MMSET up to 100  $\mu$ M. However, a larger panel of HMT profiling is still needed. It might be even more challenging to achieve selectivity within MLL-SET domain family. Potential ways to improve selectivity will be discussed in the following structural modification section.

## Appendices

## Appendix A: Experimental Section of Chapter 2

**Chemistry.** *General Information.* All reagents and solvents were used as supplied without further purification and the reactions were performed under a nitrogen atmosphere in anhydrous solvents. The final products were purified by a reverse phase HPLC (RP-HPLC) with solvent A (0.1% of TFA in H<sub>2</sub>O) and solvent B (0.1% of TFA in CH<sub>3</sub>CN) as eluents. The purity was determined by Waters ACQUITY UPLC and all the final compounds were > 95% pure. <sup>1</sup>HNMR and <sup>13</sup>CNMR spectra of the synthetic compounds were acquired at proton frequencies of 300 or 400 MHz and chemical shifts are reported in parts per million (ppm) relative to an internal standard. High resolution mass spectra (HRMS) were obtained with an Agilent Q-TOF Electrospray mass spectrometer and low resolution mass spectral analyses were performed on a Thermo-Scientific LCQ Fleet mass spectrometer.

**(S)-N<sup>1</sup>-((19S,22S)-26-amino-19-carbamoyl-1-(3',6'-dihydroxy-3-oxo-3H-spiro[isobenzofuran-1,9'-xanthen]-5-yl)-1,13,21-trioxo-5,8,11-trioxa-2,14,20-triazahexacosan-22-yl)-2-((2S,5S,8S,11S,14S,17S,20S,23S,26S)-2-(3-amino-3-oxopropyl)-5,20-bis(4-aminobutyl)-14-benzyl-23-((S)-sec-butyl)-11-(hydroxymethyl)-8,17-diisobutyl-26-(naphthalen-2-ylmethyl)-4,7,10,13,16,19,22,25,28-nonaoxo-3,6,9,12,15,18,21,24,27-**

**nonaazanonacosanamido)pentanediamide (4):** The immobilized peptide on Rink amide resin was produced on an ABI 433 peptide synthesizer. The 4-methyltrityl (Mtt)



protecting group on the Lys12 side chain was removed by treatment with 0.5% TFA in DCM (10 mL). The resulting resin was treated with 1-(9H-fluoren-9-yl)-3-oxo-2,7,10,13-tetraoxa-4-azapentadecan-15-oic acid and *N,N'*-Diisopropylcarbodiimide in DMF, and this was followed by removal of the Fmoc group with diethylamine. 5-Carboxyfluorescein succinimide ester (5-FAM, SE) was then added to a mixture of the washed resin and DIPEA in DMF and this mixture was shaken overnight. The peptide was cleaved from the resin by treatment with TFA:TES:H<sub>2</sub>O (36:1:2) cleavage cocktail and this was followed by HPLC purification, which yielded compound **4**. HRMS (ESI-MS) *m/z*: calculated for C<sub>108</sub>H<sub>153</sub>N<sub>20</sub>O<sub>26</sub><sup>+</sup> 2146.1259, found 2146.1207 [M+H]<sup>+</sup>.

**General method for the synthesis of peptides 5-13.** Peptides **5-13** were synthesized by solid-phase peptide synthesis using Fmoc chemistry shown in Scheme 2.

**(S)-6-Amino-N-((S)-1-amino-4-methyl-1-oxopentan-2-yl)-2-((2S,3S)-2-((S)-2-formamido-3-(naphthalen-2-yl)propanamido)-3-methylpentanamido)hexanamide (5):** HRMS (ESI-MS) *m/z*: calculated for C<sub>32</sub>H<sub>49</sub>N<sub>6</sub>O<sub>5</sub><sup>+</sup> 597.3759, found 597.3752 [M+H]<sup>+</sup>.

**(S)-6-Amino-2-((2S,3S)-2-((S)-2-amino-3-(naphthalen-2-yl)propanamido)-3-methylpentanamido)-N-((S)-1-amino-4-methyl-1-oxopentan-2-yl)hexanamide (6):** HRMS (ESI-MS) *m/z*: calculated for C<sub>31</sub>H<sub>49</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup> 569.3810, found 569.3804 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-((S)-1-amino-4-methyl-1-oxopentan-2-yl)-2-((2S,3S)-3-methyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)pentanamido)hexanamide (7):** HRMS (ESI-MS) *m/z*: calculated for C<sub>34</sub>H<sub>53</sub>N<sub>6</sub>O<sub>5</sub><sup>+</sup> 625.4072, found 625.4075 [M+H]<sup>+</sup>.

**N-((S)-1-(((2S,3S)-1-(((S)-6-Amino-1-(((S)-1-amino-4-methyl-1-oxopentan-2-yl)amino)-1-oxohexan-2-yl)amino)-3-methyl-1-oxopentan-2-yl)amino)-3-**

**(naphthalen-2-yl)-1-oxopropan-2-yl)cyclopropanecarboxamide (8):** HRMS (ESI-MS)

*m/z*: calculated for C<sub>35</sub>H<sub>53</sub>N<sub>6</sub>O<sub>5</sub><sup>+</sup> 637.4072, found 637.4074 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-((S)-1-amino-4-methyl-1-oxopentan-2-yl)-2-((2S,3S)-2-((S)-2-isobutyramido-3-(naphthalen-2-yl)propanamido)-3-**

**methylpentanamido)hexanamide (9):** HRMS (ESI-MS) *m/z*: calculated for C<sub>35</sub>H<sub>55</sub>N<sub>6</sub>O<sub>5</sub><sup>+</sup> 639.4228, found 639.4235 [M+H]<sup>+</sup>.

**((S)-6-Amino-N-((S)-1-amino-4-methyl-1-oxopentan-2-yl)-2-((2S,3S)-2-((S)-2-butyramido-3-(naphthalen-2-yl)propanamido)-3-methylpentanamido)hexanamide**

**(10):** HRMS (ESI-MS) *m/z*: calculated for C<sub>35</sub>H<sub>55</sub>N<sub>6</sub>O<sub>5</sub><sup>+</sup> 639.4228, found 639.4232 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-((S)-1-amino-4-methyl-1-oxopentan-2-yl)-2-((2S,3S)-3-methyl-2-((S)-2-(3-methylbutanamido)-3-(naphthalen-2-**

**yl)propanamido)pentanamido)hexanamide (11):** HRMS (ESI-MS) *m/z*: calculated for C<sub>36</sub>H<sub>57</sub>N<sub>6</sub>O<sub>5</sub><sup>+</sup> 653.4385, found 653.4389 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-((S)-1-amino-4-methyl-1-oxopentan-2-yl)-2-((2S,3S)-3-methyl-2-((S)-2-(3-methylureido)-3-(naphthalen-2-yl)propanamido)pentanamido)hexanamide**

**(12):** HRMS (ESI-MS) *m/z*: calculated for C<sub>33</sub>H<sub>52</sub>N<sub>7</sub>O<sub>5</sub><sup>+</sup> 626.4024, found 626.4024 [M+H]<sup>+</sup>.

**(Methyl ((S)-1-(((2S,3S)-1-(((S)-6-amino-1-(((S)-1-amino-4-methyl-1-oxopentan-2-yl)-amino)-1-oxohexan-2-yl)amino)-3-methyl-1-oxopentan-2-yl)amino)-3-(naphthalen-2-yl)-1-oxopropan-2-yl)carbamate (13):** HRMS (ESI-MS) *m/z*: calculated for

C<sub>33</sub>H<sub>51</sub>N<sub>6</sub>O<sub>6</sub><sup>+</sup> 627.3865, found 627.3861 [M+H]<sup>+</sup>.

**General procedure for the synthesis of compounds 14-21:** *Step 1.* Fmoc-Lys(Boc)-OH was loaded onto a 2-chlorotrityl chloride resin by shaking a mixture of 2-chlorotrityl

chloride resin (1 equiv.), Fmoc-Lys(Boc)-OH (1 equiv.) and DIPEA (3 equiv.) in DCM overnight. Fmoc chemistry was used to carry out the chain elongation. The crude carboxylic acid was cleaved from the resin with 0.5% TFA in DCM. The cleavage solution was evaporated and purified by HPLC to yield *N*<sup>6</sup>-(*tert*-butoxycarbonyl)-*N*<sup>2</sup>-(((*S*)-3-(naphthalen-2-yl)-2-propionamido-propanoyl)-*L*-isoleucyl)-*L*-lysine (**39**). <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>: CD<sub>3</sub>OD = 5:1), δ 7.78-7.75 (m, 3H), 7.64 (s, 1H), 7.45-7.44 (m, 2H), 7.34-7.30 (m, 2H), 4.82-4.80 (m, 1H), 4.51-4.40 (m, 1H), 4.21-3.98 (m, 1H), 3.25-2.80 (m, 4H), 2.18-2.15 (m, 2H), 1.82-1.26 (m, 18H), 1.04 (t, J=7.6, 3H), 0.90-0.84 (m, 6H); <sup>13</sup>CNMR (75 MHz, CDCl<sub>3</sub>), δ 175.1, 174.2, 172.1, 171.7, 156.7, 134.2, 133.5, 132.4, 128.1, 128.0, 127.6, 127.4, 126.1, 125.7, 79.3, 58.0, 57.8, 54.1, 52.3, 40.1, 38.0, 37.0, 31.4, 29.3, 28.4, 24.8, 22.7, 15.2, 10.9, 9.7. ESI-MS *m/z*: calculated for C<sub>33</sub>H<sub>49</sub>N<sub>4</sub>O<sub>7</sub><sup>+</sup> 613.36, found 613.67 [M+H]<sup>+</sup>.

*Step 2.* A solution of the acid (**39**) (0.05 mmol, 1 equiv.), the corresponding amine (0.10 mmol, 2 equiv.), HBTU (38 mg, 0.10 mmol, 2 equiv.), HOBt (14 mg, 0.10 mmol, 2 equiv.) and DIPEA (27 μL, 0.15 mmol, 3 equiv.) in THF (5 mL) was stirred at room temperature for 2 h before being concentrated. The residue was then redissolved in EtOAc, washed with saturated NaHCO<sub>3</sub>, 1.0M HCl, brine and dried over Na<sub>2</sub>SO<sub>4</sub>. After removal of the solvent under vacuum, the residue was treated with TFA (3 mL) in DCM (10 mL) at room temperature for 3 h. The reaction mixture was evaporated and the crude product was purified by RP-HPLC to provide the compounds **14-21** shown in Table 4.2 in 70% to 85% yield.

**(S)-6-Amino-2-((2S,3S)-3-methyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)pentanamido)-N-phenylhexanamide (14):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 9.74

(s, 1H), 8.32 (d, J=7.7, 1H), 8.04 (d, J=7.7, 1H), 7.83-7.77 (m, 3H), 7.70 (s, 1H), 7.57-7.54 (m, 2H), 7.45-7.38 (m, 3H), 7.33-7.27 (m, 2H), 7.12-7.08 (m, 1H), 4.81-4.76 (m, 1H), 4.50-4.45 (m, 1H), 4.24 (t, J=7.8, 1H), 3.54-3.36 (m, 1H), 3.10-3.02 (m, 1H), 2.88 (t, J=7.5, 2H), 2.21-2.14 (m, 2H), 1.87-1.29 (m, 8H), 1.23-1.16 (m, 1H), 1.01-0.88 (m, 9H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD), δ 177.2, 174.0, 173.6, 171.9, 139.4, 136.0, 134.9, 133.9, 129.9, 129.1, 128.9, 128.6, 128.4, 127.1, 126.7, 125.5, 121.3, 59.4, 56.1, 54.9, 40.6, 38.6, 38.1, 32.3, 29.9, 28.1, 26.0, 23.8, 15.9, 11.3, 10.3. HRMS (ESI-MS) *m/z*: calculated for C<sub>34</sub>H<sub>46</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup> 588.3544, found 588.3546 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-benzyl-2-((2S,3S)-3-methyl-2-((S)-3-(naphthalen-2-yl)-2-propionamido-propanamido)pentanamido)hexanamide (15):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.36 (t, J=5.8, 1H), 8.21 (d, J=7.9, 1H), 8.12 (d, J=7.7, 1H), 7.96 (d, J=7.8, 1H), 7.82-7.77 (m, 3H), 7.69 (s, 1H), 7.48-7.36 (m, 3H), 7.33-7.23 (m, 5H), 4.80-4.69 (m, 1H), 4.44-4.31 (m, 3H), 4.20-4.15 (m, 1H), 3.33-3.26 (m, 1H), 3.08-3.00 (m, 1H), 2.85 (t, J=7.7, 2H), 2.19-2.12 (m, 2H), 1.84-1.29 (m, 8H), 1.22-1.12 (m, 1H), 0.97 (t, J=7.7, 3H), 0.90-0.85 (m, 6H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD), δ 177.2, 174.1, 173.6, 173.5, 139.7, 136.0, 134.9, 133.9, 129.6, 129.1, 128.8, 128.6, 128.4, 128.3, 127.1, 126.7, 59.4, 56.1, 54.3, 44.1, 40.5, 38.5, 38.0, 32.3, 29.9, 28.0, 26.0, 23.7, 15.9, 11.3, 10.3. HRMS (ESI-MS) *m/z*: calculated for C<sub>35</sub>H<sub>48</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup> 602.3701, found 602.3700 [M+H]<sup>+</sup>.

**(S)-6-Amino-2-((2S,3S)-3-methyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido) pentanamido)-N-phenethylhexanamide (16):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.16-8.13 (m, 2H), 8.01-7.98 (m, 1H), 7.78-7.76 (m, 3H), 7.69 (s, 1H), 7.44-7.37 (m, 3H), 7.28-7.18 (m, 5H), 4.78-4.74 (m, 1H), 4.25-4.15 (m, 2H), 3.46-3.33 (m, 3H), 3.08-3.01 (m, 1H), 2.85-2.79 (m, 4H), 2.17-2.12 (m, 2H), 1.82-1.55 (m, 6H),

1.30-1.15 (m, 3H), 0.99-0.86 (m, 9H);  $^{13}\text{C}$ NMR (75 MHz,  $\text{CD}_3\text{OD}$ ),  $\delta$  177.2, 174.1, 173.6, 173.4, 140.3, 136.0, 134.9, 133.9, 129.9, 129.5, 129.1, 128.9, 128.8, 128.6, 128.3, 127.4, 127.1, 126.7, 59.4, 56.1, 54.2, 41.9, 40.5, 38.6, 38.0, 36.4, 32.4, 29.9, 28.0, 26.0, 23.7, 15.9, 11.3, 10.3. HRMS (ESI-MS)  $m/z$ : calculated for  $\text{C}_{36}\text{H}_{50}\text{N}_5\text{O}_4^+$  616.3857, found 616.3857  $[\text{M}+\text{H}]^+$ .

**(S)-6-Amino-N-benzhydryl-2-((2S,3S)-3-methyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)pentanamido)hexanamide (17):**  $^1\text{H}$ NMR (300 MHz,  $\text{CD}_3\text{OD}$ ),  $\delta$  8.71 (d,  $J=8.3$ , 1H), 8.11 (d,  $J=7.7$ , 1H), 7.95 (d,  $J=7.9$ , 1H), 7.82-7.76 (m, 3H), 7.68 (s, 1H), 7.46-7.23 (m, 13H), 6.17-6.14 (m, 1H), 4.74-4.69 (m, 1H), 4.47-4.42 (m, 1H), 4.20-4.15 (m, 1H), 3.27-3.21 (m, 1H), 3.07-2.99 (m, 1H), 2.82 (t,  $J=7.7$ , 2H), 2.19-2.12 (m, 2H), 1.83-1.73 (m, 2H), 1.70-1.48 (m, 4H), 1.38-1.34 (m, 2H), 1.20-1.10 (m, 1H), 0.97 (t,  $J=7.6$ , 3H), 0.88-0.83 (m, 6H);  $^{13}\text{C}$ NMR (75 MHz,  $\text{CD}_3\text{OD}$ ),  $\delta$  177.2, 174.0, 173.6, 172.8, 142.9, 142.7, 136.0, 134.9, 133.9, 129.6, 129.5, 129.1, 128.84, 128.77, 128.6, 128.5, 128.41, 128.37, 127.1, 126.7, 59.4, 58.3, 56.1, 54.2, 40.5, 38.5, 38.0, 32.3, 29.9, 28.0, 26.1, 23.7, 15.9, 11.3, 10.3. HRMS (ESI-MS)  $m/z$ : calculated for  $\text{C}_{41}\text{H}_{52}\text{N}_5\text{O}_4^+$  678.4014, found 678.4012  $[\text{M}+\text{H}]^+$ .

**(S)-6-Amino-N-((R)-2,3-dihydro-1H-inden-1-yl)-2-((2S,3S)-3-methyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)pentanamido)hexanamide (18):**  $^1\text{H}$ NMR (300 MHz,  $\text{CD}_3\text{OD}$ ),  $\delta$  8.25 (d,  $J=7.8$ , 1H), 8.19 (d,  $J=8.2$ , 1H), 8.10 (d,  $J=7.5$ , 1H), 7.97 (d,  $J=7.6$ , 1H), 7.81-7.75 (m, 3H), 7.66 (s, 1H), 7.46-7.34 (m, 3H), 7.23-7.11 (m, 4H), 5.35 (dd,  $J=7.7$ , 15.4, 1H), 4.73-4.68 (m, 1H), 4.37-4.33 (m, 1H), 4.21-4.16 (m, 1H), 3.28-3.21 (m, 1H), 3.04-2.96 (m, 2H), 2.89-2.80 (m, 3H), 2.52-2.42 (m, 1H), 2.17-2.10 (m, 2H), 1.90-1.80 (m, 3H), 1.72-1.37 (m, 6H), 1.28-1.14 (m, 1H), 0.98-0.87 (m, 9H);  $^{13}\text{C}$ NMR

(75 MHz, CD<sub>3</sub>OD),  $\delta$  177.2, 174.1, 173.52, 173.48, 144.5, 144.2, 136.0, 134.9, 133.9, 129.1, 129.0, 128.8, 128.6, 128.4, 127.6, 127.1, 126.7, 125.7, 125.2, 59.5, 56.1, 55.9, 54.3, 40.6, 38.5, 38.0, 34.2, 32.3, 31.1, 29.8, 28.0, 26.1, 23.8, 16.0, 11.4, 10.3. HRMS (ESI-MS)  $m/z$ : calculated for C<sub>37</sub>H<sub>50</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup> 628.3857, found 628.3859 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-((S)-2,3-dihydro-1H-inden-1-yl)-2-((2S,3S)-3-methyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)pentanamido)hexanamide (19):**

<sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD),  $\delta$  8.25-8.18 (m, 2H), 8.11 (d, J=7.6, 1H), 7.97 (d, J=7.9, 1H), 7.82-7.76 (m, 3H), 7.69 (s, 1H), 7.45-7.37 (m, 3H), 7.23-7.17 (m, 4H), 5.34 (dd, J=7.5, 15.2, 1H), 4.77-4.72 (m, 1H), 4.38-4.33 (m, 1H), 4.21-4.16 (m, 1H), 3.34-3.27 (m, 1H), 3.07-2.95 (m, 2H), 2.89-2.79 (m, 3H), 2.50-2.43 (m, 1H), 2.19-2.11 (m, 2H), 1.92-1.83 (m, 3H), 1.72-1.38 (m, 6H), 1.22-1.12 (m, 1H), 0.98-0.85 (m, 9H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD),  $\delta$  177.2, 174.0, 173.5, 173.4, 144.6, 144.2, 136.0, 134.9, 133.9, 129.1, 129.0, 128.8, 128.6, 128.4, 127.7, 127.1, 126.7, 125.8, 125.0, 59.3, 56.1, 55.9, 54.2, 40.6, 38.6, 38.0, 34.2, 32.6, 31.1, 29.9, 28.0, 26.0, 23.8, 15.9, 11.3, 10.3. HRMS (ESI-MS)  $m/z$ : calculated for C<sub>37</sub>H<sub>50</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup> 628.3857, found 628.3859 [M+H]<sup>+</sup>.

**(S)-6-Amino-2-((2S,3S)-3-methyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)pentanamido)-N-((R)-1,2,3,4-tetrahydronaphthalen-1-yl)hexanamide (20):**

<sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD),  $\delta$  8.27-8.19 (m, 2H), 8.09 (d, J=7.5, 1H), 7.97 (d, J=7.6, 1H), 7.81-7.75 (m, 3H), 7.67 (s, 1H), 7.45-7.35 (m, 3H), 7.18-7.07 (m, 4H), 5.07-5.05 (m, 1H), 4.71-4.68 (m, 1H), 4.35-4.33 (m, 1H), 4.23-4.17 (m, 1H), 3.28-3.21 (m, 1H), 3.04-2.80 (m, 5H), 2.18-2.11 (m, 2H), 2.03-1.28 (m, 12H), 1.25-1.15 (m, 1H), 0.98-0.88 (m, 9H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD),  $\delta$  177.2, 174.0, 173.5, 173.0, 138.7, 137.4, 136.0, 134.9, 133.9, 130.1, 129.5, 129.0, 128.8, 128.6, 128.4, 128.3, 127.1, 126.7, 59.4, 56.0, 54.3,

40.5, 38.5, 38.0, 32.4, 31.2, 30.2, 29.8, 28.0, 26.1, 23.8, 21.4, 16.0, 11.4, 10.3. HRMS (ESI-MS)  $m/z$ : calculated for  $C_{38}H_{52}N_5O_4^+$  642.4014, found 642.4015  $[M+H]^+$ .

**(S)-6-Amino-2-((2S,3S)-3-methyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)pentanamido)-N-((S)-1,2,3,4-tetrahydronaphthalen-1-yl)hexanamide (21):**

$^1H$ NMR (300 MHz,  $CD_3OD$ ),  $\delta$  8.26-8.18 (m, 2H), 8.11 (d,  $J=7.7$ , 1H), 7.97 (d,  $J=8.0$ , 1H), 7.82-7.77 (m, 3H), 7.70 (s, 1H), 7.45-7.37 (m, 3H), 7.13-7.08 (m, 4H), 5.04-5.03 (m, 1H), 4.78-4.73 (m, 1H), 4.38-4.33 (m, 1H), 4.22-4.17 (m, 1H), 3.27-3.21 (m, 1H), 3.08-3.00 (m, 1H), 2.89-2.79 (m, 4H), 2.18-2.09 (m, 2H), 1.96-1.29 (m, 12H), 1.22-1.12 (m, 1H), 1.02-0.86 (m, 9H);  $^{13}C$ NMR (75 MHz,  $CD_3OD$ ),  $\delta$  177.1, 174.0, 173.5, 172.8, 138.8, 137.4, 136.0, 134.9, 133.9, 130.2, 129.5, 129.1, 128.8, 128.6, 128.4, 128.3, 127.1, 126.7, 59.3, 56.0, 54.3, 40.5, 38.6, 38.0, 32.5, 31.1, 30.2, 29.9, 28.0, 26.0, 23.7, 21.1, 15.9, 11.3, 10.3. HRMS (ESI-MS)  $m/z$ : calculated for  $C_{38}H_{52}N_5O_4^+$  642.4014, found 642.4016  $[M+H]^+$ .

#### **General procedure for the synthesis of compounds 22-37.**

The amine **44** (0.1 mmol, 1 equiv.) was added to a solution of the corresponding Fmoc protected amino acid (0.1 mmol, 1 equiv.), HBTU (76 mg, 0.2 mmol, 2 equiv.), HOBt (27 mg, 0.2 mmol, 2 equiv.) and DIPEA (52  $\mu$ L, 0.3 mmol, 3 equiv.) in DMF (5 mL) and the resulting mixture was stirred at room temperature for 1 h. The solution was diluted with EtOAc and washed with saturated  $NaHCO_3$ , 1.0 M HCl, brine and dried over  $Na_2SO_4$ . After removal of the solvent under vacuum, the residue was treated with diethylamine (1.0 mL) in  $CH_3CN$  (9 mL) for 1 h. The reaction mixture was evaporated and the residue was redissolved in DCM (5 mL) and treated with propionic anhydride (39 mg, 0.3 mmol, 3 equiv.) and DIPEA (87  $\mu$ L, 0.5 mmol, 5 equiv.) for 30 min. The reaction mixture was evaporated and treated with TFA (3 mL) in DCM (10 mL) at room temperature for 3 h.

This reaction mixture was concentrated and purified by RP-HPLC to provide the compounds **22-37** in 37% to 62% yield over three steps.

**(S)-6-Amino-N-benzhydryl-2-((S)-2-cyclopentyl-2-((S)-3-(naphthalen-2-yl)-2-propion-amidopropanamido)acetamido)hexanamide (22):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.66 (d, J=8.3, 1H), 8.30 (d, J=8.0, 1H), 8.07 (d, J=7.3, 1H), 7.82-7.76 (m, 3H), 7.68 (s, 1H), 7.46-7.24 (m, 13H), 6.15 (d, J=8.2, 1H), 4.76-4.70 (m, 1H), 4.49-4.42 (m, 1H), 4.17-4.12 (m, 1H), 3.27-3.25 (m, 1H), 3.06-2.98 (m, 1H), 2.82 (t, J=7.5, 2H), 2.23-2.11 (m, 3H), 1.86-1.24 (m, 14H), 0.96 (t, J=7.6, 3H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD), δ 177.1, 173.94, 173.87, 172.8, 142.9, 142.7, 136.0, 134.9, 133.9, 129.7, 129.5, 129.0, 128.9, 128.8, 128.61, 128.58, 128.4, 127.1, 126.7, 59.0, 58.4, 56.0, 54.2, 43.4, 40.5, 38.6, 32.2, 30.5, 30.3, 29.9, 27.9, 26.2, 25.9, 23.7, 10.2. HRMS (ESI-MS) *m/z*: calculated for C<sub>42</sub>H<sub>52</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup> 690.4014, found 690.4017 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-benzhydryl-2-((S)-2-cyclohexyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)acetamido)hexanamide (23):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.69 (d, J=8.3, 1H), 8.28 (d, J=7.9, 1H), 8.10 (d, J=7.6, 1H), 7.93 (d, J=7.9, 1H), 7.82-7.77 (m, 3H), 7.69 (s, 1H), 7.46-7.24 (m, 13H), 6.15 (d, J=8.2, 1H), 4.75-4.70 (m, 1H), 4.46-4.41 (m, 1H), 4.14 (t, J=7.7, 1H), 3.27-3.26 (m, 1H), 3.07-3.00 (m, 1H), 2.82 (t, J=7.5, 2H), 2.20-2.12 (m, 2H), 1.82-1.55 (m, 9H), 1.40-1.12 (m, 6H), 1.00-0.95 (m, 5H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD), δ 177.2, 173.9, 173.4, 172.8, 142.9, 142.7, 136.0, 134.9, 133.9, 129.7, 129.6, 129.1, 128.9, 128.8, 128.6, 128.5, 128.4, 127.1, 126.7, 59.8, 58.4, 56.0, 54.2, 41.3, 40.5, 38.6, 32.3, 30.7, 30.1, 29.9, 27.9, 27.2, 27.1, 27.0, 23.7, 10.3. HRMS (ESI-MS) *m/z*: calculated for C<sub>43</sub>H<sub>54</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup> 704.4170, found 704.4168 [M+H]<sup>+</sup>.



**(S)-6-Amino-N-benzhydryl-2-((S)-3-cyclohexyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)propanamido)hexanamide (24):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.67 (d, J=8.3, 1H), 8.20-8.08 (m, 2H), 7.83-7.77 (m, 3H), 7.69 (s, 1H), 7.46-7.24 (m, 13H), 6.17 (d, J=8.3, 1H), 4.70-4.65 (m, 1H), 4.45-4.36 (m, 2H), 3.27-3.25 (m, 1H), 3.08-3.00 (m, 1H), 2.83 (t, J=7.4, 2H), 2.16 (q, J=7.7, 2H), 1.82-1.55 (m, 11H), 1.39-1.14 (m, 6H), 0.99-0.88 (m, 5H);; <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD), δ 177.2, 174.8, 174.1, 172.9, 142.9, 142.7, 135.9, 134.9, 133.9, 129.63, 129.56, 129.1, 128.9, 128.7, 128.6, 128.5, 128.44, 128.40, 127.1, 126.7, 58.3, 56.0, 54.3, 52.8, 40.5, 40.3, 38.6, 35.2, 34.8, 33.3, 32.2, 29.9, 27.9, 27.5, 27.3, 27.1, 23.7, 10.2. HRMS (ESI-MS) *m/z*: calculated for C<sub>44</sub>H<sub>56</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup> 718.4327, found 718.4326 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-benzhydryl-2-((S)-4,4-dimethyl-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)pentanamido)hexanamide (25):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.66 (d, J=8.3, 1H), 8.16-8.10 (m, 2H); 7.79-7.74 (m, 3H), 7.66 (s, 1H), 7.43-7.22 (m, 13H), 6.16-6.13 (m, 1H), 4.67-4.62 (m, 1H), 4.42-4.39 (m, 1H), 3.27-3.22 (m, 1H), 3.03-2.96 (m, 1H), 2.80 (t, J=7.4, 2H), 2.15-2.08 (m, 2H), 1.81-1.54 (m, 5H), 1.36-1.27 (m, 3H), 0.95-0.89 (m, 12H);; <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD), δ 177.2, 174.9, 173.6, 172.9, 142.9, 142.7, 136.0, 134.9, 133.9, 129.6, 129.5, 129.1, 128.8, 128.7, 128.64, 128.62, 128.5, 128.41, 128.38, 127.1, 126.7, 58.3, 56.1, 54.3, 52.7, 46.0, 40.5, 38.4, 32.3, 31.4, 30.1, 29.8, 27.9, 23.7, 10.2. HRMS (ESI-MS) *m/z*: calculated for C<sub>42</sub>H<sub>54</sub>N<sub>5</sub>O<sub>4</sub><sup>+</sup> 692.4170, found 692.4161 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-benzhydryl-2-((S)-4,4,4-trifluoro-2-((S)-3-(naphthalen-2-yl)-2-propionamidopropanamido)butanamido)hexanamide (26):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.72 (d, J=8.2, 1H), 8.18 (d, J=6.9, 1H), 7.82-7.77 (m, 3H), 7.68 (s, 1H), 7.47-

7.23 (m, 13H), 6.17 (d, J=8.3, 1H), 4.61-4.54 (m, 2H), 4.43-4.39 (m, 1H), 3.29-3.27 (m, 1H), 3.10-3.01 (m, 1H), 2.86-2.59 (m, 4H), 2.15 (q, d=7.6, 2H), 1.84-1.56 (m, 4H), 1.37-1.29 (m, 2H), 0.94 (t, J=7.6, 3H); HRMS (ESI-MS)  $m/z$ : calculated for  $C_{42}H_{54}N_5O_4^+$  704.3418, found 704.3416  $[M+H]^+$ .

**(S)-6-Amino-N-benzhydryl-2-((S)-2-((S)-3-(naphthalen-2-yl)-2-**

**propionamidopropanamido)-3-(thiophen-2-yl)propanamido)hexanamide (27):**

$^1\text{H}$ NMR (300 MHz,  $\text{CD}_3\text{OD}$ ),  $\delta$  8.70 (d, J=8.4, 1H), 8.14 (d, J=7.6, 1H), 8.07 (d, J=7.0, 1H), 7.82-7.76 (m, 3H), 7.65 (s, 1H), 7.46-7.43 (m, 2H), 7.33-7.27 (m, 11H), 7.17-7.15 (m, 1H), 6.84-6.82 (m, 2H), 6.18 (d, J=8.3, 1H), 4.61-4.53 (m, 2H), 4.47-4.40 (m, 1H), 3.28-3.22 (m, 3H), 3.05-2.98 (m, 1H), 2.83 (t, J=7.4, 2H), 2.14 (q, J=7.6, 2H), 1.80-1.58 (m, 4H), 1.37-1.32 (m, 2H), 0.94 (t, J=7.6, 3H);  $^{13}\text{C}$ NMR (75 MHz,  $\text{CD}_3\text{OD}$ ),  $\delta$  177.4, 174.0, 172.81, 172.78, 142.9, 142.7, 139.6, 135.9, 134.9, 133.9, 129.62, 129.57, 129.1, 128.8, 128.6, 128.5, 128.3, 128.0, 127.8, 127.1, 126.7, 125.6, 58.3, 56.4, 56.2, 54.4, 40.5, 38.7, 32.44, 32.38, 29.8, 27.9, 23.6, 10.2. HRMS (ESI-MS)  $m/z$ : calculated for  $C_{42}H_{54}N_5O_4^+$  718.3422, found 718.3420  $[M+H]^+$ .

**(S)-6-Amino-N-benzhydryl-2-((S)-2-((S)-3-(benzo[d]thiazol-2-yl)-2-propion-**

**amidopropanamido)-2-cyclopentylacetamido)hexanamide (28):**  $^1\text{H}$ NMR (300 MHz,  $\text{CD}_3\text{OD}$ ),  $\delta$  8.63 (d, J=7.6, 1H), 8.31 (t, J=8.2, 1H), 8.10 (d, J=6.7, 1H), 7.93 (d, J=7.9, 2H), 7.52-7.38 (m, 2H), 7.32-7.23 (m, 10H), 6.14 (d, J=7.8, 1H), 4.95-4.91 (m, 1H), 4.46-4.43 (m, 1H), 4.14-4.09 (m, 1H), 3.61-3.34 (m, 2H), 2.84 (t, J=7.0, 2H), 2.27-2.17 (m, 3H), 2.02-1.24 (m, 14H), 1.05 (t, J=7.4, 3H);  $^{13}\text{C}$ NMR (75 MHz,  $\text{CD}_3\text{OD}$ ),  $\delta$  177.2, 173.9, 172.8, 172.7, 169.2, 154.0, 142.8, 142.7, 136.5, 129.6, 129.5, 128.8, 128.6, 128.5, 128.4, 127.4, 126.5, 123.4, 122.9, 59.1, 58.3, 54.2, 43.2, 40.5, 36.22, 32.19, 30.3, 29.9, 27.9, 26.2,

25.8, 23.7, 10.1. HRMS (ESI-MS)  $m/z$ : calculated for  $C_{39}H_{49}N_6O_4S^+$  697.3531, found 697.3532  $[M+H]^+$ .

**(S)-6-Amino-N-benzhydryl-2-((S)-2-cyclopentyl-2-((S)-3-(4-fluorobenzo[d]thiazol-2-yl)-2-propionamidopropanamido)acetamido)hexanamide (29):**  $^1H$ NMR (300 MHz,  $CD_3OD$ ),  $\delta$  8.63 (d,  $J=7.3$ , 1H), 8.39-8.30 (m, 2H), 8.12 (d,  $J=7.0$ , 1H), 7.75 (d,  $J=8.0$ , 1H), 7.44-7.21 (m, 12H), 6.14 (d,  $J=8.1$ , 1H), 4.92-4.90 (m, 1H), 4.46-4.44 (m, 1H), 4.14-4.08 (m, 1H), 3.62-3.55 (m, 1H), 3.46-3.34 (m, 1H), 2.88 (t,  $J=7.5$ , 2H), 2.28-2.14 (m, 3H), 2.03-1.25 (m, 14H), 1.06 (t,  $J=7.6$ , 3H); HRMS (ESI-MS)  $m/z$ : calculated for  $C_{39}H_{48}FN_6O_4S^+$  715.3436, found 715.3436  $[M+H]^+$ .

**(S)-6-Amino-N-benzhydryl-2-((S)-2-cyclopentyl-2-((S)-3-(5-fluorobenzo[d]thiazol-2-yl)-2-propionamidopropanamido)acetamido)hexanamide (30):**  $^1H$ NMR (300 MHz,  $CD_3OD$ ),  $\delta$  8.65 (d,  $J=8.2$ , 1H), 8.39-8.30 (m, 2H), 8.09 (d,  $J=7.3$ , 1H), 7.95-7.91 (m, 1H), 7.64 (dd,  $J=2.4$ , 11.9, 1H), 7.35-7.20 (m, 11H), 6.14 (d,  $J=8.1$ , 1H), 4.93-4.90 (m, 1H), 4.47-4.44 (m, 1H), 4.16-4.11 (m, 1H), 3.61-3.54 (m, 1H), 3.45-3.37 (m, 1H), 2.86 (t,  $J=7.4$ , 2H), 2.27-2.14 (m, 3H), 1.84-1.24 (m, 14H), 1.06 (t,  $J=7.6$ , 3H); HRMS (ESI-MS)  $m/z$ : calculated for  $C_{39}H_{48}FN_6O_4S^+$  715.3436, found 715.3431  $[M+H]^+$ .

**(S)-6-Amino-N-benzhydryl-2-((S)-2-cyclopentyl-2-((S)-3-(6-fluorobenzo[d]thiazol-2-yl)-2-propionamidopropanamido)acetamido)hexanamide (31):**  $^1H$ NMR (300 MHz,  $CD_3OD$ ),  $\delta$  8.64 (d,  $J=7.9$ , 1H), 8.30 (d,  $J=8.0$ , 1H), 8.08 (d,  $J=7.5$ , 1H), 7.94-7.89 (m, 1H), 7.72 (dd,  $J=2.6$ , 8.4, 1H), 7.36-7.22 (m, 11H), 6.16-6.13 (m, 1H), 4.82-4.80 (m, 1H), 4.47-4.42 (m, 1H), 4.15-4.10 (m, 1H), 3.59-3.53 (m, 1H), 3.44-3.36 (m, 1H), 2.86 (t,  $J=7.4$ , 2H), 2.28-2.17 (m, 3H), 1.82-1.29 (m, 14H), 1.06 (t,  $J=7.6$ , 3H); HRMS (ESI-MS)  $m/z$ : calculated for  $C_{39}H_{48}FN_6O_4S^+$  715.3436, found 715.3436  $[M+H]^+$ .

**(S)-6-Amino-N-benzhydryl-2-((S)-2-((S)-3-(5-chlorobenzo[d]thiazol-2-yl)-2-propionamidopropanamido)-2-cyclopentylacetamido)hexanamide (32):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.65 (d, J=8.1, 1H), 8.30 (t, J=7.8, 1H), 8.09 (d, J=7.3, 1H), 7.92-7.89 (m, 2H), 7.40 (dd, J=1.7, 8.6, 1H), 7.34-7.20 (m, 10H), 6.13 (d, J=8.0, 1H), 4.92-4.91 (m, 1H), 4.45-4.41 (m, 1H), 4.14-4.10 (m, 1H), 3.60-3.37 (m, 2H), 2.84 (t, J=7.4, 2H), 2.26-2.13 (m, 3H), 1.83-1.23 (m, 14H), 1.05 (t, J=7.6, 3H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD), δ 177.2, 173.9, 172.8, 172.5, 171.5, 155.0, 142.8, 142.7, 135.1, 133.3, 129.6, 129.5, 128.8, 128.6, 128.5, 128.4, 126.7, 124.0, 123.1, 59.0, 58.3, 54.2, 54.1, 43.3, 40.5, 36.3, 32.2, 30.3, 29.9, 28.0, 26.2, 25.8, 23.7, 10.2. HRMS (ESI-MS) *m/z*: calculated for C<sub>39</sub>H<sub>48</sub>ClN<sub>6</sub>O<sub>4</sub>S<sup>+</sup> 731.3141, found 731.3128 [M+H]<sup>+</sup>.

**(S)-6-Amino-N-benzhydryl-2-((S)-2-((S)-3-(6-chlorobenzo[d]thiazol-2-yl)-2-propionamidopropanamido)-2-cyclopentylacetamido)hexanamide (33):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD:CCl<sub>3</sub>D=1:1), δ 8.54 (d, J=8.1, 1H), 8.35 (t, J=7.5, 1H), 8.11-8.06 (m, 3H), 7.68 (dd, J=1.9, 8.7, 1H), 7.57-7.41 (m, 10H), 6.36 (d, J=8.1, 1H), 5.09-5.04 (m, 1H), 4.66-4.61 (m, 1H), 4.30-4.25 (m, 1H), 3.77-3.61 (m, 2H), 3.07 (t, J=7.1, 2H), 2.51-2.33 (m, 3H), 2.09-1.38 (m, 14H), 1.31 (t, J=7.6, 3H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD:CCl<sub>3</sub>D=1:1), δ 176.4, 173.1, 171.8, 171.7, 168.9, 151.7, 141.7, 141.6, 136.9, 131.9, 129.1, 129.0, 128.0, 127.97, 127.6, 123.7, 121.9, 58.5, 57.6, 53.34, 53.26, 42.3, 39.9, 35.4, 31.3, 29.7, 29.6, 29.5, 27.1, 25.6, 25.3, 22.8, 9.9. HRMS (ESI-MS) *m/z*: calculated for C<sub>39</sub>H<sub>48</sub>ClN<sub>6</sub>O<sub>4</sub>S<sup>+</sup> 731.3141, found 731.3138 [M+H]<sup>+</sup>.

**(S)-6-Amino-2-((S)-2-((S)-3-(6-chlorobenzo[d]thiazol-2-yl)-2-propionamidopropanamido)-2-cyclopentylacetamido)-N-((R)-1,2,3,4-tetrahydronaphthalen-1-yl)hexanamide (34):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.00 (d,

J=2.0, 1H), 7.89 (d, J=8.7, 1H), 7.49 (dd, J=2.1, 8.7, 1H), 7.17-7.08 (m, 2H), 5.05-5.04 (m, 1H), 4.80-4.74 (m, 1H), 4.35-4.31 (m, 1H), 4.13 (d, J=8.8, 1H), 3.54-3.36 (m, 2H), 2.94-2.81 (m, 4H), 2.27-2.20 (m, 3H), 1.98-1.29 (m, 18H), 1.06 (t, J=7.6, 3H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD), δ 177.2, 173.9, 172.9, 170.1, 152.7, 138.7, 138.0, 137.5, 132.2, 130.1, 129.5, 128.3, 128.0, 127.1, 124.4, 122.5, 59.2, 54.3, 43.2, 40.6, 36.1, 32.3, 31.2, 30.4, 30.3, 30.2, 29.9, 28.0, 26.2, 25.9, 23.8, 21.4, 10.2. HRMS (ESI-MS) *m/z*: calculated for C<sub>36</sub>H<sub>48</sub>ClN<sub>6</sub>O<sub>4</sub>S<sup>+</sup> 695.3141, found 695.3134 [M+H]<sup>+</sup>.

**(S)-6-Amino-2-((S)-2-((S)-3-(6-chlorobenzo[d]thiazol-2-yl)-2-**

**propionamidopropanamido)-2-cyclopentylacetamido)-N-((R)-chroman-4-**

**yl)hexanamide (35):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.35-8.22 (m, 2H), 8.06 (d, J=6.9, 1H), 8.00 (d, J=2.0, 1H), 7.88 (d, J=8.7, 1H), 7.48 (dd, J=2.1, 8.7, 1H), 7.14-7.08 (m, 2H), 6.85-6.75 (m, 2H), 5.07-5.00 (m, 1H), 4.80-4.74 (m, 1H), 4.34-4.31 (m, 1H), 4.23-4.20 (m, 2H), 4.13-4.08 (m, 1H), 3.54-3.36 (m, 2H), 2.91 (t, J=7.5, 2H), 2.27-1.28 (m, 19H), 1.06 (t, J=7.6, 3H); <sup>13</sup>CNMR (75 MHz, DMSO-D<sub>6</sub>), δ 173.8, 171.3, 170.5, 169.6, 155.1, 151.7, 137.1, 129.9, 129.7, 129.1, 126.9, 123.8, 123.6, 122.2, 120.6, 116.9, 63.6, 56.3, 52.8, 52.4, 42.8, 36.1, 31.9, 29.1, 29.0, 28.8, 28.6, 27.2, 25.3, 25.0, 22.7, 10.2. HRMS (ESI-MS) *m/z*: calculated for C<sub>35</sub>H<sub>46</sub>ClN<sub>6</sub>O<sub>5</sub>S<sup>+</sup> 697.2933, found 697.2923 [M+H]<sup>+</sup>.

**(S)-6-Amino-2-((S)-2-((S)-3-(6-chlorobenzo[d]thiazol-2-yl)-2-**

**propionamidopropanamido)-2-(tetrahydro-2H-pyran-4-yl)acetamido)-N-((R)-**

**chroman-4-yl)hexanamide (36):** <sup>1</sup>HNMR (300 MHz, DMSO-D<sub>6</sub>), δ 8.45 (d, J=8.2, 1H), 8.36 (d, J=8.3, 1H), 8.24-8.21 (m, 2H), 7.95-7.92 (m, 2H), 7.71 (br., 3H), 7.53 (d, J=8.4, 1H), 7.17-7.08 (m, 2H), 6.86-6.76 (m, 2H), 5.01-4.99 (m, 1H), 4.85-4.84 (m, 1H), 4.31-4.20 (m, 4H), 3.86-3.83 (m, 2H), 3.35-3.18 (m, 3H), 2.78-2.73 (m, 2H), 2.15-1.89 (m, 5H),

1.65-1.14 (m, 11H), 0.94 (t, J=7.5, 3H); <sup>13</sup>CNMR (75 MHz, DMSO-D<sub>6</sub>), δ 173.9, 171.3, 170.6, 170.4, 169.5, 155.1, 151.7, 137.1, 129.9, 129.6, 129.1, 126.9, 123.8, 123.7, 122.2, 120.6, 116.9, 67.4, 67.2, 63.6, 56.9, 53.0, 52.5, 42.8, 38.2, 36.1, 31.8, 29.5, 29.1, 28.9, 28.6, 27.1, 22.8, 10.2. HRMS (ESI-MS) *m/z*: calculated for C<sub>35</sub>H<sub>46</sub>ClN<sub>6</sub>O<sub>6</sub>S<sup>+</sup> 713.2883, found 713.2870 [M+H]<sup>+</sup>.

**(S)-6-Amino-2-((S)-2-((S)-3-(6-chlorobenzo[d]thiazol-2-yl)-2-**

**propionamidopropanamido)-2-(piperidin-4-yl)acetamido)-N-((R)-chroman-4-**

**yl)hexanamide (37):** <sup>1</sup>HNMR (300 MHz, CD<sub>3</sub>OD), δ 8.49-8.40 (m, 2H), 8.25 (d, J=7.9, 1H), 7.99 (d, J=2.0, 1H), 7.90 (d, J=8.7, 1H), 7.48 (dd, J=2.0, 8.7, 1H), 7.18-7.10 (m, 2H), 6.85-6.82 (m, 1H), 6.76 (d, J=8.1, 1H), 5.08-5.00 (m, 1H), 4.80-4.73 (m, 1H), 4.37-4.32 (m, 1H), 4.23-4.14 (m, 3H), 3.62-3.34 (m, 4H), 3.06-2.88 (m, 4H), 2.29-2.21 (m, 2H), 2.09-1.23 (m, 15H), 1.07 (t, J=7.6, 3H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD), δ 177.4, 173.4, 172.8, 171.9, 169.8, 156.5, 152.8, 138.0, 132.3, 130.4, 130.1, 128.1, 124.5, 123.3, 122.6, 121.6, 118.0, 64.5, 57.8, 54.8, 54.4, 44.9, 44.8, 40.5, 37.8, 36.2, 32.4, 30.2, 29.9, 28.2, 26.9, 26.0, 24.0, 10.2. HRMS (ESI-MS) *m/z*: calculated for C<sub>35</sub>H<sub>47</sub>ClN<sub>7</sub>O<sub>5</sub>S<sup>+</sup> 712.3042, found 712.3028 [M+H]<sup>+</sup>.

**(S)-2-((S)-2-((S)-3-(6-Chlorobenzo[d]thiazol-2-yl)-2-propionamidopropanamido)-2-**  
**cyclo- pentylacetamido)-N-((R)-chroman-4-yl)-6-(dimethylamino)hexanamide (38):**

Na(AcO)<sub>3</sub>BH (4 equiv.) was added to a solution of **35** (1 equiv.) and formaldehyde (10 equiv.) in ClCH<sub>2</sub>CH<sub>2</sub>Cl and the resulting mixture was stirred at room temperature for 1 h. The reaction was quenched with 10% NaHCO<sub>3</sub> solution and evaporated to give a residue which was purified by HPLC to give **38** (81% yield). <sup>1</sup>HNMR (400 MHz, DMSO-D<sub>6</sub>), δ 9.45 (br., 1H), 8.43 (d, J=8.3, 1H), 8.34 (d, J=8.4, 1H), 8.23 (d, J=2.2, 1H), 8.14 (d, J=8.1, 1H),

7.94-7.92 (m, 2H), 7.52 (dd, J=2.2, 8.7, 1H), 7.16-7.10 (m, 2H), 6.85-6.81 (m, 1H), 6.79-6.77 (m, 1H), 5.03-4.98 (m, 1H), 4.86-4.80 (m, 1H), 4.30-4.15 (m, 4H), 3.55-3.47 (m, 1H), 3.33-3.26 (m, 1H), 3.03-2.98 (m, 2H), 2.76 (d, J=3.6, 6H), 2.21-2.00 (m, 4H), 1.88-1.83 (m, 1H), 1.70-1.24 (m, 14H), 0.94 (t, J=7.6, 3H);  $^{13}\text{C}$ NMR (100 MHz, DMSO- $\text{D}_6$ ),  $\delta$  173.8, 171.3, 171.2, 170.5, 169.6, 155.1, 151.7, 137.1, 129.9, 129.7, 129.1, 127.0, 123.8, 123.6, 122.2, 120.6, 116.9, 63.6, 56.9, 56.2, 52.7, 52.5, 42.8, 42.7, 42.6, 36.1, 31.9, 29.2, 29.0, 28.9, 28.6, 25.3, 25.0, 23.8, 22.8, 10.2. HRMS (ESI-MS)  $m/z$ : calculated for  $\text{C}_{37}\text{H}_{50}\text{ClN}_6\text{O}_5\text{S}^+$  725.3246, found 725.3257  $[\text{M}+\text{H}]^+$ .

**General procedure for the synthesis of amino acids 41a-c:** To a solution of (S)-3-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-(allyloxy)-4-oxobutanoic acid **42** (395 mg, 1 mmol, 1 equiv.) in DCM (30 mL), oxalyl chloride (257  $\mu\text{L}$ , 3 equiv.) and a catalytic amount of DMF were added at 0 °C. After gas evolution ceased, the reaction mixture was concentrated. The residue was suspended in toluene (25 mL) and treated with different substituted 2-aminobenzenethiols (1 mmol, 1 equiv.). The resultant mixture was stirred overnight at room temperature. The solution was diluted with EtOAc and washed with saturated  $\text{NaHCO}_3$ , 1.0M HCl, brine and dried over  $\text{Na}_2\text{SO}_4$ . The solvent was evaporated and the crude product was purified by flash chromatography on silica gel. Phenylsilane (130 mg, 1.2 mmol, 3 equiv.) was added to a solution of the product of the previous step (0.4 mmol, 1 equiv.) and then tetrakis(triphenylphosphine)-palladium(0) (46 mg, 0.04 mmol, 0.1 equiv.) in DCM was added. The resultant solution was stirred for 1 h before being concentrated. The residue was purified by flash chromatography on silica gel to afford the Fmoc protected amino acids (**41a-c**) in 30% to 51% yield over two steps.

**(S)-2-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-3-(4-fluorobenzo[d]thiazol-2-yl)propanoic acid (41a):** <sup>1</sup>HNMR (400 MHz, DMSO), δ 13.0 (br. 1H), 7.97-7.87 (m, 4H), 7.66-7.64 (m, 2H), 7.87 (d, J=7.5, 2H), 7.48-7.22 (m, 6H), 4.54 (dt, J=4.0, 8.8, 1H), 4.28 (d, J=6.3, 2H), 4.20 (t, J=6.8, 1H), 3.65 (dd, J=4.6, 15.1, 1H), 3.49 (dd, J=9.9, 15.1, 1H); ESI-MS *m/z*: calculated for C<sub>25</sub>H<sub>20</sub>FN<sub>2</sub>O<sub>4</sub>S<sup>+</sup> 463.11, found 463.2 [M+H]<sup>+</sup>.

**(S)-2-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-3-(5-fluorobenzo[d]thiazol-2-yl)propanoic acid (41b):** <sup>1</sup>HNMR (400 MHz, DMSO), δ 13.0 (br. 1H), 8.12-8.09 (m, 1H), 7.94 (d, J=8.6, 1H), 7.87 (d, J=7.5, 2H), 7.79 (dd, J=2.5, 9.9, 1H), 7.67-7.64 (m, 2H), 7.41-7.23 (m, 5H), 4.55 (dt, J=4.0, 9.4, 1H), 4.29 (d, J=6.7, 2H), 4.20 (t, J=6.8, 1H), 3.63 (dd, J=4.5, 15.1, 1H), 3.47 (dd, J=9.8, 15.1, 1H); <sup>13</sup>CNMR (75 MHz, CD<sub>3</sub>OD). ESI-MS *m/z*: calculated for C<sub>25</sub>H<sub>20</sub>FN<sub>2</sub>O<sub>4</sub>S<sup>+</sup> 463.1, found 463.2 [M+H]<sup>+</sup>.

**(S)-2-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-3-(6-fluorobenzo[d]thiazol-2-yl)propanoic acid (41c):** <sup>1</sup>HNMR (400 MHz, DMSO), δ 13.0 (br. 1H), 7.99-7.92 (m, 3H), 7.88 (d, J=7.5, 2H), 7.67-7.64 (m, 2H), 7.42-7.34 (m, 3H), 7.30-7.23 (m, 2H), 4.54 (dt, J=4.4, 8.6, 1H), 4.28 (d, J=7.0, 2H), 4.20 (t, J=6.8, 1H), 3.60 (dd, J=4.3, 15.1, 1H), 3.45 (dd, J=9.8, 15.1, 1H); ESI-MS *m/z*: calculated for C<sub>25</sub>H<sub>20</sub>FN<sub>2</sub>O<sub>4</sub>S<sup>+</sup> 463.11, found 463.2 [M+H]<sup>+</sup>.

**General procedure for the synthesis of compounds 43a-c:** The amine (10 mmol, 1 equiv.) was added to a solution of Fmoc-Lys(Boc)-OH **40** (4.7 g, 10 mmol, 1 equiv.), HBTU (7.6 g, 20 mmol, 2 equiv.), HOBt (2.7 g, 20 mmol, 2 equiv.) and DIPEA (5.2 mL, 30 mmol, 3 equiv.) in DMF (50 mL) and the resultant mixture was stirred at room temperature for 2 h. The solution was diluted with EtOAc and washed with saturated NaHCO<sub>3</sub>, 1.0M HCl, brine then dried over Na<sub>2</sub>SO<sub>4</sub>. After removal of the solvent under



vacuum, the residue was treated with diethylamine (10 mL) in CH<sub>3</sub>CN (90 mL) for 1 h. The solvent was evaporated and the crude product was purified by flash chromatography on silica gel to afford **43**.

**tert-Butyl (S)-(5-amino-6-(benzhydrylamino)-6-oxohexyl)carbamate (43a):** **43a** was produced in 76% yield over two steps. <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>), δ 8.25 (d, J=8.6, 1H), 7.30-7.18 (m, 10H), 6.19 (d, J=8.6, 1H), 5.05 (brd. 1H), 3.33-3.29 (m, 1H), 3.02-3.00 (m, 2H), 1.87-1.76 (m, 3H), 1.57-1.30 (m, 14H); <sup>13</sup>CNMR (75 MHz, CDCl<sub>3</sub>), δ 174.4, 156.2, 141.8, 128.58, 128.56, 127.4, 127.3, 78.8, 56.3, 54.9, 40.1, 34.6, 29.7, 28.5, 22.8. ESI-MS *m/z*: calculated for C<sub>24</sub>H<sub>33</sub>N<sub>3</sub>O<sub>3</sub><sup>+</sup> 412.26, found 412.67 [M+H]<sup>+</sup>.

**tert-Butyl ((S)-5-amino-6-oxo-6-(((R)-1,2,3,4-tetrahydronaphthalen-1-yl)amino)hexyl)carbamate (43b):** <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>), δ 7.63 (d, J=8.8, 1H), 7.32-7.05 (m, 4H), 5.15-5.10 (m, 1H), 4.96 (brd., 1H), 3.37-3.33 (m, 1H), 3.11-3.09 (m, 2H), 2.82-2.70 (m, 2H), 2.03-1.99 (m, 1H), 1.84-1.75 (m, 6H), 1.57-1.41 (m, 14H); <sup>13</sup>CNMR (75 MHz, CDCl<sub>3</sub>), δ 174.3, 156.1, 137.4, 136.9, 129.1, 128.4, 127.1, 126.1, 78.9, 55.0, 46.9, 40.1, 34.7, 30.2, 29.8, 29.2, 28.4, 22.9, 20.1. ESI-MS *m/z*: calculated for C<sub>21</sub>H<sub>34</sub>N<sub>3</sub>O<sub>3</sub><sup>+</sup> 376.26, found 376.83 [M+H]<sup>+</sup>.

**tert-Butyl ((S)-5-amino-6-(((R)-chroman-4-yl)amino)-6-oxohexyl)carbamate (43c):** <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>), δ 7.74 (d, J=8.1, 1H), 7.16-7.12 (m, 2H), 6.89-6.78 (m, 2H), 5.08-5.06 (m, 1H), 4.25-4.12 (m, 1H), 3.35-3.31 (m, 1H), 3.11-3.09 (m, 2H), 2.19-2.17 (m, 1H), 2.00-1.84 (m, 2H), 1.62-1.41 (m, 16H); <sup>13</sup>CNMR (75 MHz, CDCl<sub>3</sub>), δ 174.7, 156.1, 154.9, 129.1, 128.9, 122.3, 120.6, 116.9, 78.8, 63.3, 54.9, 42.9, 40.1, 34.6, 29.8, 29.1, 28.4, 22.9. ESI-MS *m/z*: calculated for C<sub>20</sub>H<sub>32</sub>N<sub>3</sub>O<sub>4</sub><sup>+</sup> 378.24, found 378.67 [M+H]<sup>+</sup>.

**General procedure for the synthesis of compounds 44a-j:** Compound **43** (0.3 mmol, 1 equiv.) was added to a solution of the corresponding Fmoc protected amino acid (0.3 mmol, 1 equiv.), HBTU (228 mg, 0.6 mmol, 2 equiv.), HOBt (81 mg, 0.6 mmol, 2 equiv.) and DIPEA (157  $\mu$ L, 0.9 mmol, 3 equiv.) in DMF (10 mL) and the resultant mixture was stirred at room temperature for 1 h. The solution was diluted with EtOAc and washed with saturated NaHCO<sub>3</sub>, 1.0M HCl, brine then dried over Na<sub>2</sub>SO<sub>4</sub>. After removal of the solvent under vacuum, the residue was treated with 1 mL diethylamine in CH<sub>3</sub>CN (9 mL) for 1 h. The reaction mixture was evaporated and the crude product was purified by flash chromatography on silica gel to afford compounds **44** in 59-75% yield over two steps.

**tert-Butyl ((S)-5-((S)-2-amino-2-cyclopentylacetamido)-6-(benzhydrylamino)-6-oxohexyl)-carbamate (44a):** <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>),  $\delta$  8.16 (d, J=8.4, 1H), 7.76 (d, J=8.5, 1H), 7.35-7.16 (m, 10H), 6.23 (d, J=8.4, 1H), 4.81 (brd. 1H), 4.74-4.67 (m, 1H), 3.04-2.98 (m, 3H), 2.09-2.04 (m, 1H), 1.89-1.78 (m, 1H), 1.73-1.43 (m, 20H), 1.36-1.18 (m, 4H); <sup>13</sup>CNMR (75 MHz, CDCl<sub>3</sub>),  $\delta$  175.4, 171.0, 156.1, 141.7, 141.4, 128.6, 128.5, 127.6, 127.43, 127.38, 127.2, 79.0, 58.4, 56.7, 52.5, 43.7, 40.2, 32.4, 29.5, 29.3, 28.5, 27.6, 25.6, 25.4, 22.7. ESI-MS *m/z*: calculated for C<sub>31</sub>H<sub>45</sub>N<sub>4</sub>O<sub>4</sub><sup>+</sup> 537.34, found 537.92 [M+H]<sup>+</sup>.

**tert-Butyl ((S)-5-((S)-2-amino-2-cyclohexylacetamido)-6-(benzhydrylamino)-6-oxohexyl)-carbamate (44b):** <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>),  $\delta$  7.80 (d, J=8.2, 1H), 7.72 (d, J=8.3, 1H), 7.35-7.18 (m, 10H), 6.21 (d, J=8.3, 1H), 4.65 (brd. 1H), 4.59-4.51 (m, 1H), 3.06-2.95 (m, 3H), 1.95-1.58 (m, 8H), 1.45-1.42 (m, 11H), 1.36-1.00 (m, 9H); <sup>13</sup>CNMR (75 MHz, CDCl<sub>3</sub>),  $\delta$  175.0, 170.9, 156.1, 141.6, 141.4, 128.64, 128.56, 127.5, 127.3, 79.1,

59.9, 56.9, 52.7, 41.0, 40.2, 31.6, 30.2, 29.5, 28.4, 26.7, 26.3, 26.14, 26.07, 22.8. ESI-MS  $m/z$ : calculated for  $C_{32}H_{47}N_4O_4^+$  551.36, found 551.75  $[M+H]^+$ .

**tert-Butyl ((S)-5-((S)-2-amino-3-cyclohexylpropanamido)-6-(benzhydrylamino)-6-oxo-hexyl)carbamate (44c):**  $^1H$ NMR (300 MHz,  $CDCl_3$ ),  $\delta$  8.02 (d,  $J=8.5$ , 1H), 7.85 (d,  $J=8.5$ , 1H), 7.35-7.15 (m, 10H), 6.22 (d,  $J=8.3$ , 1H), 4.72 (t,  $J=5.8$ , 1H), 4.67-4.59 (m, 1H), 3.17-3.13 (m, 1H), 3.04-2.93 (m, 2H), 1.90-1.79 (m, 1H), 1.66-1.58 (m, 9H), 1.49-1.42 (m, 11H), 1.35-1.08 (m, 7H), 0.98-0.77 (m, 2H);  $^{13}C$ NMR (75 MHz,  $CDCl_3$ ),  $\delta$  176.3, 171.0, 156.0, 141.6, 141.5, 128.6, 128.5, 127.6, 127.4, 127.2, 78.9, 56.7, 52.63, 52.56, 42.5, 40.2, 34.2, 34.1, 32.2, 32.0, 29.6, 28.5, 26.4, 26.3, 26.1, 22.7. ESI-MS  $m/z$ : calculated for  $C_{33}H_{49}N_4O_4^+$  565.37, found 565.58  $[M+H]^+$ .

**tert-Butyl ((S)-5-((S)-2-amino-4,4-dimethylpentanamido)-6-(benzhydrylamino)-6-oxo-hexyl)carbamate (44d):**  $^1H$ NMR (300 MHz,  $CDCl_3$ ),  $\delta$  7.91 (d,  $J=8.4$ , 1H), 7.83 (d,  $J=8.4$ , 1H), 7.32-7.16 (m, 10H), 6.22 (d,  $J=8.3$ , 1H), 4.68 (brd., 1H), 4.61-4.53 (m, 1H), 3.15-3.02 (m, 3H), 1.94-1.79 (m, 2H), 1.69-1.42 (m, 13H), 1.32-1.09 (m, 4H), 0.93 (s, 9H);  $^{13}C$ NMR (75 MHz,  $CDCl_3$ ),  $\delta$  176.8, 170.9, 156.0, 141.6, 141.5, 128.6, 128.5, 127.5, 127.4, 127.2, 79.0, 56.7, 52.9, 52.7, 49.2, 40.1, 31.9, 30.7, 30.0, 29.6, 28.5, 22.8. ESI-MS  $m/z$ : calculated for  $C_{33}H_{49}N_4O_4^+$  539.36, found 539.58  $[M+H]^+$ .

**tert-Butyl ((S)-5-((S)-2-amino-4,4,4-trifluorobutanamido)-6-(benzhydrylamino)-6-oxo-hexyl)carbamate (44e):**  $^1H$ NMR (300 MHz,  $CDCl_3$ ),  $\delta$  8.12 (d,  $J=8.5$ , 1H), 7.80 (d,  $J=8.3$ , 1H), 7.36-7.14 (m, 10H), 6.20 (d,  $J=8.3$ , 1H), 4.74 (d,  $J=5.9$ , 1H), 4.63-4.56 (m, 1H), 3.42-3.25 (m, 2H), 3.04-2.95 (m, 2H), 2.78-2.60 (m, 1H), 1.91-1.61 (m, 4H), 1.47-1.42 (m, 11H), 1.35-1.23 (m, 2H); ESI-MS  $m/z$ : calculated for  $C_{28}H_{38}F_3N_4O_4^+$  551.28, found 551.92  $[M+H]^+$ .

**tert-Butyl ((S)-5-((S)-2-amino-3-(thiophen-2-yl)propanamido)-6-(benzhydrylamino)-6-oxo-hexyl)carbamate (44f):** <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>), δ 8.07 (d, J=8.2, 1H), 7.95 (d, J=8.4, 1H), 7.34-7.15 (m, 11H), 6.93-6.90 (m, 1H), 6.78-6.77 (m, 1H), 6.22 (d, J=8.4, 1H), 4.76-4.64 (m, 2H), 3.36-2.94 (m, 5H), 1.84-1.63 (m, 3H), 1.45-1.42 (m, 11H), 1.24-1.08 (m, 3H); <sup>13</sup>CNMR (75 MHz, CDCl<sub>3</sub>), δ 174.0, 170.8, 156.1, 141.6, 141.4, 139.2, 128.64, 128.56, 127.54, 127.46, 127.3, 127.0, 126.7, 124.7, 79.0, 56.7, 55.9, 52.6, 40.3, 35.0, 32.3, 29.5, 28.5, 22.6. ESI-MS *m/z*: calculated for C<sub>28</sub>H<sub>38</sub>F<sub>3</sub>N<sub>4</sub>O<sub>4</sub><sup>+</sup> 565.28, found 565.25 [M+H]<sup>+</sup>.

**tert-Butyl ((5S)-5-((S)-2-amino-2-cyclopentylacetamido)-6-oxo-6-((1,2,3,4-tetrahydro-naphthalen-1-yl)amino)hexyl)carbamate (44g):** <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>), δ 7.77 (d, J=8.5, 1H), 7.58 (d, J=8.1, 1H), 7.18-7.04 (m, 4H), 5.14-5.09 (m, 1H), 4.92 (br, 1H), 4.62-4.54 (m, 1H), 3.08-3.06 (m, 2H), 2.76-2.72 (m, 3H), 2.03-2.01 (m, 2H), 1.82-1.77 (m, 4H), 1.66-1.20 (m, 24H); <sup>13</sup>CNMR (75 MHz, CDCl<sub>3</sub>), δ 175.1, 171.1, 156.1, 137.4, 136.5, 129.0, 128.5, 127.0, 126.1, 78.9, 58.3, 52.3, 47.4, 43.5, 40.3, 32.9, 30.2, 29.4, 29.3, 28.4, 27.5, 25.6, 25.4, 22.7, 20.2. ESI-MS *m/z*: calculated for C<sub>28</sub>H<sub>45</sub>N<sub>4</sub>O<sub>4</sub><sup>+</sup> 501.34, found 501.83 [M+H]<sup>+</sup>.

**tert-Butyl ((5S)-5-((S)-2-amino-2-cyclopentylacetamido)-6-(chroman-4-ylamino)-6-oxo-hexyl)carbamate (44h):** <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>), δ 7.95 (d, J=7.4, 1H), 7.80 (d, J=8.5, 1H), 7.14-7.10 (m, 2H), 6.87-6.77 (m, 2H), 5.11-5.05 (m, 1H), 4.94 (br, 1H), 4.67-4.57 (m, 1H), 4.26-4.15 (m, 2H), 3.07-3.05 (m, 2H), 2.71 (d, J=6.1, 1H), 2.19-2.13 (m, 1H), 2.03-1.99 (m, 2H), 1.81-1.77 (m, 1H), 1.67-1.17 (m, 24H); <sup>13</sup>CNMR (75 MHz, CDCl<sub>3</sub>), δ 175.2, 171.3, 156.1, 155.1, 129.4, 128.9, 121.9, 120.6, 116.9, 78.9, 77.5, 77.1, 76.7,

63.3, 58.1, 52.2, 43.5, 43.4, 40.3, 32.9, 29.4, 29.3, 29.1, 28.4, 27.5, 25.6, 25.4, 22.7. ESI-MS  $m/z$ : calculated for  $C_{27}H_{43}N_4O_5^+$  503.32, found 503.83  $[M+H]^+$ .

**tert-Butyl ((S)-5-((S)-2-amino-2-(tetrahydro-2H-pyran-4-yl)acetamido)-6-(((R)-chroman-4-yl)amino)-6-oxohexyl)carbamate (44i):**  $^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ ),  $\delta$  7.91-7.81 (m, 2H), 7.13-7.08 (m, 2H), 6.85-6.75 (m, 2H), 5.09-5.07 (m, 1H), 4.90 (br, 1H), 4.58-4.51 (m, 1H), 4.24-4.14 (m, 2H), 3.96-3.92 (m, 2H), 3.36-3.26 (m, 2H), 3.04-3.02 (m, 1H), 2.70-2.69 (m, 1H), 2.17-2.14 (m, 1H), 2.01-1.97 (m, 1H), 1.82-1.75 (m, 2H), 1.66-1.29 (m, 20H);  $^{13}\text{C NMR}$  (75 MHz,  $\text{CDCl}_3$ ),  $\delta$  174.0, 171.1, 156.1, 155.0, 129.3, 129.0, 121.9, 120.6, 117.0, 79.0, 67.7, 63.3, 58.9, 52.4, 43.4, 40.2, 38.3, 32.9, 29.5, 29.2, 28.4, 26.8, 22.8. ESI-MS  $m/z$ : calculated for  $C_{27}H_{43}N_4O_6^+$  519.31, found 519.92  $[M+H]^+$ .

**tert-Butyl 4-((S)-1-amino-2-(((S)-6-((tert-butoxycarbonyl)amino)-1-(((R)-chroman-4-yl)-amino)-1-oxohexan-2-yl)amino)-2-oxoethyl)piperidine-1-carboxylate (44j):**  $^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ ),  $\delta$  7.92-7.76 (m, 2H), 7.13-7.08 (m, 2H), 6.84-6.74 (m, 2H), 5.08-5.06 (m, 1H), 4.94 (br, 1H), 4.53-4.51 (m, 1H), 4.19-4.07 (m, 4H), 3.15-3.02 (m, 4H), 2.84-2.82 (m, 1H), 2.62-2.58 (m, 1H), 2.15-2.13 (m, 1H), 2.00-1.92 (m, 2H), 1.77-1.17 (m, 30H);  $^{13}\text{C NMR}$  (75 MHz,  $\text{CDCl}_3$ ),  $\delta$  173.9, 171.2, 156.1, 155.0, 154.6, 129.3, 129.0, 121.8, 120.7, 117.0, 79.5, 79.0, 63.3, 58.9, 52.6, 43.5, 40.1, 39.3, 32.6, 29.5, 29.1, 28.8, 28.4, 26.1, 22.7. ESI-MS  $m/z$ : calculated for  $C_{32}H_{52}N_5O_7^+$  618.39, found 618.75  $[M+H]^+$ .

**Computational Modeling.** The crystal structure of DCN1 in a complex with the UBC12 peptide and cullin 1 (PDB ID 3TDU<sup>15</sup>) was used to construct the binding models of designed molecules with DCN1. Coordinates of DCN1 protein were extracted from the co-crystal structure<sup>15</sup> and the structurally unresolved side chains in DCN1 were added using the MOE program.<sup>80</sup> Hydrogen atoms not determined in the crystal structure were

added using the “protonate 3D” module in MOE in the physiological conditions. In the docking simulations, the designed compounds were first prepared using the MOE program and then subject to the docking calculations using the GOLD program (version 5.1).<sup>95, 96</sup> The binding site was centered at F164 in DCN1 and had a radius of 12 Å. In each genetic algorithm (GA) run, a maximum number of 200,000 operations were performed on a population of 5 islands each with 100 individuals. Operator weights for crossover, mutation and migration were set to 95, 95 and 10 respectively. The selection of docking poses were assessed by ChemScore (the fitness function) implemented in Gold 5.1. The top-ranked conformations based on ChemScore were analyzed to determine the predicted docking modes.

**Fluorescence Polarization (FP) competitive binding assays.** The immunoprecipitation assay was performed as described previously. The Fluorescence Polarization (FP) competitive binding assays were performed similarly as described previously<sup>16</sup> to determine accurately the binding affinities of DCN1 inhibitors. A novel FAM labeled fluorescent probe compound (**4**) was designed and synthesized based on the optimized UBC12 peptide. Equilibrium dissociation constant ( $K_D$ ) values of **4** to DCN1 were determined from protein saturation experiments by monitoring the total FP values of mixtures composed with the fluorescent probe at a fixed concentration and proteins with increasing concentrations up to full saturation. Serial dilutions of proteins were mixed with **4** to a final volume of 200  $\mu$ l in the assay buffer (100 mM phosphate buffer, pH = 6.5, with 0.02% Tween-20 and 2% DMSO). The final probe concentration was 5 nM for both assays. Plates were incubated at room temperature for 30 min with gentle shaking to ensure equilibration. FP values in millipolarization units (mP) were measured using the

Infinite M-1000 plate reader (Tecan U.S., Research Triangle Park, NC) in Microfluor 1 96-well, black, round-bottom plates (Thermo Scientific, Waltham, MA) at an excitation wavelength of 485 nm and an emission wavelength of 530 nm. The  $K_D$  values of **4** were calculated by fitting the sigmoidal dose-dependent FP increases to a function of protein concentrations using Graphpad Prism 6.0 software (Graphpad Software, San Diego, CA). The  $IC_{50}$  and  $K_i$  values of compounds were determined in competitive binding experiments. Mixtures of 4  $\mu$ l of compounds in DMSO and 196  $\mu$ l of pre-incubated protein/probe complex solution in the assay buffer were added into assay plates which were incubated at room temperature for 30 minutes with gentle shaking. Final concentrations of DCN1 protein and fluorescent probe (**4**) were 200 and 5 nM, respectively. Negative controls containing protein/probe complex only (equivalent to 0% inhibition), and positive controls containing only free probes (equivalent to 100% inhibition), were included in each assay plate. FP values were measured as described above.  $IC_{50}$  values were determined by nonlinear regression fitting of the competition curves.  $K_i$  values of competitive inhibitors were obtained directly by nonlinear regression fitting, based upon the  $K_D$  values of the probe (**4**), concentrations of the protein and probe in the competitive assays. All the FP competitive experiments were performed in duplicate in three independent experiments.

**Bio-Layer Interferometry (BLI) Method.** Bio-Layer Interferometry (BLI) assay was performed similarly as described previously.<sup>16</sup> Purified recombinant DCN1 protein was biotinylated using the Thermo EZ-Link long-chain biotinylation reagent. Briefly, protein and biotinylation reagent were mixed in a 1:1 molar ratio in PBS at 4 °C. A low concentration of biotinylation reagent was applied to avoid protein over-biotinylation. This

reaction mixture was incubated at 4 °C for 2 h to allow the reaction to complete. The reaction mixture was then dialyzed using Fisherscientific 10K MWCO dialysis cassettes to remove unreacted biotinylation reagent.

BLI experiments were performed using an OctetRED96 instrument from PALL/ForteBio. All assays were run at 30 °C with continuous 1000 RPM shaking. PBS with 0.1% BSA, 0.01% Tween-20 and 2% DMSO was used as the assay buffer. Biotinylated DCN1 protein was tethered to Super Streptavidin (SSA) biosensors (ForteBio) by dipping the sensors into 10 µg/mL protein solutions. An average saturation response level of 11-12 nm was achieved in 15 min. Biotinylated blocked Streptavidin (SAV B4) sensors were prepared by following the protocol provided by the manufacturer and were used as the inactive reference controls. Sensors with tethered proteins were washed in assay buffer for 10 min to eliminate loose nonspecific bound protein molecules and establish a stable base line before starting association-dissociation cycles of the test compound. Different association and dissociation times were used depending on the potencies of compounds being tested. A DMSO-only reference was included in all assays. Raw kinetic data collected were processed in the Data Analysis software provided by the manufacturer using double reference subtraction in which both DMSO-only reference and inactive protein reference were subtracted. The resulting data were analyzed based on 1:1 binding model from which  $k_{on}$  and  $k_{off}$  values were obtained and then  $K_D$  values were calculated. All the BLI experiments were performed in two independent experiments.



**Aqueous Solubility Assay.** Aqueous Solubility Assay was performed similarly as described previously.<sup>16</sup> The equilibrium aqueous solubility of an individual compound at pH 7.4 was determined with a small-scale shake-flask method in Eppendorf tubes. Compounds were added as powder in excess to phosphate buffer at pH 7.4 until heterogeneous suspensions were obtained. Suspensions containing excessive compound powder were sonicated in a water bath for 30 min followed by incubation at room temperature with shaking for 24 h to achieve thermodynamic equilibrium. Suspensions were centrifuged at 15000 RPM for 15 min. Supernatants obtained were filtered using Millipore MultiScreen filter plates equipped with 0.45  $\mu\text{m}$  membranes. Compound concentrations in the saturated solutions were measured by UV spectroscopy at the maximum absorption wavelength of each compound, pre-determined by a spectrum scan.

**Cloning and Purification of DCN1 proteins.** DCN1 proteins were cloned and purified similarly as described previously.<sup>16</sup> Human DCN1 (residues 58-259) was expressed and purified as described previously. Briefly, the N-terminal His<sub>6</sub> tagged protein was expressed in Rosetta2 cells and purified by Ni-NTA (Qiagen) followed by a Superdex 75 (GE Healthcare) column pre-equilibrated with 25 mM Tris pH 7.5, 200 mM NaCl and 1 mM DTT. The uncleaved DCN1 protein was stored at -80 °C without glycerol.

**Crystallization and Structure Determination of 38-DCN1 Complex.** Crystallization experiments were performed similarly as described previously.<sup>16</sup> Prior to crystallization, DCN1 in 25 mM Tris-HCl pH 7.5, 200 mM NaCl and 1 mM DTT was concentrated to 10

mg/mL and incubated for 1 h at 4 °C with compound **38** in a protein to compound molar ratio of 1:1.3. Crystals were grown at 20 °C from sitting drop vapor diffusion experiments. The drops contained 1 µL of DCN1:**38** and 1 µL of well solution (37 % PEG 300, 0.1 M phosphate-citrate pH 4.2). The crystal growth conditions were cryoprotective, therefore crystals were mounted directly from the drops and stored in liquid nitrogen prior to data collection.

Diffraction data were collected on a Mar225 detector mounted on the LS-CAT 21-ID-F beamline at the Advanced Photon Source and processed with HKL2000<sup>97</sup>. The structure was solved by molecular replacement (Molrep<sup>98</sup>) using an in-house DCN1 structure lacking its bound ligand as the search model. The resulting DCN1:**38** structure had two protein molecules in the asymmetric unit, each containing one bound compound **38** molecule. The structure was fit iteratively and refined to 2.16 Å resolution using Coot<sup>99</sup> and Buster<sup>100</sup>, respectively. The coordinates and restraints for the compound were determined using Grade<sup>100</sup> with the mogul+qm option. Residues 62 – 251 from the A chain and residues 62-75; 80-251 from chain B were visible in the electron density maps. Data collection and refinement statistics are listed in Table A1.

**Table A1 Crystallography Data Collection and Refinement Statistics**

<b>Data Collection</b>	<b>DCN1-38</b>
PDBID	6B5Q
Space Group	C222 <sub>1</sub>
Unit Cell (Å)	a = 88.87 b= 90.841 c = 105.010
	$\alpha = \beta = \gamma = 90^\circ$
Wavelength (Å)	0.9787
Resolution (Å) <sup>1</sup>	2.16 (2.2 - 2.16)
Rmerge <sup>2</sup>	0.048 (0.226)
$\langle I/\sigma I \rangle^3$	20 (5)
Completeness (%) <sup>4</sup>	99.8 (100)
Redundancy	7.4 (7.3)
<b>Refinement</b>	
Resolution (Å)	2.16
R-Factor <sup>5</sup>	0.1955
Rfree <sup>6</sup>	0.2417
Protein atoms	3025
Ligands	2
Water Molecules	110
Unique Reflections	22721
R.m.s.d. <sup>7</sup>	
Bonds	0.01
Angles	0.99
MolProbity Score <sup>8</sup>	1.07
Clash Score <sup>8</sup>	2.12
RSCC <sup>9</sup>	0.94/0.94
RSR <sup>9</sup>	0.12/0/11

<sup>1</sup>Statistics for highest resolution bin of reflections in parentheses.

<sup>2</sup> $R_{\text{merge}} = \sum_h \sum_j |I_{hj} - \langle I_h \rangle| / \sum_h \sum_j I_{hj}$ , where  $I_{hj}$  is the intensity of observation  $j$  of reflection  $h$  and  $\langle I_h \rangle$  is the mean intensity for multiply recorded reflections.

<sup>3</sup>Intensity signal-to-noise ratio.

<sup>4</sup>Completeness of the unique diffraction data.

<sup>5</sup>R-factor =  $\sum_h | |F_o| - |F_c| | / \sum_h |F_o|$ , where  $F_o$  and  $F_c$  are the observed and calculated structure factor amplitudes for reflection  $h$ .

<sup>6</sup> $R_{\text{free}}$  is calculated against a 5% random sampling of the reflections that were removed before structure refinement.

<sup>7</sup>Root mean square deviation of bond lengths and bond angles.

<sup>8</sup>Chen et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica D66:12-21.

<sup>9</sup>wwPDB Validation Server.

**Cellular Thermal Shift Assay.** Cellular thermal shift assay was performed similarly as described previously.<sup>16</sup> The Cellular Thermal Shift Assay was performed according to the reported method<sup>16, 72</sup>. H2170 cells ( $5 \times 10^5$  per sample) were treated with a compound or

with DMSO for 1 h, washed with PBS three times, and dissolved in 50  $\mu$ l PBS supplemented with a protease inhibitor, followed by heating at indicated temperatures in a Mastercycler gradient (Eppendorf, New York, USA). Treated cells were then subjected to snap-freezing in liquid nitrogen and thawed on ice for 3 cycles. The protein levels of DCN1 in equal amounts of the supernatant were examined by western blots. GAPDH was used as the loading control.

**Co-immunoprecipitation Assay.** The immunoprecipitation assay was performed as described previously.<sup>101</sup> Briefly, to immunoprecipitate endogenous DCN1, the pre-cleared cell lysates were incubated with DCN1 antibody for 3 h, followed by incubation with protein A&G beads (Santa Cruz) at 4 °C overnight with rotation. The beads were washed three times with lysis buffer, and the immunoprecipitation complexes were subjected to SDS-PAGE.

**Western blot analysis.** For direct analysis, cells were lysed in a Triton X-100 or RIPA buffer with phosphatase inhibitors. The antibodies used were as follows: UBE2M (Santa Cruz), CUL1 (Santa Cruz), CUL-2 (BD Bioscience), CUL3 (Cell Signal), CUL4A (Cell Signal), CUL4B (Proteintech), CUL5 (Santa Cruz), NRF2 (Santa Cruz), CDT-1 (Santa Cruz), DCN1 (Genway) and  $\beta$ -actin (Sigma).

## **Appendix B: Experimental Section of Chapter 3**

### **Peptide Synthesis**

All the peptides in Table 1 and Table 2 were synthesized using Fmoc solid phase peptide synthesis (SPPS). The synthesis was conducted on the ABI 433A peptide synthesizer. Rink amide resin was used as the solid support. The TFA/TES/H<sub>2</sub>O (36:1:2) cleavage cocktail was applied to cleave the peptide from resin. After removal of solvents by evaporation, the peptide was purified by RP-HPLC.

**Chemistry. General Information.** All reagents and solvents were used as supplied without further purification and the reactions were performed under a nitrogen atmosphere in anhydrous solvents. The final products were purified by a reverse phase HPLC (RP-HPLC) with solvent A (0.1% of TFA in H<sub>2</sub>O) and solvent B (0.1% of TFA in CH<sub>3</sub>CN) as eluents. The purity was determined by Waters ACQUITY UPLC and all the final compounds were > 95% pure. <sup>1</sup>HNMR spectra of the synthetic compounds were acquired at a proton frequency of 400 MHz and chemical shifts are reported in parts per million (ppm) relative to an internal standard. High resolution mass spectra (HRMS) were obtained with an Agilent Q-TOF Electrospray mass spectrometer and low resolution mass spectral analyses were performed on a Thermo-Scientific LCQ Fleet mass spectrometer.

### **General Procedure for the Synthesis of Compounds 1-7.**

Step 1. A solution of compound 37 (2.43 g, 5 mmol, 1 equiv), an amine (5 mmol, 1 equiv), HATU (3.8g, 10 mmol, 2 equiv), and DIPEA (2.6 ml, 15 mmol, 3 equiv) in DMF (25 ml)

was stirred at rt for 3 h. After the reaction was complete, the solution was diluted with EtOAc and washed with saturated NaHCO<sub>3</sub>, 1.0 M HCl and brine then dried over Na<sub>2</sub>SO<sub>4</sub>. The organic solvent was removed under vacuum, the residue was treated with diethylamine in DCM (1:9) for 1 h. The reaction mixture was evaporated, and the residue was dissolved in DCM (25 mL) and treated with propionic anhydride (1.3 ml, 15 mmol, 3 equiv) and DIPEA (2.6 ml, 15 mmol, 3 equiv) for 30 min. The solvent was evaporated, and the crude product was purified by flash chromatography on silica gel to afford the common intermediate (39).

Step 2. Compound 39 (0.1 mmol, 1 equiv), boronic acid (0.2 mmol, 2 equiv), K<sub>2</sub>CO<sub>3</sub> (0.3 mmol, 3 equiv) were dissolved in dioxane (1ml). The flask was sealed and charged with nitrogen gas, then a catalyst, tetrakis(triphenylphosphine)palladium(0) (0.1 equiv) was quickly added. The flask was heated in oil bath at 120 °C and stirred for 2 h. After the solution cooled to rt, it was diluted with EtOAc and washed with saturated NaHCO<sub>3</sub>, 1.0 M HCl and brine and then dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was evaporated, and the crude product was purified by flash chromatography on silica gel to afford compound 1-7.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-**

**2-propionamidopropanamide (1):** <sup>1</sup>H NMR (400 MHz, MeOD) δ 7.54 - 7.40 (m, 3H), 7.39 - 7.29 (m, 3H), 7.28 - 7.15 (m, 3H), 4.70 (dd, J = 7.8, 5.2 Hz, 1H), 4.42-4.39 (m, 1H), 3.45 (dd, J = 15.1, 5.2 Hz, 1H), 3.31 - 3.23 (m, 1H), 3.18 (dd, J = 13.0, 3.4 Hz, 1H), 3.07 (dd, J = 15.8, 7.6 Hz, 1H), 2.98 - 2.84 (m, 2H), 2.43 (s, 3H), 2.26 (q, J = 7.6 Hz, 2H), 1.10 (t, J = 7.6 Hz, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 485.17, found = 485.36.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(4-methyl-5-(m-tolyl)thiazol-2-yl)-2-**

**propionamidopropanamide (2):** <sup>1</sup>H NMR (400 MHz, MeOH-*d*<sub>4</sub>) δ 7.33 (dt, J = 10.5, 7.4

Hz, 3H), 7.28 – 7.20 (m, 6H), 4.69 (dd,  $J = 7.8, 5.1$  Hz, 1H), 4.45 – 4.36 (m, 1H), 3.45 (ddd,  $J = 15.1, 5.2, 1.8$  Hz, 1H), 3.31 – 3.25 (m, 1H), 3.18 (dd,  $J = 13.0, 3.7$  Hz, 1H), 3.11 (s, 1H), 2.91 (dd,  $J = 7.5, 3.6$  Hz, 2H), 2.40 (d,  $J = 9.1$  Hz, 6H), 2.26 (q,  $J = 7.6$  Hz, 2H), 1.10 (t,  $J = 7.6$  Hz, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 465.22, found = 465.17.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(4-methyl-5-(3-**

**(trifluoromethyl)phenyl)thiazol-2-yl)-2-propionamidopropanamide (3):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.72 (q,  $J = 5.1, 4.6$  Hz, 4H), 7.35 – 7.21 (m, 5H), 4.71 (dd,  $J = 7.9, 5.2$  Hz, 1H), 4.45 – 4.34 (m, 1H), 3.46 (dd,  $J = 15.1, 5.2$  Hz, 1H), 3.32 – 3.26 (m, 1H), 3.18 (dd,  $J = 13.1, 3.7$  Hz, 1H), 3.08 (dd,  $J = 13.0, 10.3$  Hz, 1H), 2.97 – 2.85 (m, 2H), 2.44 (s, 3H), 2.31 – 2.21 (m, 2H), 1.11 (t,  $J = 7.6$  Hz, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 519.20, found = 519.11.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(4-methyl-5-phenylthiazol-2-yl)-2-**

**propionamidopropanamide (4):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.51 – 7.37 (m, 5H), 7.35 – 7.21 (m, 5H), 4.69 (dd,  $J = 7.8, 5.2$  Hz, 1H), 4.41 (d,  $J = 7.5$  Hz, 1H), 3.45 (dd,  $J = 15.1, 5.1$  Hz, 1H), 3.27 (dd,  $J = 15.8, 7.7$  Hz, 1H), 3.21 – 3.15 (m, 1H), 3.13 – 3.02 (m, 1H), 2.98 – 2.84 (m, 2H), 2.47 – 2.38 (m, 3H), 2.26 (q,  $J = 7.6$  Hz, 2H), 1.10 (t,  $J = 7.6$  Hz, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 451.21, found = 451.32.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(2-chlorophenyl)-4-methylthiazol-2-yl)-**

**2-propionamidopropanamide (5):**  $^1\text{H}$  NMR (400 MHz, MeOD)  $\delta$  7.59 – 7.53 (m, 1H), 7.48 – 7.37 (m, 3H), 7.35 – 7.19 (m, 5H), 4.59 (dd,  $J = 8.5, 5.8$  Hz, 1H), 4.49 – 4.32 (m, 1H), 3.22 – 3.09 (m, 3H), 3.00 (dd,  $J = 12.9, 10.6$  Hz, 1H), 2.92 (dd,  $J = 13.9, 6.4$  Hz, 1H), 2.82 (dd,  $J = 13.9, 8.7$  Hz, 1H), 2.34 – 2.16 (m, 5H), 1.10 (t,  $J = 7.6$  Hz, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 485.17, found = 485.32.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(2-chlorophenyl)-4-methylthiazol-2-yl)-2-propionamidopropanamide (6):**  $^1\text{H}$  NMR (400 MHz, MeOD)  $\delta$  7.92 – 7.78 (m, 2H), 7.68 – 7.59 (m, 2H), 7.37 – 7.28 (m, 2H), 7.28 – 7.18 (m, 3H), 4.71 (dd,  $J$  = 8.1, 5.1 Hz, 1H), 4.41-4.38 (m, 1H), 3.46 (dd,  $J$  = 15.2, 5.1 Hz, 1H), 3.31 – 3.25 (m, 1H), 3.18 (dd,  $J$  = 13.1, 3.7 Hz, 1H), 3.11 – 3.02 (m, 1H), 2.92-2.90 (m, 2H), 2.47 (s, 3H), 2.25 (q,  $J$  = 7.5 Hz, 2H), 1.10 (t,  $J$  = 7.6 Hz, 3H). ESI-MS  $m/z$   $[\text{M} + \text{H}]^+$  calculated = 485.17, found = 485.36.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-cyclopentyl-4-methylthiazol-2-yl)-2-propionamidopropanamide (7):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.46 – 7.12 (m, 5H), 4.64 (dd,  $J$  = 8.1, 5.1 Hz, 1H), 4.38 (d,  $J$  = 8.4 Hz, 1H), 3.45 (dd,  $J$  = 15.1, 5.1 Hz, 1H), 3.25 (dd,  $J$  = 15.1, 8.1 Hz, 1H), 3.16 (dd,  $J$  = 13.0, 3.8 Hz, 1H), 3.05 (dd,  $J$  = 13.0, 10.3 Hz, 1H), 2.89 (dt,  $J$  = 10.4, 4.5 Hz, 2H), 2.33 (d,  $J$  = 0.9 Hz, 2H), 2.23 (qd,  $J$  = 7.4, 2.4 Hz, 2H), 2.05 – 1.81 (m, 4H), 1.78 (d,  $J$  = 12.8 Hz, 1H), 1.46 (dt,  $J$  = 15.8, 11.1 Hz, 2H), 1.40 – 1.21 (m, 3H), 1.08 (td,  $J$  = 7.6, 1.0 Hz, 3H). ESI-MS  $m/z$   $[\text{M} + \text{H}]^+$  calculated = 443.24, found = 443.36.

#### General Procedure for the Synthesis of Compounds 8-10.

An acyl chloride (0.2 mmol, 2 equiv) and DIPEA (0.3 mmol, 3 equiv) were added to a solution of compound **40** (52.7 mg, 0.1 mmol, 1 equiv) in DCM and stirred for 1 h at rt. The reaction mixture was concentrated and treated with TFA (3 ml) in DCM (10 ml) at rt for 1 h. This reaction mixture was concentrated and purified by RP-HPLC to provide the compounds **8-10** in 45-76% yield over two steps.

**(2S)-2-Acetamido-N-(1-amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)propanamide (8):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.49 – 7.36 (m,



4H), 7.35 – 7.29 (m, 2H), 7.29 – 7.22 (m, 3H), 4.70 (dd,  $J = 7.8, 5.2$  Hz, 1H), 4.40 (td,  $J = 10.9, 10.4, 4.9$  Hz, 1H), 3.44 (dd,  $J = 15.2, 5.2$  Hz, 1H), 3.28 (ddd,  $J = 15.1, 7.8, 1.4$  Hz, 1H), 3.18 (dd,  $J = 13.0, 3.7$  Hz, 1H), 3.08 (dd,  $J = 13.2, 10.3$  Hz, 1H), 2.97 – 2.85 (m, 2H), 2.43 (s, 3H), 1.99 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 471.15, found = 471.26.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-isobutyramidopropanamide (9):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.50 – 7.34 (m, 4H), 7.33 – 7.20 (m, 5H), 4.70 (dd,  $J = 7.8, 5.2$  Hz, 1H), 4.40 (ddd,  $J = 10.9, 7.3, 3.6$  Hz, 1H), 3.46 (dd,  $J = 15.1, 5.2$  Hz, 1H), 3.29 (dd,  $J = 7.9, 1.2$  Hz, 1H), 3.18 (dd,  $J = 13.0, 3.7$  Hz, 1H), 3.09 (t,  $J = 11.8$  Hz, 1H), 2.92 (d,  $J = 7.4$  Hz, 2H), 2.50 (p,  $J = 6.8$  Hz, 1H), 2.42 (s, 3H), 1.09 (t,  $J = 6.8$  Hz, 6H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 499.19, found = 499.02.

**N-((2S)-1-((1-Amino-3-phenylpropan-2-yl)amino)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-1-oxopropan-2-yl)benzamide (10):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.88 – 7.83 (m, 2H), 7.63 – 7.58 (m, 1H), 7.54 – 7.50 (m, 2H), 7.46 – 7.39 (m, 3H), 7.35 (dt,  $J = 7.1, 1.7$  Hz, 1H), 7.24 – 7.14 (m, 5H), 4.47 – 4.36 (m, 1H), 3.56 (dd,  $J = 15.1, 5.2$  Hz, 1H), 3.46 (dd,  $J = 15.1, 8.2$  Hz, 1H), 3.19 (dd,  $J = 13.0, 3.7$  Hz, 1H), 3.13 – 3.06 (m, 1H), 2.97 – 2.82 (m, 3H), 2.43 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 533.17, found = 533.21.

#### General Procedure for the Synthesis of Compounds 11-19.

A solution of compound **40** (52.7 mg, 0.1 mmol, 1 equiv.), aryl iodide (0.1 mmol, 1 equiv.), copper iodide (20 mol%), 2-isobutyrylcyclohexanone (30 mol%),  $\text{Cs}_2\text{CO}_3$  (0.1 mmol, 2 equiv.) in DMF in a round bottom flask was sealed and was stirred in oil bath at

80 °C for 8 h. After cooling down to rt, the mixture was diluted with EtOAc and washed with saturated NaHCO<sub>3</sub>, 1.0 M HCl and brine and then dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was evaporated, and the crude product was purified by purified by RP-HPLC to provide compounds **11-19**.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-(phenylamino)propanamide (11):** <sup>1</sup>H NMR (400 MHz, MeOH-*d*<sub>4</sub>) δ 7.55 – 7.36 (m, 4H), 7.36 – 7.16 (m, 6H), 7.17 – 7.07 (m, 2H), 6.72 (t, *J* = 7.3 Hz, 1H), 6.63 – 6.59 (m, 1H), 4.58 (s, 1H), 4.41 (s, 1H), 3.50 (d, *J* = 2.3 Hz, 1H), 3.18 – 3.12 (m, 1H), 3.04 (d, *J* = 11.5 Hz, 1H), 2.96 – 2.81 (m, 2H), 2.42 (s, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 505.18, found = 505.53.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((3-chlorophenyl)amino)propanamide (12):** <sup>1</sup>H NMR (400 MHz, MeOH-*d*<sub>4</sub>) δ 7.47 – 7.37 (m, 3H), 7.33 (dq, *J* = 7.3, 1.6 Hz, 1H), 7.31 – 7.19 (m, 5H), 7.05 (t, *J* = 8.0 Hz, 1H), 6.69 – 6.64 (m, 2H), 6.51 – 6.46 (m, 1H), 4.46 – 4.37 (m, 1H), 4.34 (ddd, *J* = 6.5, 5.0, 1.3 Hz, 1H), 3.49 (dd, *J* = 14.9, 5.0 Hz, 1H), 3.32 – 3.26 (m, 1H), 3.17 (dd, *J* = 13.0, 4.1 Hz, 1H), 3.06 (dd, *J* = 12.8, 9.5 Hz, 1H), 2.90 (qd, *J* = 14.0, 7.3 Hz, 2H), 2.41 (s, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 539.14, found = 539.20.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((2-chlorophenyl)amino)propanamide (13):** <sup>1</sup>H NMR (400 MHz, MeOH-*d*<sub>4</sub>) δ 7.47 – 7.38 (m, 3H), 7.33 (dt, *J* = 7.3, 1.7 Hz, 1H), 7.30 – 7.22 (m, 5H), 7.21 (dd, *J* = 6.7, 2.0 Hz, 1H), 7.12 – 7.06 (m, 1H), 6.70 (t, *J* = 7.6 Hz, 1H), 6.58 (d, *J* = 8.1 Hz, 1H), 4.45 (q, *J* = 7.7, 5.4 Hz, 2H), 3.54 (dd, *J* = 14.8, 4.8 Hz, 1H), 3.43 (dd, *J* = 14.8, 6.2 Hz, 1H), 3.20

(dd,  $J = 13.1, 3.8$  Hz, 1H), 3.08 (t,  $J = 11.3$  Hz, 1H), 3.00 – 2.81 (m, 2H), 2.41 (s, 3H).

ESI-MS  $m/z$   $[M + H]^+$  calculated = 539.14, found = 539.21.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((2-chlorophenyl)amino)propanamide (14):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.48 – 7.38 (m, 3H), 7.34 (dt,  $J = 7.2, 1.6$  Hz, 1H), 7.31 – 7.21 (m, 5H), 7.10 – 7.00 (m, 2H), 6.56 – 6.47 (m, 2H), 4.41 (d,  $J = 6.2$  Hz, 1H), 4.27 (dd,  $J = 7.3, 5.1$  Hz, 1H), 3.52 – 3.44 (m, 1H), 3.22 – 3.14 (m, 1H), 3.11 – 3.01 (m, 2H), 2.95 (dd,  $J = 14.0, 5.9$  Hz, 1H), 2.85 (dd,  $J = 14.1, 9.0$  Hz, 1H), 2.42 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 539.14, found = 539.22.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((3-fluorophenyl)amino)propanamide (15):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.46 – 7.39 (m, 3H), 7.34 (dt,  $J = 7.2, 1.7$  Hz, 1H), 7.31 – 7.22 (m, 5H), 7.08 (td,  $J = 8.3, 6.7$  Hz, 1H), 6.38 (tdd,  $J = 11.7, 6.4, 2.3$  Hz, 3H), 4.42 (td,  $J = 7.5, 5.9, 2.8$  Hz, 1H), 4.32 (dd,  $J = 7.3, 5.1$  Hz, 1H), 3.52 – 3.46 (m, 1H), 3.17 (dd,  $J = 13.0, 4.0$  Hz, 1H), 3.06 (d,  $J = 10.7$  Hz, 1H), 2.98 – 2.92 (m, 1H), 2.93 – 2.80 (m, 2H), 2.42 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 523.17, found = 523.45.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((2-fluorophenyl)amino)propanamide (16):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.50 – 7.37 (m, 3H), 7.33 (dq,  $J = 7.2, 1.5$  Hz, 1H), 7.31 – 7.15 (m, 5H), 7.05 – 6.87 (m, 2H), 6.70 (tdd,  $J = 7.8, 4.9, 1.5$  Hz, 1H), 6.61 (td,  $J = 8.4, 1.5$  Hz, 1H), 4.49 – 4.34 (m, 2H), 3.52 (dd,  $J = 14.9, 5.0$  Hz, 1H), 3.39 (dd,  $J = 14.8, 6.9$  Hz, 1H), 3.18 (dd,  $J = 13.1, 4.0$  Hz, 1H), 3.07 (dd,  $J = 12.9, 9.7$  Hz, 1H), 3.00 – 2.79 (m, 2H), 2.41 (d,  $J = 0.9$  Hz, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 523.17, found = 523.41.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((4-fluorophenyl)amino)propanamide (17):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.48 – 7.38 (m, 3H), 7.33 (dt,  $J = 7.3, 1.7$  Hz, 1H), 7.29 – 7.16 (m, 5H), 7.03 – 6.90 (m, 2H), 6.73 – 6.66 (m, 1H), 6.60 (td,  $J = 8.4, 1.5$  Hz, 1H), 4.50 – 4.33 (m, 2H), 3.52 (dd,  $J = 14.9, 5.0$  Hz, 1H), 3.39 (dd,  $J = 14.9, 6.9$  Hz, 1H), 3.18 (dd,  $J = 13.0, 4.0$  Hz, 1H), 3.07 (dd,  $J = 13.0, 9.7$  Hz, 1H), 2.99 – 2.81 (m, 2H), 2.41 (s, 3H). ESI-MS  $m/z$   $[\text{M} + \text{H}]^+$  calculated = 523.17, found = 523.38.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((2,3-difluorophenyl)amino)propanamide (18):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.47 – 7.38 (m, 3H), 7.35 (dt,  $J = 7.3, 1.7$  Hz, 1H), 7.30 – 7.18 (m, 5H), 6.95 (dt,  $J = 10.6, 9.0$  Hz, 1H), 6.50 (ddd,  $J = 13.0, 6.7, 2.8$  Hz, 1H), 6.30 (ddd,  $J = 9.0, 3.3, 1.5$  Hz, 1H), 4.42 (tdt,  $J = 8.2, 5.7, 2.4$  Hz, 1H), 4.29 – 4.22 (m, 1H), 3.48 (dd,  $J = 15.0, 5.1$  Hz, 1H), 3.18 (dd,  $J = 13.0, 4.1$  Hz, 1H), 3.12 – 3.01 (m, 1H), 3.00 – 2.80 (m, 2H), 2.42 (s, 3H). ESI-MS  $m/z$   $[\text{M} + \text{H}]^+$  calculated = 541.16, found = 541.37.

**(2S)-N-(1-Amino-3-phenylpropan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)propanamide (19):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.48 – 7.39 (m, 3H), 7.35 (dd,  $J = 7.2, 1.8$  Hz, 1H), 7.30 – 7.18 (m, 5H), 6.24 – 6.14 (m, 3H), 4.47 – 4.38 (m, 1H), 4.32 (ddd,  $J = 6.6, 5.1, 1.3$  Hz, 1H), 3.49 (dd,  $J = 15.0, 5.1$  Hz, 1H), 3.35 (s, 1H), 3.18 (dd,  $J = 13.0, 4.0$  Hz, 1H), 3.06 (dd,  $J = 13.0, 9.4$  Hz, 1H), 2.98 – 2.82 (m, 2H), 2.42 (s, 3H). ESI-MS  $m/z$   $[\text{M} + \text{H}]^+$  calculated = 541.16, found = 541.39.

#### General Procedure for the Synthesis of Compounds 20-34.

A solution of compound **42** (40 mg, 0.1 mmol, 1 equiv), an amine (0.12 mmol, 1.2 equiv), HATU (76 mg, 0.2 mmol, 2 equiv), and DIPEA (26 ml, 0.2 mmol, 2 equiv) in

DMF (25 ml) was stirred at rt for 3 h. The reaction mixture was diluted with EtOAc and washed with saturated NaHCO<sub>3</sub>, 1.0 M HCl and brine and then dried over with Na<sub>2</sub>SO<sub>4</sub>. The solvent was evaporated, and the crude product was purified by purified by RP-HPLC to provide the compounds **20-34**.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-((S)-1-(dimethylamino)-3-(p-tolyl)propan-2-yl)propanamide (20):** <sup>1</sup>H NMR (400 MHz, MeOH-*d*<sub>4</sub>) δ 7.46 – 7.38 (m, 3H), 7.35 – 7.19 (m, 6H), 6.19 (dd, *J* = 10.2, 7.3 Hz, 3H), 4.57 (s, 1H), 4.39 – 4.32 (m, 1H), 3.42 (dd, *J* = 15.1, 3.9 Hz, 2H), 3.28 – 3.14 (m, 2H), 2.93 (s, 7H), 2.85 (dd, *J* = 13.9, 8.2 Hz, 1H), 2.41 (s, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 569.19, found = 583.12.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-((S)-1-(dimethylamino)propan-2-yl)propanamide (21):** <sup>1</sup>H NMR (400 MHz, Chloroform-*d*) δ 9.10 (s, 1H), 7.86 (d, *J* = 8.2 Hz, 1H), 7.52 – 7.41 (m, 2H), 7.34 (dt, *J* = 10.4, 1.9 Hz, 1H), 7.25 (ddt, *J* = 10.7, 7.1, 1.8 Hz, 1H), 6.20 (t, *J* = 9.5 Hz, 3H), 4.67 – 4.36 (m, 2H), 4.20 (d, *J* = 15.1 Hz, 1H), 3.60 (t, *J* = 12.6 Hz, 1H), 3.46 (t, *J* = 12.2 Hz, 1H), 3.06 (s, 1H), 2.98 (s, 6H), 2.51 (d, *J* = 17.0 Hz, 3H), 1.23 (dd, *J* = 14.0, 6.5 Hz, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 493.16, found = 493.28.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-((S)-1-(dimethylamino)-4-methylpentan-2-yl)propanamide (22):** <sup>1</sup>H NMR (400 MHz, Chloroform-*d*) δ 10.56 (s, 1H), 8.30 (d, *J* = 8.9 Hz, 1H), 7.43 (dd, *J* = 4.9, 1.8 Hz, 2H), 7.36 (d, *J* = 1.9 Hz, 1H), 7.26 (dt, *J* = 5.7, 2.0 Hz, 1H), 6.21 (dd, *J* = 9.7, 6.7 Hz, 3H), 4.56 (d, *J* = 8.9 Hz, 1H), 4.43 (s, 1H), 4.04 (d, *J* = 14.8 Hz, 1H), 3.61 – 3.41 (m, 2H), 2.96 (d, *J* = 13.9 Hz, 6H), 2.88 (d, *J* = 12.1 Hz, 1H), 2.52 (s, 3H), 1.53 (dd, *J* = 27.2,

14.9 Hz, 2H), 1.27 – 1.14 (m, 1H), 0.86 (dd,  $J = 17.8, 6.4$  Hz, 6H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 535.20, found = 535.11.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-N-((S)-1-cyclohexyl-2-**

**(dimethylamino)-ethyl)-2-((3,5-difluorophenyl)amino)propanamide (23):**  $^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  10.67 (s, 1H), 8.56 (d,  $J = 9.3$  Hz, 1H), 7.45 – 7.34 (m, 3H), 7.27 (ddd,  $J = 4.9, 3.8, 1.8$  Hz, 1H), 6.34 – 6.14 (m, 3H), 4.56 (d,  $J = 8.4$  Hz, 1H), 4.23 (s, 1H), 3.99 (d,  $J = 14.8$  Hz, 1H), 3.73 (td,  $J = 6.5, 3.3$  Hz, 1H), 3.55 (q,  $J = 13.5$  Hz, 2H), 3.15 (dt,  $J = 7.4, 3.8$  Hz, 1H), 2.90 (d,  $J = 14.9$  Hz, 6H), 2.51 (s, 3H), 1.71 (d,  $J = 12.9$  Hz, 2H), 1.62 (d,  $J = 11.4$  Hz, 3H), 1.35 (t,  $J = 7.3$  Hz, 1H), 1.23 – 0.98 (m, 4H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 561.22, found = 561.35.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-N-((S)-1-cyclohexyl-3-**

**(dimethylamino)-propan-2-yl)-2-((3,5-difluorophenyl)amino)propanamide (24):**  $^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  11.25 (s, 1H), 8.35 (s, 1H), 7.46 – 7.36 (m, 3H), 7.27 (q,  $J = 3.0$  Hz, 1H), 6.36 – 6.15 (m, 3H), 4.47 (d,  $J = 19.9$  Hz, 2H), 3.88 (d,  $J = 14.9$  Hz, 1H), 3.64 – 3.39 (m, 2H), 2.90 (s, 7H), 2.51 (s, 3H), 1.76 (d,  $J = 12.5$  Hz, 1H), 1.62 (d,  $J = 13.4$  Hz, 5H), 1.28 (s, 1H), 1.08 (dq,  $J = 27.6, 13.3, 12.4$  Hz, 4H), 0.91 (d,  $J = 8.3$  Hz, 1H), 0.78 (q,  $J = 11.3$  Hz, 1H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 575.23, found = 575.12.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-N-((S)-1-cyclohexyl-3-**

**(dimethylamino)-propan-2-yl)-2-((3,5-difluorophenyl)amino)propanamide (25):**  $^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  11.00 (s, 1H), 8.67 (d,  $J = 8.6$  Hz, 1H), 7.39 – 7.36 (m, 2H), 7.35 (dt,  $J = 2.3, 1.1$  Hz, 2H), 7.28 – 7.22 (m, 2H), 7.21 – 7.15 (m, 2H), 6.23 – 6.10 (m, 3H), 4.62 (s, 1H), 4.33 (dd,  $J = 7.6, 4.3$  Hz, 1H), 4.00 (d,  $J = 11.7$  Hz, 1H), 3.74 (ddt,

$J = 10.3, 6.7, 3.7$  Hz, 1H), 3.70 – 3.61 (m, 1H), 3.58 – 3.46 (m, 2H), 3.38 (dd,  $J = 15.2, 7.6$  Hz, 1H), 3.14 (qd,  $J = 7.5, 4.3$  Hz, 1H), 3.07 (dd,  $J = 7.1, 3.9$  Hz, 1H), 2.90 (dd,  $J = 13.0, 3.7$  Hz, 2H), 2.84 (s, 2H), 2.45 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 603.15, found = 603.20.

**(S)-N-((S)-1-(3-Chlorophenyl)-3-(dimethylamino)propan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)propanamide (26):**  $^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  11.26 (s, 1H), 8.81 (d,  $J = 8.5$  Hz, 1H), 7.42 – 7.39 (m, 2H), 7.38 – 7.35 (m, 1H), 7.28 – 7.24 (m, 1H), 7.24 – 7.20 (m, 3H), 7.08 (d,  $J = 5.4$  Hz, 1H), 6.21 (t,  $J = 9.0$  Hz, 1H), 6.14 (d,  $J = 8.3$  Hz, 2H), 4.55 (s, 1H), 4.44 (d,  $J = 7.3$  Hz, 1H), 3.89 (d,  $J = 14.8$  Hz, 1H), 3.59 (t,  $J = 12.1$  Hz, 1H), 3.44 (dd,  $J = 15.2, 8.7$  Hz, 1H), 3.00 (dd,  $J = 13.8, 6.6$  Hz, 1H), 2.81 (dd,  $J = 29.9, 14.9$  Hz, 8H), 2.51 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 603.15, found = 603.28.

**(S)-N-((S)-1-(4-Chlorophenyl)-3-(dimethylamino)propan-2-yl)-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)propanamide (27):**  $^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  11.58 (s, 1H), 8.84 (d,  $J = 8.3$  Hz, 1H), 7.38 (td,  $J = 4.1, 1.4$  Hz, 3H), 7.28 – 7.20 (m, 3H), 7.13 (d,  $J = 8.0$  Hz, 2H), 6.23 – 6.07 (m, 3H), 4.52 (s, 1H), 4.40 (d,  $J = 7.5$  Hz, 1H), 3.75 (d,  $J = 14.9$  Hz, 1H), 3.56 (t,  $J = 11.9$  Hz, 1H), 3.41 (dd,  $J = 15.2, 7.9$  Hz, 1H), 2.99 (dd,  $J = 13.9, 6.7$  Hz, 1H), 2.91 – 2.62 (m, 8H), 2.48 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 603.15, found = 603.25.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-((S)-1-(dimethylamino)-3-(4-fluorophenyl)propan-2-yl)propanamide (28):**  $^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  10.83 (s, 1H), 8.76 (s, 1H), 7.42 (d,  $J = 5.7$  Hz, 2H), 7.36 (s, 1H), 7.26 (d,  $J = 5.6$  Hz, 1H), 7.14 (s, 2H), 6.95 (t,  $J = 7.4$  Hz, 2H), 6.30 – 6.02 (m, 3H),

4.52 (d,  $J = 37.2$  Hz, 2H), 4.10 – 3.92 (m, 1H), 3.72 – 3.51 (m, 1H), 3.49 – 3.34 (m, 1H), 3.10 – 2.61 (m, 9H), 2.52 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 587.18, found = 587.38.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-((S)-1-(dimethylamino)-3-(4-(trifluoromethyl)phenyl)propan-2-yl)propanamide (29):**

$^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  11.06 (s, 1H), 8.91 (s, 1H), 7.53 (s, 2H), 7.42 (s, 2H), 7.34 (dd,  $J = 12.9, 7.0$  Hz, 3H), 7.26 (s, 1H), 6.28 – 6.07 (m, 3H), 4.70 (s, 2H), 4.01 (dd,  $J = 27.9, 10.2$  Hz, 1H), 3.67 (s, 1H), 3.52 – 3.35 (m, 1H), 3.17 – 3.03 (m, 1H), 3.01 – 2.65 (m, 8H), 2.52 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 637.17, found = 637.28.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-((S)-1-(dimethylamino)-3-(*p*-tolyl)propan-2-yl)propanamide (30):**  $^1\text{H}$  NMR (400 MHz,

Chloroform- $d$ )  $\delta$  11.59 (s, 1H), 8.68 (s, 1H), 7.38 (d,  $J = 5.3$  Hz, 3H), 7.27 (t,  $J = 3.9$  Hz, 1H), 7.09 (q,  $J = 7.6$  Hz, 4H), 6.26 – 6.05 (m, 3H), 4.46 (d,  $J = 46.6$  Hz, 2H), 3.75 (d,  $J = 14.1$  Hz, 1H), 3.61 – 3.32 (m, 2H), 2.99 (s, 1H), 2.80 (d,  $J = 22.4$  Hz, 8H), 2.48 (s, 3H), 2.32 (s, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 583.20, found = 583.35.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-**

**((S)-1-(dimethylamino)-3-(4-ethylphenyl)propan-2-yl)propanamide (31):**  $^1\text{H}$  NMR (400 MHz, Chloroform- $d$ )  $\delta$  11.04 (s, 1H), 8.65 (s, 1H), 7.45 – 7.39 (m, 2H), 7.36 (s, 1H), 7.28 – 7.22 (m, 1H), 7.11 (q,  $J = 7.7$  Hz, 4H), 6.19 (d,  $J = 11.8$  Hz, 3H), 4.51 (d,  $J = 30.7$  Hz, 2H), 3.92 (s, 1H), 3.45 (s, 2H), 2.98 (s, 1H), 2.84 (d,  $J = 20.1$  Hz, 8H), 2.62 (q,  $J = 7.6$  Hz, 2H), 2.50 (s, 3H), 1.22 (t,  $J = 7.6$  Hz, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 597.22, found = 597.10.



**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-((S)-1-(dimethylamino)-3-(4-hydroxyphenyl)propan-2-yl)propanamide (32):** <sup>1</sup>H NMR (400 MHz, Chloroform-*d*) δ 10.67 (s, 1H), 8.35 (d, *J* = 8.6 Hz, 1H), 7.86 (s, 1H), 7.44 – 7.35 (m, 3H), 7.28 – 7.24 (m, 1H), 6.97 (dd, *J* = 27.1, 8.4 Hz, 2H), 6.77 – 6.70 (m, 2H), 6.25 – 6.14 (m, 3H), 4.47 (dd, *J* = 9.4, 4.1 Hz, 2H), 3.94 (d, *J* = 15.5 Hz, 1H), 3.54 – 3.39 (m, 3H), 2.87 (d, *J* = 12.8 Hz, 5H), 2.81 (s, 3H), 2.50 (d, *J* = 12.9 Hz, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 585.18, found = 585.27.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-((S)-1-(2,3-dihydro-1H-inden-2-yl)-2-(dimethylamino)ethyl)propanamide (33):** <sup>1</sup>H NMR (400 MHz, Chloroform-*d*) δ 10.84 (s, 1H), 8.66 (d, *J* = 9.3 Hz, 1H), 7.45 – 7.35 (m, 3H), 7.27 (d, *J* = 5.4 Hz, 1H), 7.18 – 7.08 (m, 4H), 6.28 – 6.13 (m, 3H), 4.55 (d, *J* = 8.1 Hz, 2H), 3.99 (dd, *J* = 15.4, 3.2 Hz, 1H), 3.62 (t, *J* = 12.3 Hz, 1H), 3.50 (dd, *J* = 15.3, 9.0 Hz, 1H), 2.97 (d, *J* = 21.3 Hz, 9H), 2.79 – 2.61 (m, 3H), 2.51 (s, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 595.20, found = 595.33.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)-N-((S)-1-(dimethylamino)-3-(1H-indol-3-yl)propan-2-yl)propanamide (34):** <sup>1</sup>H NMR (400 MHz, Chloroform-*d*) δ 10.63 (s, 1H), 8.54 (d, *J* = 8.6 Hz, 1H), 8.28 (s, 1H), 7.44 – 7.36 (m, 3H), 7.36 – 7.32 (m, 1H), 7.26 – 7.21 (m, 2H), 7.18 – 7.08 (m, 3H), 6.23 – 6.13 (m, 3H), 4.70 (s, 1H), 4.54 (d, *J* = 9.0 Hz, 1H), 4.02 (d, *J* = 15.0 Hz, 1H), 3.63 – 3.40 (m, 2H), 3.15 (dd, *J* = 14.6, 5.8 Hz, 1H), 3.02 – 2.90 (m, 2H), 2.82 (d, *J* = 17.1 Hz, 6H), 2.49 (d, *J* = 11.9 Hz, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 608.20, found = 608.39.

**tert-Butyl (S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-amino-4-thioxobutanoate (36):** Step 1: Fmoc-Asp-OtBu (10.6 g, 27.3 mmol) was dissolved in

DMF then ammonium carbonate (4.3 g, 54.6 mmol) and carbonyldiimidazole (13.2 g, 81.9 mmol) were added. The mixture was stirred at rt and monitored by TLC. After the reaction was complete, the mixture was poured into saturated aqueous NaHCO<sub>3</sub> and extract with DCM. The organic layer was washed with brine and then the solvent was evaporated for the next reaction.

Step 2: The residues were dissolved in THF (500 mL), Lawesson reagent (11.56 g, 28.6 mmol) was added, and the mixture was stirred at rt for 18 h. The reaction mixture was then poured into saturated aqueous NaHCO<sub>3</sub> (300 ml), then extracted with EtOAc (2 x 100 ml). The organic fractions were combined, dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated. The crude product was purified by flash column chromatography. <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 9.56 (s, 1H), 9.18 (s, 1H), 7.90 (dd, *J* = 7.5, 2.7 Hz, 2H), 7.72 (dd, *J* = 7.2, 3.6 Hz, 2H), 7.69 – 7.64 (m, 1H), 7.43 (td, *J* = 7.5, 2.7 Hz, 2H), 7.37 – 7.29 (m, 2H), 4.61 (dq, *J* = 8.3, 2.7 Hz, 1H), 4.35 – 4.18 (m, 3H), 2.83 (dddd, *J* = 60.8, 14.5, 7.3, 2.7 Hz, 2H), 1.38 (d, *J* = 2.8 Hz, 9H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 427.16, found = 427.29.

**(S)-2-(((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-3-(5-bromo-4-methylthiazol-2-yl)propanoic acid (37):** A solution of compound **36** (408 mg, 1 mmol) and chloroacetone (227mg, 3 mmol) in DMF was refluxed for 3 h. The reaction mixture was evaporated under vacuum and the residue was dissolved in EtOAc. The solution was washed with saturated NaHCO<sub>3</sub> solution and brine, dried over Na<sub>2</sub>SO<sub>4</sub>, then concentrated for the next reaction. The residue was redissolved in THF and NBS (356mg, 2 mmol) was added then the mixture was stirred for 5 h at rt. The reaction

mixture was then poured into saturated aqueous NaHCO<sub>3</sub>. The mixture was extracted with EtOAc. The organic fractions were combined, dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated. The crude was purified by flash column chromatography. ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 487.02, found = 487.20.

**(S)-2-(((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-3-(5-(3-chlorophenyl)-4-**

**methylthiazol-2-yl)propanoic acid (38):** Compound **37** (1 mmol, 1 equiv), 3-chlorophenylboronic acid (2 mmol, 2 eq), K<sub>2</sub>CO<sub>3</sub> (3 mmol, 3 eq) were dissolved in dioxane. The flask was sealed and charged with nitrogen gas, then tetrakis(triphenylphosphine)palladium(0) (0.1 equiv) was quickly added to the flask. The flask was heated in oil bath at a temperature of 120 °C and stirred for 2 h. After the solution cooled to rt, the solution was diluted with EtOAc and washed with saturated NaHCO<sub>3</sub>, 1.0 M HCl and brine and then dried over with Na<sub>2</sub>SO<sub>4</sub>. The solvent was evaporated, and the crude product was purified by flash chromatography on silica gel to afford compound **38**. ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 519.11, found = 519.28.

**tert-Butyl ((S)-2-((S)-3-(5-bromo-4-methylthiazol-2-yl)-2-**

**propionamidopropanamido)-3-phenylpropyl)carbamate (39):** <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ 8.07 (d, *J* = 8.6 Hz, 1H), 7.92 (d, *J* = 8.4 Hz, 1H), 7.31 – 7.10 (m, 5H), 6.82 (t, *J* = 5.9 Hz, 1H), 4.55 (td, *J* = 8.8, 4.8 Hz, 1H), 3.94 (dq, *J* = 12.1, 6.7, 5.3 Hz, 1H), 3.27 (dd, *J* = 14.9, 5.0 Hz, 1H), 3.05 (dd, *J* = 14.6, 8.5 Hz, 2H), 2.99 – 2.89 (m, 1H), 2.75 – 2.69 (m, 1H), 2.63 (dd, *J* = 13.8, 8.1 Hz, 1H), 2.27 (s, 2H), 2.22 (q, *J* = 7.5 Hz, 1H), 2.13 – 2.05 (m, 2H), 1.38 (s, 9H), 0.95 (t, *J* = 7.6 Hz, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 553.14, found = 553.35.

***tert*-Butyl ((S)-2-((S)-2-amino-3-(5-(3-chlorophenyl)-4-methylthiazol-2-**

**yl)propanamido)-3-phenylpropyl)carbamate (40):**  $^1\text{H}$  NMR (400 MHz, MeOH- $d_4$ )  $\delta$  7.45 (ddd,  $J$  = 16.6, 9.9, 5.2 Hz, 4H), 7.36 – 7.21 (m, 5H), 4.24 (s, 2H), 3.62 (dd,  $J$  = 16.6, 4.3 Hz, 1H), 3.46 – 3.40 (m, 1H), 3.23 (q,  $J$  = 7.1 Hz, 1H), 3.14 (dd,  $J$  = 14.1, 7.0 Hz, 1H), 2.90 (dd,  $J$  = 14.0, 5.7 Hz, 1H), 2.81 (dd,  $J$  = 14.1, 8.5 Hz, 1H), 2.49 (s, 3H), 1.44 (s, 9H). ESI-MS  $m/z$   $[\text{M} + \text{H}]^+$  calculated = 529.20, found = 529.33.

**Allyl (S)-2-amino-3-(5-(3-chlorophenyl)-4-methylthiazol-2-yl)propanoate (41):** ESI-MS  $m/z$   $[\text{M} + \text{H}]^+$  calculated = 337.07, found = 337.21.

**(S)-3-(5-(3-Chlorophenyl)-4-methylthiazol-2-yl)-2-((3,5-difluorophenyl)amino)propanoic acid (41):** ESI-MS  $m/z$   $[\text{M} + \text{H}]^+$  calculated = 409.05, found = 409.14.

### **Biolayer Interferometry (BLI)**

The biolayer interferometry (BLI) assay developed for DCN1 was conducted as described previously.<sup>75, 76</sup> A BLI assay for DCN3 was developed to test new synthesized peptides and screen our DCN1 library as described previously. The BLI kinetic experiments were performed using the OctetRED96 instrument from PALL/ForteBio. All assays were run at 30 °C, sample plates were continuously shaken at 1000 RPM to avoid any mass transport effect. PBS (pH 7.4) was used as assay buffer, in which 0.1% BSA and 0.01% Tween-20 were added to reduce non-specific interactions. 1% DMSO was introduced to increase compound solubility. Assays were run in Greiner 96-well blackflat-bottom microplates in which protein solutions, pure assay buffer for dissociation, and serial dilutions of a tested compound were loaded. The biotinylation of purified recombinant DCN3 protein was performed using EZ-Link long-chain biotinylation reagent (Thermo) to allow tethering on

Super Streptavidin (SSA) biosensors. A 1:1 molar ratio of protein and biotinylation reagent was applied to avoid overbiotinylation of the proteins. The reaction mixture of DCN3 protein and biotinylation reagent in PBS was incubated at 4 °C for 2 h to allow the reaction to complete. The mixture was then dialyzed to remove unreacted biotinylation reagent, using Fisher Scientific 10K MWCO dialysis cassettes.

Biotinylated DCN3 protein was tethered to super streptavidin (SSA) biosensors (ForteBio) by dipping the sensors into 10 µg/mL protein solutions. The sensor saturation response level of 9-10 nm was achieved in 20 min. Biotinylated blocked streptavidin (SAV B4) sensors were prepared following the protocol provided by the manufacturer and were used as the inactive reference controls. Sensors with tethered proteins were washed in assay buffer for 10 min to eliminate loose nonspecific bound protein molecules and establish a stable baseline before starting association–dissociation cycles of the test compound. Association and dissociation times of 10 min were used in the assay. A DMSO-only reference was included in all assays. Raw kinetic data collected were processed in the data analysis software provided by the manufacturer using double reference subtraction in which both DMSO-only reference and inactive protein reference were subtracted. The resulting data were analyzed based on 1:1 binding model from which  $k_{on}$  and  $k_{off}$  values were obtained, and then  $K_D$  values were calculated. All the BLI experiments were performed in at least two independent experiments.

### **Fluorescence Polarization (FP) Competitive Binding Assays**

The fluorescence polarization (FP) competitive binding assays for accurately testing newly designed DCN1 inhibitors were performed as described previously.<sup>75, 76</sup> In this

case, we developed a new FP competitive binding assay to determine the binding affinity of all synthesized DCN3 inhibitors. A novel FAM tethered fluorescent probe compound (DCN3-Flu1) was designed and synthesized based on the optimized UBE2F peptide. Equilibrium dissociation constant ( $K_D$ ) values of DCN3-Flu1 to DCN3 were determined from protein saturation experiments by monitoring the total FP values of mixtures composed with the fluorescent probe at a fixed concentration and proteins with increasing concentrations up to full saturation. Serial dilutions of proteins were mixed with DCN3-Flu1 to a final volume of 200  $\mu$ L in the assay buffer (100 mM phosphate buffer, pH = 6.5, with 0.02% Tween-20 and 1% DMSO). The final probe concentration was 5 nM. Plates were incubated at rt for 30 min with gentle shaking to ensure equilibration. FP values in millipolarization units (mP) were measured using the Infinite M-1000 plate reader (Tecan U.S., Research Triangle Park, NC) in Microfluor 1 96-well, black round-bottom plates (Thermo Scientific, Waltham, MA) at an excitation wavelength of 485 nm and an emission wavelength of 530 nm. The  $K_D$  values of DCN3-Flu1 were calculated by fitting the sigmoidal dose-dependent FP increases to a function of protein concentration using Graphpad Prism 7.0 software (Graphpad Software, San Diego, CA).

The  $IC_{50}$  and  $K_i$  values of compounds were determined in competitive binding experiments. Mixtures of 4  $\mu$ L of compounds in DMSO and 196  $\mu$ L of preincubated protein/probe complex solution in the assay buffer were added into assay plates which were incubated at rt for 30 min with gentle shaking. Final concentrations of DCN3 protein and fluorescent probe DCN3-Flu1 were 400 and 5 nM, respectively. Negative controls containing protein/probe complex only (equivalent to 0% inhibition) and positive controls containing only free probes (equivalent to 100% inhibition) were included in each assay

plate. FP values were measured as described above.  $IC_{50}$  values were determined by nonlinear regression fitting of the competition curves.  $K_i$  values of competitive inhibitors were obtained directly by nonlinear regression fitting, based upon the  $K_D$  values of the probe DCN3-Flu1 and concentrations of the protein and probe in the competitive assays.  $K_i$  values were calculated using the methods described previously.<sup>102</sup> All the FP competitive experiments were performed in duplicate in three independent experiments.

**Computational Modeling.** Coordinates of DCN3 protein in the co-crystal structure of DCN3 and the NEDD8 E2 peptide (PDB ID 4GBA<sup>82</sup>) were extracted to build the binding models of designed inhibitors with DCN3. To accommodate our designed compounds, the rotamer state of F142 in DCN3 was modeled based on F117 in the structure of the DCN1/DI-591 complex (PDB ID 5UFI<sup>16</sup>) and the side chain atoms of the modeled DCN3 structure were minimized while the backbone atoms were constrained using default parameters in the MOE program.<sup>80</sup> Hydrogen atoms in DCN3 were added using the “protonate 3D” module in MOE in the physiological conditions. In the docking calculations, the structures of the designed compounds were prepared using MOE and then subject to the docking simulations using default parameters in the GOLD program (version 5.6.2).<sup>95</sup> <sup>96</sup> The center of the binding site was set at M139 in DCN3 and the radius of the binding site was 13 Å. The docking poses were assessed by the GoldScore fitness function and the top-ranked poses were analyzed to determine the predicted docking modes.

### **The Cellular Thermal Shift Assay**

Cellular Thermal Shift Assay was performed according to the reported method.<sup>103</sup> Briefly, cells ( $5 \times 10^5$  per sample) were treated with a compound as indicated or with DMSO for

1 h, washed with PBS three times, and dissolved in 50  $\mu$ l PBS supplemented with a protease inhibitor, followed by heating at the indicated temperatures in a Mastercycler gradient (Eppendorf, New York, USA). Treated cells were then subjected to snap-freezing in liquid nitrogen and thawed on ice for 3 cycles. The protein levels of DCN1 and DCN3 in equal amounts of the supernatant were examined by western blots. GAPDH was used as the control.

### **Western Blotting Analysis**

Treated cells were lysed by RIPA buffer supplemented with protease inhibitors. The expression levels of indicated proteins were examined by western blotting analysis. GAPDH was used as the loading control. Anti-Cullin 1 (sc-11384) and anti-Cullin 2 (sc-10781) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA); anti-Cullin 4A (PA5-14542), anti-Cullin 4B (PA5-35239) and anti-DCN3 antibodies (DCUN1D3, PA5-44000) were from ThermoFisher Scientific (Wayne, MI); anti-Cullin 3 (2759) antibodies from Cell Signaling Technology (Boston, MA); anti-DCN1 (GWB-E3D700) antibody from GenWay Biotech (San Diego, CA). Anti-Cullin 5 antibody (A302-173A) from Bethyl Laboratories (Montgomery, TX).



## **Appendix C: Experimental Section of Chapter 4**

**In vitro histone methyltransferase (HMT) assay.** The HMT assay was performed as described previously (Dou et al., 2005). Small molecule inhibitors were tested in reaction buffer consisting of 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 10% glycerol. Equal stoichiometry of human MLL1 SET domain, WDR5, RbBP5 and ASH2L were mixed to form the active MLL1 complex at a final concentration of 125 nM. MLL1 complex were mixed with small molecules inhibitors at a final concentration of 10uM and incubated on ice. After 10 minutes the methyltransferase reaction was initiated by adding 1  $\mu$ M of 3H-SAM and 4  $\mu$ M of histone H3 (1–21) peptide. The reactions proceeded for 1 h at 22 °C and was terminated by adding 10  $\mu$ L of reaction mixture onto Whatman P81 ion exchange filter paper. The filter paper was air dried and washed in three 10-min washes in 50 mM sodium bicarbonate pH 9.0 to remove excess SAM. Filter paper was heat-dried for 20 min and placed in 10 mL Ultima Gold scintillation fluid for 3H signal acquisition in the unit of counts per minute (c.p.m.).

### **Alpha LISA assay**

The method for Alpha LISA assay developed by our group has been described.<sup>86</sup>

### **Preparation of tert-butyl (S)-4-((((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methyl)amino)-2-((tert-butoxycarbonyl)amino)butanoate (20)**

Sodium triacetoxyborohydride (1.5 eq) was added to a solution of 5'-amino-5'-deoxy-2', 3'-o-(1-methylethylidene)-adenosine (0.9 eq) and *t*-butyl (S)-2-[(tert-butoxycarbonyl)

amino]-4-oxo-butanoate (1 eq) in DCE and stirred at rt for 4 h. Then saturated sodium bicarbonate solution was added and stirring was continued for 20 min. The reaction mixture was extracted with DCM and the combined organic layers were dried over MgSO<sub>4</sub> and concentrated under reduced pressure. The compound was purified by preparative HPLC.

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.41 (s, 1H), 8.40 (s, 1H), 6.37 (d, *J* = 2.2 Hz, 1H), 5.42 (dd, *J* = 6.3, 2.2 Hz, 1H), 5.19 (dd, *J* = 6.3, 3.9 Hz, 1H), 4.55 (dt, *J* = 9.8, 3.5 Hz, 1H), 4.04 (dd, *J* = 9.3, 5.0 Hz, 1H), 3.61 (dd, *J* = 13.1, 9.8 Hz, 1H), 3.51 (dd, *J* = 13.1, 3.3 Hz, 1H), 3.08 (t, *J* = 7.7 Hz, 2H), 2.14 (dt, *J* = 13.3, 4.7 Hz, 1H), 1.96 – 1.83 (m, 1H), 1.64 (s, 3H), 1.46 (s, 10H), 1.43 (s, 8H), 1.42 – 1.40 (m, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 564.31, found = 564.22.

**Preparation of tert-butyl (S)-4-((((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methyl)(azetidino-3-ylmethyl)amino)-2-((tert-butoxycarbonyl)amino)butanoate (21)**

Benzyl 3-formylazetidino-1-carboxylate (1.2 eq) and sodium triacetoxyborohydride (1.5 eq) in DCE were added to a solution of 5'-deoxy-5'-[[(3S)-4-(1,1-dimethylethoxy)-3-[[[(1,1-dimethylethoxy)carbonyl]amino]-4-oxobutyl]amino]-2',3'-O-(1-methylethylidene)adenosine (1 eq) and stirred for 3 h at rt. Then saturated sodium bicarbonate solution was added and stirring was continued for 20 min. The reaction mixture was extracted with DCM and the combined organic layers were dried over MgSO<sub>4</sub> and concentrated under reduced pressure. Then the crude product was dissolved in MeOH, Pd/C (0.05%) was added and the mixture was stirred under a hydrogen atmosphere. The black particles were filtered off and washed with MeOH several times.

The filtrate was then concentrated under reduced pressure and purified by preparative HPLC.

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.52 – 8.39 (m, 2H), 6.38 (d,  $J$  = 2.1 Hz, 1H), 5.36 (d,  $J$  = 6.3 Hz, 1H), 5.13 (d,  $J$  = 5.5 Hz, 1H), 4.64 (s, 1H), 4.25 – 4.08 (m, 2H), 3.97 (s, 3H), 3.60 (s, 2H), 3.54 – 3.36 (m, 3H), 3.25 (s, 1H), 3.18 – 2.92 (m, 2H), 2.13 (d,  $J$  = 63.5 Hz, 2H), 1.81 (s, 1H), 1.64 (d,  $J$  = 6.2 Hz, 2H), 1.52 – 1.50 (m, 1H), 1.45 (d,  $J$  = 1.9 Hz, 15H), 1.40 (s, 3H), 1.32 (d,  $J$  = 9.8 Hz, 1H). ESI-MS  $m/z$   $[\text{M} + \text{H}]^+$  calculated = 633.36, found = 633.28.

**Synthesis of compound 11-19.** Sodium triacetoxyborohydride (1.5 eq) was added to a solution of *t*-butyl (S)-4-((((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methyl)(azetidin-3-ylmethyl)amino)-2-((tert-butoxycarbonyl)amino)butanoate (1 eq) and the aldehyde (1.2 eq) in DCE and stirred for 2 h at rt. Then saturated sodium bicarbonate solution was added and stirring was continued for 20 min. The reaction mixture was extracted with DCM. The combined organic layers dried over  $\text{MgSO}_4$  and concentrated under reduced pressure. Then the crude product was dissolved in water/trifluoroacetic acid solution (1:4) and stirred for 16 h. The solvent was removed *in vacuo* and the compounds were purified directly by preparative HPLC.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(1-(3-cyclopentylpropyl)azetidin-3-yl)amino)butanoic acid (1)**

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.36 (s, 1H), 8.34 (s, 1H), 6.02 (dd,  $J$  = 3.9, 1.4 Hz, 1H), 4.70 (t,  $J$  = 4.6 Hz, 1H), 4.29 (t,  $J$  = 5.9 Hz, 1H), 4.21 (s, 2H), 4.07 (t,  $J$  = 6.1 Hz, 1H),

3.93 (s, 3H), 3.16 – 3.08 (m, 3H), 2.83 (t,  $J = 6.6$  Hz, 2H), 2.10 (dq,  $J = 31.9, 7.2$  Hz, 2H), 1.77 (dt,  $J = 14.4, 8.8$  Hz, 4H), 1.70 – 1.44 (m, 8H), 1.19 – 1.03 (m, 3H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 533.31, found = 533.22.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(2-aminoethyl)amino)butanoic acid (2)**

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.45 (s, 1H), 8.43 (s, 1H), 6.07 (d,  $J = 3.6$  Hz, 1H), 4.65 (dd,  $J = 5.5, 3.7$  Hz, 1H), 4.32 (t,  $J = 5.9$  Hz, 1H), 4.26 (d,  $J = 7.4$  Hz, 1H), 4.05 (t,  $J = 6.3$  Hz, 1H), 3.07 (q,  $J = 4.6$  Hz, 4H), 2.99 – 2.85 (m, 4H), 2.17 (td,  $J = 17.3, 14.8, 7.4$  Hz, 1H), 2.05 (s, 1H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 411.20, found = 411.12.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(3-((3-cyclopentylpropyl)amino)cyclobutyl)amino)butanoic acid (3)**

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.38 (s, 1H), 8.37 (s, 1H), 6.08 (d,  $J = 3.6$  Hz, 1H), 4.71 (dd,  $J = 5.4, 3.6$  Hz, 1H), 4.44 (t,  $J = 5.8$  Hz, 1H), 4.40 – 4.33 (m, 1H), 4.20 – 4.09 (m, 1H), 4.01 (dd,  $J = 7.8, 5.1$  Hz, 1H), 3.80 (tdd,  $J = 7.4, 4.2, 1.7$  Hz, 1H), 3.54 (dd,  $J = 14.5, 9.6$  Hz, 1H), 3.43 (d,  $J = 14.1$  Hz, 1H), 3.28 (t,  $J = 6.9$  Hz, 2H), 2.95 – 2.88 (m, 2H), 2.84 – 2.73 (m, 2H), 2.66 – 2.48 (m, 2H), 2.35 – 2.27 (m, 1H), 2.19 – 2.09 (m, 1H), 1.87 – 1.78 (m, 3H), 1.76 – 1.57 (m, 6H), 1.41 (dt,  $J = 11.1, 6.3$  Hz, 2H), 1.19 – 1.08 (m, 2H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 547.33, found = 547.45.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-butylazetidin-3-yl)methyl)amino)butanoic acid (4)**

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.45 (s, 1H), 8.42 (d, *J* = 5.4 Hz, 1H), 6.11 (d, *J* = 3.8 Hz, 1H), 4.79 (dt, *J* = 27.8, 4.9 Hz, 1H), 4.68 (t, *J* = 4.1 Hz, 1H), 4.42 (d, *J* = 5.2 Hz, 2H), 4.29 (ddd, *J* = 17.4, 9.5, 4.7 Hz, 1H), 4.20 – 4.08 (m, 1H), 4.03 (dt, *J* = 7.7, 4.0 Hz, 1H), 3.98 – 3.83 (m, 1H), 3.59 – 3.38 (m, 5H), 3.37 (d, *J* = 8.0 Hz, 1H), 3.29 (t, *J* = 8.7 Hz, 1H), 3.17 (d, *J* = 21.3 Hz, 2H), 2.37 (dq, *J* = 15.0, 7.5 Hz, 1H), 2.17 (dd, *J* = 14.5, 6.2 Hz, 1H), 1.58 – 1.45 (m, 2H), 1.38 (q, *J* = 7.9, 7.3 Hz, 2H), 0.97 (t, *J* = 7.2 Hz, 3H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 493.28, found = 493.11.

**(S)-2-amino-4-(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(furan-3-ylmethyl)azetidin-3-yl)methyl)amino)butanoic acid (5)**

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.36 (dd, *J* = 8.8, 3.8 Hz, 2H), 7.73 (s, 1H), 7.64 – 7.57 (m, 1H), 6.54 (d, *J* = 1.8 Hz, 1H), 6.07 (t, *J* = 3.7 Hz, 1H), 4.69 (dd, *J* = 5.2, 3.7 Hz, 1H), 4.39 (qd, *J* = 9.5, 8.5, 4.3 Hz, 3H), 4.23 (s, 3H), 4.16 (d, *J* = 25.5 Hz, 2H), 3.97 (dd, *J* = 7.8, 4.9 Hz, 3H), 3.27 – 3.14 (m, 3H), 2.28 (dd, *J* = 15.8, 8.2 Hz, 1H), 2.11 – 2.01 (m, 1H), 1.42 (d, *J* = 4.9 Hz, 2H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 517.24, found = 517.15.

**(S)-2-amino-4-(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(3-cyclopentylpropyl)azetidin-3-yl)methyl)amino)butanoic acid (6)**

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.37 (s, 1H), 8.35 (s, 1H), 6.08 (d, *J* = 3.9 Hz, 1H), 4.71 (t, *J* = 4.5 Hz, 1H), 4.38 (qd, *J* = 7.3, 4.8, 3.9 Hz, 2H), 4.26 (s, 1H), 4.17 – 4.03 (m, 1H), 3.96 (dd, *J* = 7.9, 4.8 Hz, 1H), 3.86 (d, *J* = 22.2 Hz, 1H), 3.36 (d, *J* = 9.1 Hz, 3H), 3.32 – 3.04 (m, 6H), 2.30 (h, *J* = 7.3 Hz, 1H), 2.06 (dq, *J* = 15.1, 5.2 Hz, 1H), 1.85 – 1.73

(m, 3H), 1.69 – 1.60 (m, 2H), 1.59 – 1.48 (m, 4H), 1.34 (q,  $J = 7.2$  Hz, 2H), 1.18 – 1.05 (m, 2H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 547.33, found = 547.45.

**(S)-2-amino-4-(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(3-phenylpropyl)azetidino-3-yl)methyl)amino)butanoic acid (7)**

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.32 (d,  $J = 2.8$  Hz, 2H), 7.34 – 7.28 (m, 2H), 7.24 – 7.19 (m, 3H), 6.08 – 6.03 (m, 1H), 4.71 (dd,  $J = 5.3, 3.7$  Hz, 1H), 4.39 (t,  $J = 5.8$  Hz, 1H), 4.33 (s, 2H), 4.13 – 4.02 (m, 1H), 3.95 (t,  $J = 6.3$  Hz, 2H), 3.27 (s, 4H), 3.14 (s, 5H), 2.68 (t,  $J = 7.5$  Hz, 2H), 2.25 (d,  $J = 14.3$  Hz, 1H), 2.03 (s, 1H), 1.84 (q,  $J = 7.9$  Hz, 2H), 1.44 – 1.29 (m, 1H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 555.30, found = 555.22.

**(S)-2-amino-4-(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(3-((3-phenylpropyl)amino)propyl)amino)butanoic acid (8)**

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.27 (d,  $J = 4.1$  Hz, 2H), 7.32 – 7.26 (m, 2H), 7.19 (dd,  $J = 12.6, 7.1$  Hz, 3H), 6.04 (d,  $J = 3.8$  Hz, 1H), 4.78 – 4.73 (m, 1H), 4.46 – 4.42 (m, 1H), 3.85 (dd,  $J = 8.9, 3.8$  Hz, 1H), 3.50 (dt,  $J = 3.3, 1.6$  Hz, 1H), 3.42 – 3.39 (m, 1H), 3.30 – 3.25 (m, 1H), 3.15 (dt,  $J = 3.3, 1.6$  Hz, 1H), 2.98 – 2.92 (m, 2H), 2.67 (t,  $J = 7.0$  Hz, 2H), 2.23 – 2.19 (m, 1H), 2.09 – 1.95 (m, 5H), 1.67 (s, 5H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 543.30, found = 543.21.

**(S)-2-amino-4-(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(3-((3-phenylpropyl)amino)methyl)cyclobutyl)amino)butanoic acid (9)**

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.29 (d, *J* = 2.7 Hz, 2H), 7.29 (t, *J* = 7.5 Hz, 2H), 7.25 – 7.14 (m, 3H), 6.06 (d, *J* = 3.2 Hz, 1H), 4.70 (t, *J* = 4.3 Hz, 1H), 4.51 (t, *J* = 5.9 Hz, 1H), 4.39 (t, *J* = 8.2 Hz, 1H), 3.85 (d, *J* = 6.4 Hz, 2H), 3.64 (t, *J* = 12.2 Hz, 2H), 3.09 (s, 1H), 3.07 (s, 1H), 3.02 – 2.93 (m, 2H), 2.70 (t, *J* = 7.2 Hz, 2H), 2.61 – 2.54 (m, 2H), 2.24 – 2.16 (m, 1H), 2.04 (ddt, *J* = 23.9, 19.0, 10.1 Hz, 4H), 1.71 (q, *J* = 5.4, 3.3 Hz, 4H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 569.31, found = 569.49.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(3-((3-phenylpropyl)amino)methyl)cyclobutyl)amino)butanoic acid (10)**

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.30 (s, 2H), 7.79 (s, 1H), 7.29 – 7.19 (m, 3H), 7.16 – 7.08 (m, 2H), 6.06 (d, *J* = 4.1 Hz, 1H), 4.76 – 4.71 (m, 1H), 4.71 – 4.62 (m, 2H), 4.47 – 4.41 (m, 1H), 4.34 (t, *J* = 5.4 Hz, 1H), 4.22 (s, 2H), 3.87 (t, *J* = 6.4 Hz, 1H), 3.27 (d, *J* = 11.6 Hz, 1H), 3.21 (t, *J* = 6.9 Hz, 3H), 3.17 – 3.13 (m, 1H), 2.35 – 2.25 (m, 1H), 2.10 – 2.02 (m, 1H), 1.42 – 1.36 (m, 1H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 553.26, found = 553.19.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-benzylazetidino-3-yl)methyl)amino)butanoic acid (11)**

This title compound was prepared following General Procedure A using benzaldehyde (6.0 mg, 0.056 mmol) and 5 (30 mg, 0.047 mmol). <sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.34 – 8.19 (m, 2H), 7.57 – 7.37 (m, 5H), 6.01 (d, *J* = 3.8 Hz, 1H), 4.80 – 4.70 (m, 1H), 4.44 – 4.29 (m, 3H), 4.30 – 4.01 (m, 4H), 3.94 (dt, *J* = 24.7, 9.0 Hz, 2H), 3.80 (s, 1H), 3.22 –

3.13 (m, 1H), 3.11 – 2.83 (m, 5H), 2.23 – 2.11 (m, 1H), 1.97 (dt,  $J = 14.8, 6.0$  Hz, 1H).

ESI-MS  $m/z$   $[M + H]^+$  calculated = 527.27, found = 527.15.

**(S)-2-amino-4-(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(3-chlorobenzyl)azetidin-3-yl)methyl)amino)butanoic acid (12)**

This title compound was prepared following General Procedure A using benzaldehyde (7.8 mg, 0.056 mmol) and 5 (30 mg, 0.047 mmol)  $^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.43 (s, 1H), 8.40 (s, 1H), 7.54 – 7.36 (m, 4H), 6.10 (d,  $J = 3.7$  Hz, 1H), 4.67 (dd,  $J = 4.8, 3.7$  Hz, 1H), 4.44 – 4.34 (m, 4H), 4.29 – 4.14 (m, 2H), 4.12 – 3.96 (m, 3H), 3.44 (tdd,  $J = 20.2, 9.2, 6.4$  Hz, 5H), 3.26 (dq,  $J = 13.2, 6.8, 6.2$  Hz, 2H), 2.34 (dq,  $J = 14.9, 7.4$  Hz, 1H), 2.13 (dq,  $J = 14.7, 5.8$  Hz, 1H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 561.23, found = 561.12.

**(S)-2-amino-4-(((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(2-chlorobenzyl)azetidin-3-yl)methyl)amino)butanoic acid (13)**

$^1\text{H}$  NMR (400 MHz, Methanol- $d_4$ )  $\delta$  8.41 (s, 1H), 8.38 (s, 1H), 7.52 (ddd,  $J = 7.4, 3.0, 1.6$  Hz, 2H), 7.42 (dtd,  $J = 23.7, 7.4, 1.5$  Hz, 2H), 6.07 (d,  $J = 3.6$  Hz, 1H), 4.62 (dd,  $J = 4.9, 3.5$  Hz, 1H), 4.54 (s, 2H), 4.36 (dd,  $J = 5.7, 3.1$  Hz, 2H), 4.32 – 4.27 (m, 1H), 4.22 (td,  $J = 8.8, 7.8, 4.9$  Hz, 1H), 4.11 (ddd,  $J = 19.8, 10.6, 7.3$  Hz, 2H), 3.99 (dd,  $J = 7.7, 5.2$  Hz, 1H), 3.47 – 3.34 (m, 5H), 3.25 (dd,  $J = 13.8, 6.9$  Hz, 2H), 2.30 (dt,  $J = 14.9, 7.4$  Hz, 1H), 2.13 (ddd,  $J = 11.8, 9.7, 6.0$  Hz, 1H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 561.23, found = 561.12.



**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(4-chlorobenzyl)azetidin-3-yl)methyl)amino)butanoic acid (14)**

This title compound was prepared following General Procedure A using 2-chlorobenzaldehyde (11.6 mg, 0.064 mmol) and 5 (30 mg, 0.069 mmol) <sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.43 (s, 1H), 8.40 (s, 1H), 7.49 – 7.43 (m, 4H), 6.10 (d, *J* = 3.7 Hz, 1H), 4.67 (dd, *J* = 4.9, 3.6 Hz, 1H), 4.43 – 4.34 (m, 4H), 4.26 – 4.20 (m, 1H), 4.16 (td, *J* = 8.5, 7.6, 4.5 Hz, 1H), 4.10 – 3.98 (m, 3H), 3.42 (ddd, *J* = 20.3, 9.0, 4.5 Hz, 5H), 3.24 (dq, *J* = 13.3, 6.4 Hz, 2H), 2.33 (dd, *J* = 14.9, 7.4 Hz, 1H), 2.16 – 2.08 (m, 1H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 561.23, found = 561.10.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(3,3-diphenylpropyl)azetidin-3-yl)methyl)amino)butanoic acid (18)**

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.39 (s, 1H), 8.37 (s, 1H), 7.32 – 7.26 (m, 8H), 7.20 (ddd, *J* = 8.6, 5.6, 2.3 Hz, 2H), 6.07 (d, *J* = 3.7 Hz, 1H), 4.65 (dd, *J* = 4.9, 3.7 Hz, 1H), 4.42 – 3.71 (m, 9H), 3.41 – 3.33 (m, 3H), 3.27 – 3.06 (m, 5H), 2.28 (dq, *J* = 13.8, 7.5 Hz, 3H), 2.06 (dt, *J* = 15.3, 5.4 Hz, 1H). ESI-MS *m/z* [M + H]<sup>+</sup> calculated = 631.33, found = 631.20.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(benzofuran-3-ylmethyl)azetidin-3-yl)methyl)amino)butanoic acid (17)**

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>) δ 8.34 (s, 1H), 8.32 (s, 1H), 7.66 (dt, *J* = 7.8, 1.0 Hz, 1H), 7.53 (dd, *J* = 8.3, 1.0 Hz, 1H), 7.39 (ddd, *J* = 8.5, 7.2, 1.4 Hz, 1H), 7.30 (td, *J* = 7.5,

1.0 Hz, 1H), 7.06 (s, 1H), 6.05 (d, J = 3.9 Hz, 1H), 4.70 (dd, J = 5.2, 3.8 Hz, 1H), 4.60 (s, 2H), 4.40 – 4.22 (m, 5H), 4.14 (dd, J = 10.8, 7.1 Hz, 1H), 4.08 (dd, J = 10.7, 7.0 Hz, 1H), 3.94 (dd, J = 7.6, 5.1 Hz, 1H), 3.31 – 3.20 (m, 4H), 3.17 – 3.11 (m, 2H), 2.25 (dt, J = 14.8, 7.5 Hz, 1H), 2.08 – 1.98 (m, 1H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 567.26, found = 567.14.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(thiophen-2-ylmethyl)azetidin-3-yl)methyl)amino)butanoic acid (16)**

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>)  $\delta$  8.33 (d, J = 5.3 Hz, 2H), 7.60 (dd, J = 5.1, 1.2 Hz, 1H), 7.30 – 7.23 (m, 1H), 7.13 (dd, J = 5.1, 3.5 Hz, 1H), 6.05 (d, J = 3.7 Hz, 1H), 4.69 (dd, J = 5.4, 3.7 Hz, 1H), 4.57 (s, 2H), 4.38 (t, J = 5.9 Hz, 1H), 4.35 – 4.29 (m, 1H), 4.23 (t, J = 8.9 Hz, 1H), 4.14 (t, J = 9.3 Hz, 1H), 3.99 (dt, J = 22.6, 7.7 Hz, 3H), 3.30 – 3.08 (m, 7H), 2.25 (dd, J = 14.9, 7.3 Hz, 1H), 2.08 – 1.96 (m, 1H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 533.22, found = 533.13.

**(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(thiophen-2-ylmethyl)azetidin-3-yl)methyl)amino)butanoic acid (15)**

<sup>1</sup>H NMR (400 MHz, Methanol-*d*<sub>4</sub>)  $\delta$  8.34 (s, 1H), 8.32 (s, 1H), 7.62 (s, 1H), 7.57 (dd, J = 5.0, 3.1 Hz, 1H), 7.16 (d, J = 5.2 Hz, 1H), 6.05 (d, J = 3.9 Hz, 1H), 4.70 (dd, J = 5.2, 3.5 Hz, 1H), 4.48 – 3.89 (m, 9H), 3.31 – 3.07 (m, 6H), 2.35 – 2.19 (m, 1H), 2.09 – 1.97 (m, 1H), 1.32 (s, 1H). ESI-MS  $m/z$   $[M + H]^+$  calculated = 533.22, found = 533.10.

## Bibliography

1. Hershko, A.; Ciechanover, A. The ubiquitin system. *Annu Rev Biochem* **1998**, *67*, 425-79.
2. Hershko, A. The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. *Cell Death Differ* **2005**, *12*, 1191-7.
3. Watson, I. R.; Irwin, M. S.; Ohh, M. NEDD8 pathways in cancer, Sine Quibus Non. *Cancer Cell* **2011**, *19*, 168-76.
4. Cappadocia, L.; Lima, C. D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem Rev* **2018**, *118*, 889-918.
5. Enchev, R. I.; Schulman, B. A.; Peter, M. Protein neddylation: beyond cullin-RING ligases. *Nat Rev Mol Cell Biol* **2015**, *16*, 30-44.
6. Petroski, M. D.; Deshaies, R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nat Rev Mol Cell Biol* **2005**, *6*, 9-20.
7. Nakayama, K. I.; Nakayama, K. Ubiquitin ligases: cell-cycle control and cancer. *Nat Rev Cancer* **2006**, *6*, 369-81.
8. Soucy, T. A.; Smith, P. G.; Milhollen, M. A.; Berger, A. J.; Gavin, J. M.; Adhikari, S.; Brownell, J. E.; Burke, K. E.; Cardin, D. P.; Critchley, S.; Cullis, C. A.; Doucette, A.; Garnsey, J. J.; Gaulin, J. L.; Gershman, R. E.; Lublinsky, A. R.; McDonald, A.; Mizutani, H.; Narayanan, U.; Olhava, E. J.; Peluso, S.; Rezaei, M.; Sintchak, M. D.; Talreja, T.; Thomas, M. P.; Traore, T.; Vyskocil, S.; Weatherhead, G. S.; Yu, J.; Zhang, J.; Dick, L.

- R.; Claiborne, C. F.; Rolfe, M.; Bolen, J. B.; Langston, S. P. An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **2009**, 458, 732-6.
9. Scott, D. C.; Sviderskiy, V. O.; Monda, J. K.; Lydeard, J. R.; Cho, S. E.; Harper, J. W.; Schulman, B. A. Structure of a RING E3 trapped in action reveals ligation mechanism for the ubiquitin-like protein NEDD8. *Cell* **2014**, 157, 1671-84.
10. Chairatvit, K.; Ngamkitidechakul, C. Control of cell proliferation via elevated NEDD8 conjugation in oral squamous cell carcinoma. *Mol Cell Biochem* **2007**, 306, 163-9.
11. Salon, C.; Brambilla, E.; Brambilla, C.; Lantuejoul, S.; Gazzeri, S.; Eymin, B. Altered pattern of Cul-1 protein expression and neddylation in human lung tumours: relationships with CAND1 and cyclin E protein levels. *J Pathol* **2007**, 213, 303-10.
12. Wang, X.; Li, L.; Liang, Y.; Li, C.; Zhao, H.; Ye, D.; Sun, M.; Jeong, L. S.; Feng, Y.; Fu, S.; Jia, L.; Guo, X. Targeting the neddylation pathway to suppress the growth of prostate cancer cells: therapeutic implication for the men's cancer. *Biomed Res Int* **2014**, 2014, 974309.
13. Tanaka, T.; Nakatani, T.; Kamitani, T. Inhibition of NEDD8-conjugation pathway by novel molecules: potential approaches to anticancer therapy. *Mol Oncol* **2012**, 6, 267-75.
14. Soucy, T. A.; Smith, P. G.; Rolfe, M. Targeting NEDD8-activated cullin-RING ligases for the treatment of cancer. *Clin Cancer Res* **2009**, 15, 3912-6.
15. Scott, D. C.; Monda, J. K.; Bennett, E. J.; Harper, J. W.; Schulman, B. A. N-Terminal Acetylation Acts as an Avidity Enhancer Within an Interconnected Multiprotein Complex. *Science* **2011**, 334, 674-678.

16. Zhou, H.; Lu, J.; Liu, L.; Bernard, D.; Yang, C.-Y.; Fernandez-Salas, E.; Chinnaswamy, K.; Layton, S.; Stuckey, J.; Yu, Q.; Zhou, W.; Pan, Z.; Sun, Y.; Wang, S. A potent small-molecule inhibitor of the DCN1-UBC12 interaction that selectively blocks cullin 3 neddylation. *Nature Communications* **2017**, *8*, 1150.
17. Scott, D. C.; Hammill, J. T.; Min, J.; Rhee, D. Y.; Connelly, M.; Sviderskiy, V. O.; Bhasin, D.; Chen, Y.; Ong, S. S.; Chai, S. C.; Goktug, A. N.; Huang, G.; Monda, J. K.; Low, J.; Kim, H. S.; Paulo, J. A.; Cannon, J. R.; Shelat, A. A.; Chen, T.; Kelsall, I. R.; Alpi, A. F.; Pagala, V.; Wang, X.; Peng, J.; Singh, B.; Harper, J. W.; Schulman, B. A.; Guy, R. K. Blocking an N-terminal acetylation-dependent protein interaction inhibits an E3 ligase. *Nat Chem Biol* **2017**, *13*, 850-857.
18. Rao, R. C.; Dou, Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat Rev Cancer* **2015**, *15*, 334-46.
19. Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem* **2012**, *81*, 65-95.
20. Dou, Y.; Milne, T. A.; Tackett, A. J.; Smith, E. R.; Fukuda, A.; Wysocka, J.; Allis, C. D.; Chait, B. T.; Hess, J. L.; Roeder, R. G. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **2005**, *121*, 873-85.
21. Djabali, M.; Selleri, L.; Parry, P.; Bower, M.; Young, B. D.; Evans, G. A. A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nat Genet* **1992**, *2*, 113-8.

22. Tkachuk, D. C.; Kohler, S.; Cleary, M. L. Involvement of a homolog of *Drosophila* trithorax by 11q23 chromosomal translocations in acute leukemias. *Cell* **1992**, *71*, 691-700.
23. Muntean, A. G.; Hess, J. L. The pathogenesis of mixed-lineage leukemia. *Annu Rev Pathol* **2012**, *7*, 283-301.
24. Thiel, A. T.; Blessington, P.; Zou, T.; Feather, D.; Wu, X.; Yan, J.; Zhang, H.; Liu, Z.; Ernst, P.; Koretzky, G. A.; Hua, X. MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele. *Cancer Cell* **2010**, *17*, 148-59.
25. Dou, Y.; Milne, T. A.; Ruthenburg, A. J.; Lee, S.; Lee, J. W.; Verdine, G. L.; Allis, C. D.; Roeder, R. G. Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* **2006**, *13*, 713-9.
26. Southall, S. M.; Wong, P. S.; Odho, Z.; Roe, S. M.; Wilson, J. R. Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell* **2009**, *33*, 181-91.
27. Cao, F.; Townsend, E. C.; Karatas, H.; Xu, J.; Li, L.; Lee, S.; Liu, L.; Chen, Y.; Ouillette, P.; Zhu, J.; Hess, J. L.; Atadja, P.; Lei, M.; Qin, Z. S.; Malek, S.; Wang, S.; Dou, Y. Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol Cell* **2014**, *53*, 247-61.
28. Senisterra, G.; Wu, H.; Allali-Hassani, A.; Wasney, G. A.; Barsyte-Lovejoy, D.; Dombrovski, L.; Dong, A.; Nguyen, K. T.; Smil, D.; Bolshan, Y.; Hajian, T.; He, H.; Seitova, A.; Chau, I.; Li, F.; Poda, G.; Couture, J. F.; Brown, P. J.; Al-Awar, R.; Schapira, M.; Arrowsmith, C. H.; Vedadi, M. Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5. *Biochem J* **2013**, *449*, 151-9.

29. Grebien, F.; Vedadi, M.; Getlik, M.; Giambruno, R.; Grover, A.; Avellino, R.; Skucha, A.; Vittori, S.; Kuznetsova, E.; Smil, D.; Barsyte-Lovejoy, D.; Li, F.; Poda, G.; Schapira, M.; Wu, H.; Dong, A.; Senisterra, G.; Stukalov, A.; Huber, K. V. M.; Schonegger, A.; Marcellus, R.; Bilban, M.; Bock, C.; Brown, P. J.; Zuber, J.; Bennett, K. L.; Al-Awar, R.; Delwel, R.; Nerlov, C.; Arrowsmith, C. H.; Superti-Furga, G. Pharmacological targeting of the Wdr5-MLL interaction in C/EBPalpha N-terminal leukemia. *Nat Chem Biol* **2015**, *11*, 571-578.
30. Li, D. D.; Chen, W. L.; Wang, Z. H.; Xie, Y. Y.; Xu, X. L.; Jiang, Z. Y.; Zhang, X. J.; You, Q. D.; Guo, X. K. High-affinity small molecular blockers of mixed lineage leukemia 1 (MLL1)-WDR5 interaction inhibit MLL1 complex H3K4 methyltransferase activity. *Eur J Med Chem* **2016**, *124*, 480-489.
31. Cantoni, G. L. The nature of the active methyl donor formed enzymatically from L-methionine and adenosinetriphosphate. *Journal of the American Chemical Society* **1952**, *74*, 2942-2943.
32. Cantoni, G. L. S-Adenosylmethionine; a new intermediate formed enzymatically from L-methionine and adenosinetriphosphate. *J. Biol. Chem.* **1953**, *204*, 403-416.
33. Fontecave, M.; Atta, M.; Mulliez, E. S-adenosylmethionine: nothing goes to waste. *Trends Biochem Sci* **2004**, *29*, 243-9.
34. Deguchi, T.; Barchas, J. Inhibition of transmethylations of biogenic amines by S-adenosylhomocysteine. Enhancement of transmethylation by adenosylhomocysteinase. *J Biol Chem* **1971**, *246*, 3175-81.

35. Coward, J. K.; D'Urso-Scott, M.; Sweet, W. D. Inhibition of catechol-O-methyltransferase by S-adenosylhomocysteine and S-adenosylhomocysteine sulfoxide, a potential transition-state analog. *Biochem Pharmacol* **1972**, 21, 1200-3.
36. Coward, J. K.; Sweet, W. D. Analogs of S-adenosylhomocysteine as potential inhibitors of biological transmethylation. Synthesis and biological activity of homocysteine derivatives bridged to adenine. *J Med Chem* **1972**, 15, 381-4.
37. Coward, J. K.; Slisz, E. P. Analogs of S-adenosylhomocysteine as potential inhibitors of biological transmethylation. Specificity of the S-adenosylhomocysteine binding site. *J Med Chem* **1973**, 16, 460-3.
38. Borchardt, R. T.; Huber, J. A.; Wu, Y. S. Potential inhibitor of S-adenosylmethionine-dependent methyltransferases. 2. Modification of the base portion of S-adenosylhomocysteine. *J Med Chem* **1974**, 17, 868-73.
39. Borchardt, R. T.; Wu, Y. S. Potential inhibitors of S-adenosylmethionine-dependent methyltransferases. 1. Modification of the amino acid portion of S-adenosylhomocysteine. *J Med Chem* **1974**, 17, 862-8.
40. Borchardt, R. T.; Wu, Y. S. Potential inhibitors of S-adenosylmethionine-dependent methyltransferases. 3. Modifications of the sugar portion of S-adenosylhomocysteine. *J Med Chem* **1975**, 18, 300-4.
41. Chang, C. D.; Coward, J. K. Analogues of S-adenosylhomocysteine as potential inhibitors of biological transmethylation. Synthesis of analogues with modifications at the 5'-thioether linkage. *J Med Chem* **1976**, 19, 684-91.
42. Parang, K.; Cole, P. A. Designing bisubstrate analog inhibitors for protein kinases. *Pharmacol Ther* **2002**, 93, 145-57.



43. Lavogina, D.; Enkvist, E.; Uri, A. Bisubstrate inhibitors of protein kinases: from principle to practical applications. *ChemMedChem* **2010**, *5*, 23-34.
44. Yu, M.; Magalhaes, M. L.; Cook, P. F.; Blanchard, J. S. Bisubstrate inhibition: Theory and application to N-acetyltransferases. *Biochemistry* **2006**, *45*, 14788-94.
45. Rejman, D.; Panova, N.; Klener, P.; Maswabi, B.; Pohl, R.; Rosenberg, I. N-phosphonocarbonylpyrrolidine derivatives of guanine: a new class of bi-substrate inhibitors of human purine nucleoside phosphorylase. *J Med Chem* **2012**, *55*, 1612-21.
46. Mori, S.; Iwase, K.; Iwanami, N.; Tanaka, Y.; Kagechika, H.; Hirano, T. Development of novel bisubstrate-type inhibitors of histone methyltransferase SET7/9. *Bioorganic & Medicinal Chemistry* **2010**, *18*, 8158-8166.
47. Niwa, H.; Handa, N.; Tomabechei, Y.; Honda, K.; Toyama, M.; Ohsawa, N.; Shirouzu, M.; Kagechika, H.; Hirano, T.; Umehara, T.; Yokoyama, S. Structures of histone methyltransferase SET7/9 in complexes with adenosylmethionine derivatives. *Acta Cryst (2013). D69*, 595-602 [doi:10.1107/S0907444912052092] **2013**, *69*, 1-8.
48. Tisi, D.; Chiarparin, E.; Tamanini, E.; Pathuri, P.; Coyle, J. E.; Hold, A.; Holding, F. P.; Amin, N.; Martin, A. C. L.; Rich, S. J.; Berdini, V.; Yon, J.; Acklam, P.; Burke, R.; Drouin, L.; Harmer, J. E.; Jeganathan, F.; van Montfort, R. L. M.; Newbatt, Y.; Tortorici, M.; Westlake, M.; Wood, A.; Hoelder, S.; Heightman, T. D. Structure of the Epigenetic Oncogene MMSET and Inhibition by N-Alkyl Sinefungin Derivatives. *ACS Chemical Biology* **2016**, *11*, 3093-3105.
49. Van Aller, G. S.; Graves, A. P.; Elkins, P. A.; Bonnette, W. G.; McDevitt, P. J.; Zappacosta, F.; Annan, R. S.; Dean, T. W.; Su, D.-S.; Carpenter, C. L.; Mohammad, H.

P.; Kruger, R. G. Structure-Based Design of a Novel SMYD3 Inhibitor that Bridges the SAM-and MEKK2-Binding Pockets. *Structure/Folding and Design* **2016**, 24, 774-781.

50. Zheng, W.; Ibáñez, G.; Wu, H.; Blum, G.; Zeng, H.; Dong, A.; Li, F.; Hajian, T.; Allali-Hassani, A.; Amaya, M. F.; Siarheyeva, A.; Yu, W.; Brown, P. J.; Schapira, M.; Vedadi, M.; Min, J.; Luo, M. Sinefungin Derivatives as Inhibitors and Structure Probes of Protein Lysine Methyltransferase SETD2. *Journal of the American Chemical Society* **2012**, 134, 18004-18014.

51. Kung, P. P.; Huang, B.; Zehnder, L.; Tatlock, J.; Bingham, P.; Krivacic, C.; Gajiwala, K.; Diehl, W.; Yu, X.; Maegley, K. A. SAH derived potent and selective EZH2 inhibitors. *Bioorg Med Chem Lett* **2015**, 25, 1532-7.

52. Daigle, S. R.; Olhava, E. J.; Therkelsen, C. A.; Basavapathruni, A.; Jin, L.; Boriack-Sjodin, P. A.; Allain, C. J.; Klaus, C. R.; Raimondi, A.; Scott, M. P.; Waters, N. J.; Chesworth, R.; Moyer, M. P.; Copeland, R. A.; Richon, V. M.; Pollock, R. M. Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood* **2013**, 122, 1017-25.

53. Anglin, J. L.; Song, Y. A medicinal chemistry perspective for targeting histone H3 lysine-79 methyltransferase DOT1L. *J Med Chem* **2013**, 56, 8972-83.

54. Van Aller, G. S.; Graves, A. P.; Elkins, P. A.; Bonnette, W. G.; McDevitt, P. J.; Zappacosta, F.; Annan, R. S.; Dean, T. W.; Su, D. S.; Carpenter, C. L.; Mohammad, H. P.; Kruger, R. G. Structure-Based Design of a Novel SMYD3 Inhibitor that Bridges the SAM-and MEKK2-Binding Pockets. *Structure* **2016**, 24, 774-781.

55. Ciechanover, A.; Schwartz, A. L. The ubiquitin-proteasome pathway: the complexity and myriad functions of proteins death. *Proc Natl Acad Sci U S A* **1998**, 95, 2727-30.

56. Hershko, A. The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. *Cell Death Differ* **2005**, 12, 1191-1197.
57. Bedford, L.; Lowe, J.; Dick, L. R.; Mayer, R. J.; Brownell, J. E. Ubiquitin-like protein conjugation and the ubiquitin-proteasome system as drug targets. *Nat Rev Drug Discov* **2011**, 10, 29-46.
58. Bulatov, E.; Ciulli, A. Targeting Cullin-RING E3 ubiquitin ligases for drug discovery: structure, assembly and small-molecule modulation. *The Biochemical journal* **2015**, 467, 365-86.
59. Nalepa, G.; Rolfe, M.; Harper, J. W. Drug discovery in the ubiquitin-proteasome system. *Nat Rev Drug Discov* **2006**, 5, 596-613.
60. Zhao, Y. C.; Sun, Y. Cullin-RING Ligases as Attractive Anti-cancer Targets. *Curr Pharm Design* **2013**, 19, 3215-3225.
61. Zhao, Y. C.; Morgan, M. A.; Sun, Y. Targeting Neddylolation Pathways to Inactivate Cullin-RING Ligases for Anticancer Therapy. *Antioxid Redox Sign* **2014**, 21, 2383-2400.
62. Gong, L. M.; Yeh, E. T. H. Identification of the activating and conjugating enzymes of the NEDD8 conjugation pathway. *Journal of Biological Chemistry* **1999**, 274, 12036-12042.
63. Huang, D. T.; Miller, D. W.; Mathew, R.; Cassell, R.; Holton, J. M.; Roussel, M. F.; Schulman, B. A. A unique E1-E2 interaction required for optimal conjugation of the ubiquitin-like protein NEDD8. *Nature Structural & Molecular Biology* **2004**, 11, 927-935.
64. Pan, Z. Q.; Kentsis, A.; Dias, D. C.; Yamoah, K.; Wu, K. Nedd8 on cullin: building an expressway to protein destruction. *Oncogene* **2004**, 23, 1985-1997.

65. Enchev, R. I.; Schulman, B. A.; Peter, M. Protein neddylation: beyond cullin-RING ligases. *Nat Rev Mol Cell Bio* **2015**, 16, 30-44.
66. Oost, T. K.; Sun, C.; Armstrong, R. C.; Al-Assaad, A. S.; Betz, S. F.; Deckwerth, T. L.; Ding, H.; Elmore, S. W.; Meadows, R. P.; Olejniczak, E. T.; Oleksijew, A.; Oltersdorf, T.; Rosenberg, S. H.; Shoemaker, A. R.; Tomaselli, K. J.; Zou, H.; Fesik, S. W. Discovery of potent antagonists of the antiapoptotic protein XIAP for the treatment of cancer. *J Med Chem* **2004**, 47, 4417-26.
67. Sun, H.; Stuckey, J. A.; Nikolovska-Coleska, Z.; Qin, D.; Meagher, J. L.; Qiu, S.; Lu, J.; Yang, C. Y.; Saito, N. G.; Wang, S. Structure-based design, synthesis, evaluation, and crystallographic studies of conformationally constrained Smac mimetics as inhibitors of the X-linked inhibitor of apoptosis protein (XIAP). *J Med Chem* **2008**, 51, 7169-80.
68. Hashimoto, K.; Saito, B.; Miyamoto, N.; Oguro, Y.; Tomita, D.; Shiokawa, Z.; Asano, M.; Kakei, H.; Taya, N.; Kawasaki, M.; Sumi, H.; Yabuki, M.; Iwai, K.; Yoshida, S.; Yoshimatsu, M.; Aoyama, K.; Kosugi, Y.; Kojima, T.; Morishita, N.; Dougan, D. R.; Snell, G. P.; Imamura, S.; Ishikawa, T. Design and synthesis of potent inhibitor of apoptosis (IAP) proteins antagonists bearing an octahydropyrrolo[1,2-a]pyrazine scaffold as a novel proline mimetic. *J Med Chem* **2013**, 56, 1228-46.
69. Matter, H.; Nazare, M.; Gussregen, S.; Will, D. W.; Schreuder, H.; Bauer, A.; Urmann, M.; Ritter, K.; Wagner, M.; Wehner, V. Evidence for C-Cl/C-Br center dot center dot center dot pi Interactions as an Important Contribution to Protein-Ligand Binding Affinity. *Angew Chem Int Edit* **2009**, 48, 2911-2916.

70. Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *Journal of Medicinal Chemistry* **2013**, 56, 1363-1388.
71. Sarkaria, I.; P, O. c.; Talbot, S. G.; Reddy, P. G.; Ngai, I.; Maghami, E.; Patel, K. N.; Lee, B.; Yonekawa, Y.; Dudas, M.; Kaufman, A.; Ryan, R.; Ghossein, R.; Rao, P. H.; Stoffel, A.; Ramanathan, Y.; Singh, B. Squamous cell carcinoma related oncogene/DCUN1D1 is highly conserved and activated by amplification in squamous cell carcinomas. *Cancer Res* **2006**, 66, 9437-44.
72. Molina, D. M.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y. H.; Nordlund, P. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science* **2013**, 341, 84-87.
73. Rudrawar, S.; Kondaskar, A.; Chakraborti, A. K. An efficient acid- and metal-free one-pot synthesis of benzothiazoles from carboxylic acids. *Synthesis-Stuttgart* **2005**, 2521-2526.
74. Jia, L.; Sun, Y. SCF E3 ubiquitin ligases as anticancer targets. *Curr Cancer Drug Targets* **2011**, 11, 347-56.
75. Zhou, H.; Lu, J.; Liu, L.; Bernard, D.; Yang, C. Y.; Fernandez-Salas, E.; Chinnaswamy, K.; Layton, S.; Stuckey, J.; Yu, Q.; Zhou, W.; Pan, Z.; Sun, Y.; Wang, S. A potent small-molecule inhibitor of the DCN1-UBC12 interaction that selectively blocks cullin 3 neddylation. *Nat Commun* **2017**, 8, 1150.
76. Zhou, H.; Zhou, W.; Zhou, B.; Liu, L.; Chern, T. R.; Chinnaswamy, K.; Lu, J.; Bernard, D.; Yang, C. Y.; Li, S.; Wang, M.; Stuckey, J.; Sun, Y.; Wang, S. High-Affinity

Peptidomimetic Inhibitors of the DCN1-UBC12 Protein-Protein Interaction. *J Med Chem* **2018**, 61, 1934-1950.

77. Kurz, T.; Chou, Y. C.; Willems, A. R.; Meyer-Schaller, N.; Hecht, M. L.; Tyers, M.; Peter, M.; Sicheri, F. Dcn1 functions as a scaffold-type E3 ligase for cullin neddylation. *Mol Cell* **2008**, 29, 23-35.

78. Bommelje, C. C.; Weeda, V. B.; Huang, G.; Shah, K.; Bains, S.; Buss, E.; Shaha, M.; Gonen, M.; Ghossein, R.; Ramanathan, S. Y.; Singh, B. Oncogenic function of SCCRO5/DCUN1D5 requires its Neddylation E3 activity and nuclear localization. *Clin Cancer Res* **2014**, 20, 372-81.

79. Meyer-Schaller, N.; Chou, Y. C.; Sumara, I.; Martin, D. D.; Kurz, T.; Katheder, N.; Hofmann, K.; Berthiaume, L. G.; Sicheri, F.; Peter, M. The human Dcn1-like protein DCNL3 promotes Cul3 neddylation at membranes. *Proc Natl Acad Sci U S A* **2009**, 106, 12365-70.

80. Huang, G.; Stock, C.; Bommelje, C. C.; Weeda, V. B.; Shah, K.; Bains, S.; Buss, E.; Shaha, M.; Rechler, W.; Ramanathan, S. Y.; Singh, B. SCCRO3 (DCUN1D3) antagonizes the neddylation and oncogenic activity of SCCRO (DCUN1D1). *J Biol Chem* **2014**, 289, 34728-42.

81. Ma, T.; Shi, T.; Huang, J.; Wu, L.; Hu, F.; He, P.; Deng, W.; Gao, P.; Zhang, Y.; Song, Q.; Ma, D.; Qiu, X. DCUN1D3, a novel UVC-responsive gene that is involved in cell cycle progression and cell growth. *Cancer Sci* **2008**, 99, 2128-35.

82. Monda, J. K.; Scott, D. C.; Miller, D. J.; Lydeard, J.; King, D.; Harper, J. W.; Bennett, E. J.; Schulman, B. A. Structural conservation of distinctive N-terminal

acetylation-dependent interactions across a family of mammalian NEDD8 ligation enzymes. *Structure* **2013**, 21, 42-53.

83. Matsson, P.; Doak, B. C.; Over, B.; Kihlberg, J. Cell permeability beyond the rule of 5. *Adv Drug Deliv Rev* **2016**, 101, 42-61.

84. Sharma, K. K.; Sharma, S.; Kudwal, A.; Jain, R. Room temperature N-arylation of amino acids and peptides using copper(I) and beta-diketone. *Org Biomol Chem* **2015**, 13, 4637-41.

85. Frey, P. A.; Magnusson, O. T. S-Adenosylmethionine: a wolf in sheep's clothing, or a rich man's adenosylcobalamin? *Chem Rev* **2003**, 103, 2129-48.

86. Karatas, H.; Li, Y.; Liu, L.; Ji, J.; Lee, S.; Chen, Y.; Yang, J.; Huang, L.; Bernard, D.; Xu, J.; Townsend, E. C.; Cao, F.; Ran, X.; Li, X.; Wen, B.; Sun, D.; Stuckey, J. A.; Lei, M.; Dou, Y.; Wang, S. Discovery of a Highly Potent, Cell-Permeable Macrocyclic Peptidomimetic (MM-589) Targeting the WD Repeat Domain 5 Protein (WDR5)-Mixed Lineage Leukemia (MLL) Protein-Protein Interaction. *J Med Chem* **2017**, 60, 4818-4839.

87. Campagna-Slater, V. r.; Mok, M. W.; Nguyen, K. T.; Feher, M.; Najmanovich, R.; Schapira, M. Structural Chemistry of the Histone Methyltransferases Cofactor Binding Site. *Journal of Chemical Information and Modeling* **2011**, 51, 612-623.

88. Southall, S. M.; Wong, P.-S.; Odho, Z.; Roe, S. M.; Wilson, J. R. Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks. *Molecular cell* **2009**, 33, 181-191.

89. Zhang, Y.; Mittal, A.; Reid, J.; Reich, S.; Gamblin, S. J.; Wilson, J. R. Evolving Catalytic Properties of the MLL Family SET Domain. *Structure/Folding and Design* **2015**, 23, 1921-1933.

90. Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* **2009**, 49, 377-89.
91. Cave, D. D.; Desiderio, V.; Mosca, L.; Ilisso, C. P.; Mele, L.; Caraglia, M.; Cacciapuoti, G.; Porcelli, M. S-Adenosylmethionine-mediated apoptosis is potentiated by autophagy inhibition induced by chloroquine in human breast cancer cells. *J Cell Physiol* **2018**, 233, 1370-1383.
92. Parashar, S.; Cheishvili, D.; Arakelian, A.; Hussain, Z.; Tanvir, I.; Khan, H. A.; Szyf, M.; Rabbani, S. A. S-adenosylmethionine blocks osteosarcoma cells proliferation and invasion in vitro and tumor metastasis in vivo: therapeutic and diagnostic clinical applications. *Cancer Med* **2015**, 4, 732-44.
93. Wang, Y.; Sun, Z.; Szyf, M. S-adenosyl-methionine (SAM) alters the transcriptome and methylome and specifically blocks growth and invasiveness of liver cancer cells. *Oncotarget* **2017**, 8, 111866-111881.
94. Churcher, I. Protac-Induced Protein Degradation in Drug Discovery: Breaking the Rules or Just Making New Ones? *J Med Chem* **2018**, 61, 444-452.
95. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **1997**, 267, 727-48.
96. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, 52, 609-623.
97. Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* **1997**, 276, 307-326.



98. Vagin, A.; Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D* **2010**, 66, 22-25.
99. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **2004**, 60, 2126-32.
100. Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Sharff, A.; Smart, O.; Vonrhein, C.; Womack, T. *BUSTER, version 2.9*, Global Phasing Ltd.: Cambridge, United Kingdom, 2010.
101. Zhou, W.; Xu, J.; Li, H.; Xu, M.; Chen, Z. J.; Wei, W.; Pan, Z.; Sun, Y. Neddylation E2 UBE2F Promotes the Survival of Lung Cancer Cells by Activating CRL5 to Degrade NOXA via the K11 Linkage. *Clin Cancer Res* **2017**, 23, 1104-1116.
102. Nikolovska-Coleska, Z.; Wang, R.; Fang, X.; Pan, H.; Tomita, Y.; Li, P.; Roller, P. P.; Krajewski, K.; Saito, N. G.; Stuckey, J. A.; Wang, S. Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal Biochem* **2004**, 332, 261-73.
103. Martinez Molina, D.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P. Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **2013**, 341, 84-7.